Forecasting

Petropoulos, Fotios; Apiletti, Daniele; Assimakopoulos, Vassilios; Babai, Mohamed Zied; Barrow, Devon K.; Ben taieb, Souhaib; Bergmeir, Christoph; Bessa, Ricardo J.; Bijak, Jakub; Boylan, John E.; Browell, Jethro; Carnevale, Claudio; Castle, Jennifer L.; Cirillo, Pasquale; Clements, Michael P.; Cordeiro, Clara; Cyrino Oliveira, Fernando Luiz; De Baets, Shari; Dokumentov, Alexander; Ellison, Joanne

DOI:
10.1016/j.ijforecast.2021.11.001

International Journal of Forecasting 38 (2022) 705–871

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Review

Forecasting: theory and practice

Fotios Petropoulos [1,*], Daniele Apiletti [2], Vassilios Assimakopoulos [3], Mohamed Zied Babai [4], Devon K. Barrow [5], Souhaib Ben Taieb [6], Christoph Bergmeir [7], Ricardo J. Bessa [8], Jakub Bijak [9], John E. Boylan [10], Jethro Browell [11], Claudio Carnevale [12], Jennifer L. Castle [13], Pasquale Cirillo [14], Michael P. Clements [15], Clara Cordeiro [16,17], Fernando Luiz Cyrino Oliveira [18], Shari De Baets [19], Alexander Dokumentov [20], Joanne Ellison [9], Piotr Fiszeder [21], Philip Hans Franses [22], David T. Frazier [23], Michael Gilliland [24], M. Sinan Gönül [25], Paul Goodwin [1], Luigi Grossi [26], Yael Grushka-Cockayne [27], Mariangela Guidolin [26], Massimo Guidolin [28], Ulrich Gunter [29], Xiaojia Guo [30], Renato Guseo [26], Nigel Harvey [31], David F. Hendry [32], Ross Hollyman [1], Tim Januschowski [33], Jooyoung Jeon [34], Victor Richmond R. Jose [35], Yanfei Kang [36], Anne B. Koehler [37], Stephan Kolassa [38,10], Nikolaos Kourentzes [39,10], Sonia Leva [40], Feng Li [41], Konstantia Litsiou [42], Spyros Makridakis [43], Gael M. Martin [23], Andrew B. Martinez [44,45], Sheik Meeran [1], Theodore Modis [46], Konstantinos Nikolopoulos [47], Dilek Önkal [25], Alessia Paccagnini [48,49], Anastasios Panagiotelis [50], Ioannis Panapakidis [51], Jose M. Pavía [52], Manuela Pedio [53,54], Diego J. Pedregal [55], Pierre Pinson [56], Patrícia Ramos [57], David E. Rapach [58], J. James Reade [59], Bahman Rostami-Tabar [60], Michał Rubaszek [61], Georgios Sermpinis [62], Han Lin Shang [63], Evangelos Spiliotis [3], Aris A. Syntetos [60], Priyanga Dilini Talagala [64], Thiyanga S. Talagala [65], Len Tashman [66], Dimitrios Thomakos [67], Thordis Thorarinsdottir [68], Ezio Todini [69,70], Juan Ramón Trapero Arenas [55], Xiaoqian Wang [36], Robert L. Winkler [71], Alisa Yusupova [10], Florian Ziel [72]

[1] *School of Management, University of Bath, UK*
[2] *Politecnico di Torino, Turin, Italy*
[3] *Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Greece*
[4] *Kedge Business School, France*
[5] *Department of Management, Birmingham Business School, University of Birmingham, UK*
[6] *Big Data and Machine Learning Lab, Université de Mons (UMONS), Belgium*
[7] *Faculty of Information Technology, Monash University, Melbourne, Australia*
[8] *INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal*
[9] *Department of Social Statistics and Demography, University of Southampton, UK*
[10] *Centre for Marketing Analytics and Forecasting, Lancaster University Management School, Lancaster University, UK*
[11] *School of Mathematics and Statistics, University of Glasgow, UK*
[12] *Department of Mechanical and Industrial Engineering, University of Brescia, Italy*
[13] *Magdalen College, University of Oxford, UK*
[14] *ZHAW School of Management and Law, Zurich University of Applied Sciences, Switzerland*

* Corresponding author.
E-mail addresses: f.petropoulos@bath.ac.uk, fotios@bath.edu (F. Petropoulos).

https://doi.org/10.1016/j.ijforecast.2021.11.001
0169-2070/

[15] ICMA Centre, Henley Business School, University of Reading, UK
[16] Faculdade de Ciências e Tecnologia, Universidade do Algarve, Portugal
[17] CEAUL, Faculdade de Ciências, Universidade de Lisboa, Portugal
[18] Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
[19] Department of Business Informatics and Operations Management, Faculty of Economics and Business Administration, Universiteit Gent, Belgium
[20] Let's Forecast, Australia
[21] Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Torun, Poland
[22] Econometric Institute, Erasmus School of Economics, Rotterdam, The Netherlands
[23] Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia
[24] SAS, USA
[25] Newcastle Business School, Northumbria University, Newcastle upon Tyne, UK
[26] Department of Statistical Sciences, University of Padua, Italy
[27] Darden School of Business, University of Virginia, USA
[28] Finance Department, Bocconi University and Baffi-CAREFIN Centre, Milan, Italy
[29] Department of Tourism and Service Management, MODUL University Vienna, Austria
[30] Robert H. Smith School of Business, University of Maryland, USA
[31] Department of Experimental Psychology, University College London, UK
[32] Nuffield College and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK
[33] Amazon Research, Germany
[34] Korea Advanced Institute of Science and Technology, South Korea
[35] McDonough School of Business, Georgetown University, USA
[36] School of Economics and Management, Beihang University, Beijing, China
[37] Miami University, OH, USA
[38] SAP, Switzerland
[39] Skövde Artificial Intelligence Lab, School of Informatics, University of Skövde, Sweden
[40] Department of Energy, Politecnico di Milano, Italy
[41] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, China
[42] Manchester Metropolitan University Business School, UK
[43] M Open Forecasting Center & Institute for the Future, University of Nicosia, Nicosia, Cyprus
[44] Office of Macroeconomic Analysis, US Department of the Treasury, Washington DC, USA
[45] GWU Research Program on Forecasting, Washington DC, USA
[46] Growth Dynamics, Lugano, Switzerland
[47] Durham University Business School, Durham University, UK
[48] Michael Smurfit Business School, University College Dublin, Ireland
[49] Centre for Applied Macroeconomic Analysis, Australia
[50] Discipline of Business Analytics, The University of Sydney Business School, Australia
[51] Department of Electrical and Computer Engineering, University of Thessaly, Volos, Greece
[52] GIPEyOP, UMMICS, Department of Applied Economics, Universitat de Valencia, Spain
[53] School of Accounting and Finance, University of Bristol, UK
[54] Baffi-CAREFIN Centre, Bocconi University, Italy
[55] ETSI Industrial, Universidad de Castilla-La Mancha, Ciudad Real, Spain
[56] Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[57] Porto Accounting and Business School, Polytechnic of Porto, Portugal
[58] Department of Economics, Saint Louis University, USA
[59] Department of Economics, School of Politics, Economics and International Relations, University of Reading, UK
[60] Cardiff Business School, Cardiff University, UK
[61] SGH Warsaw School of Economics, Collegium of Economic Analysis, Poland
[62] Adam Smith Business School, University of Glasgow, UK
[63] Department of Actuarial Studies and Business Analytics, Macquarie University, Australia
[64] Department of Computational Mathematics, University of Moratuwa, Sri Lanka
[65] Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka
[66] Foresight, International Institute of Forecasters, USA
[67] School of Economics and Political Science, National and Kapodistrian University of Athens, Greece
[68] Norwegian Computing Center, Oslo, Norway
[69] University of Bologna, Italy
[70] Italian Hydrological Society, Bologna, Italy
[71] Fuqua School of Business, Duke University, Durham, USA
[72] House of Energy Markets and Finance, University of Duisburg–Essen, Germany

A R T I C L E I N F O

Keywords:
Review
Encyclopedia
Methods
Applications
Principles
Time series
Prediction

A B S T R A C T

Forecasting has always been at the forefront of decision making and planning. The uncertainty that surrounds the future is both exciting and challenging, with individuals and organisations seeking to minimise risks and maximise utilities. The large number of forecasting applications calls for a diverse set of forecasting methods to tackle real-life challenges. This article provides a non-systematic review of the theory and the practice of forecasting. We provide an overview of a wide range of theoretical, state-of-the-art models, methods, principles, and approaches to prepare, produce, organise, and evaluate forecasts. We then demonstrate how such theoretical concepts are applied in a variety of real-life contexts.

We do not claim that this review is an exhaustive list of methods and applications. However, we wish that our encyclopedic presentation will offer a point of reference for the rich work that has been undertaken over the last decades, with some key insights for the future of forecasting theory and practice. Given its encyclopedic nature, the intended mode of reading is non-linear. We offer cross-references to allow the readers to navigate through the various topics. We complement the theoretical concepts and applications covered by large lists of free or open-source software implementations and publicly-available databases.

Contents

1. Introduction 709
2. Theory 710
 2.1. Introduction to forecasting theory 710
 2.2. Pre-processing data 711
  2.2.1. Box–Cox transformations 711
  2.2.2. Time series decomposition 712
  2.2.3. Anomaly detection and time series forecasting 713
  2.2.4. Robust handling of outliers in time series forecasting 713
  2.2.5. Exogenous variables and feature engineering 714
 2.3. Statistical and econometric models 715
  2.3.1. Exponential smoothing models 715
  2.3.2. Time-series regression models 715
  2.3.3. Theta method and models 716
  2.3.4. Autoregressive integrated moving average (ARIMA) models 717
  2.3.5. Forecasting for multiple seasonal cycles 718
  2.3.6. State-space models 718
  2.3.7. Models for population processes 719
  2.3.8. Forecasting count time series 720
  2.3.9. Forecasting with many variables 721
  2.3.10. Functional time series models 722
  2.3.11. ARCH/GARCH models 723
  2.3.12. Markov switching models 723
  2.3.13. Threshold models 724
  2.3.14. Low and high prices in volatility models 725
  2.3.15. Forecasting with DSGE models 725
  2.3.16. Robust equilibrium-correction forecasting devices 726
  2.3.17. Forecasting with data subject to revision 727
  2.3.18. Innovation diffusion models 727
  2.3.19. The natural law of growth in competition 728
  2.3.20. Synchronic and diachronic competition 729
  2.3.21. Estimation and representation of uncertainty 730
  2.3.22. Forecasting under fat tails 730
 2.4. Bayesian forecasting 731
  2.4.1. Foundations of Bayesian forecasting 731
  2.4.2. Implementation of Bayesian forecasting 732
  2.4.3. Bayesian forecasting with copulas 732
 2.5. Variable and model selection 733
  2.5.1. Leading indicators and Granger causality 733
  2.5.2. Model complexity 734
  2.5.3. Variable selection 734
  2.5.4. Model selection 735
  2.5.5. Cross-validation for time-series data 735
 2.6. Combining forecasts 736
  2.6.1. Forecast combination: a brief review of statistical approaches 736
  2.6.2. Density forecast combinations 737
  2.6.3. Ensembles and predictive probability post processors 737
  2.6.4. The wisdom of crowds 738
 2.7. Data-driven methods 739
  2.7.1. Forecasting with big data 739

2.7.2. Forecasting on distributed systems 739
2.7.3. Agent-based models 740
2.7.4. Feature-based time series forecasting 741
2.7.5. Forecasting with bootstrap 742
2.7.6. Bagging for time series forecasting 742
2.7.7. Multi-step ahead forecasting 743
2.7.8. Neural networks 744
2.7.9. Deep probabilistic forecasting models 744
2.7.10. Machine learning 745
2.7.11. Machine learning with (very) noisy data 746
2.7.12. Clustering-based forecasting 746
2.7.13. Hybrid methods 747
2.8. Methods for intermittent demand 748
2.8.1. Parametric methods for intermittent demand forecasting 748
2.8.2. Non-parametric intermittent demand methods 748
2.8.3. Classification methods 749
2.8.4. Peak over the theshold 749
2.9. Reasoning and mining 750
2.9.1. Fuzzy logic 750
2.9.2. Association rule mining 750
2.9.3. Forecasting with text information 751
2.10. Forecasting by aggregation 752
2.10.1. Cross-sectional hierarchical forecasting 752
2.10.2. Temporal aggregation 753
2.10.3. Cross-temporal hierarchies 753
2.10.4. Ecological inference forecasting 754
2.11. Forecasting with judgment 754
2.11.1. Judgmental forecasting 754
2.11.2. Judgmental adjustments of computer-based forecasts 755
2.11.3. Judgmental model selection 756
2.11.4. Panels of experts 756
2.11.5. Scenarios and judgmental forecasting 757
2.11.6. Trusting model and expert forecasts 758
2.12. Evaluation, validation, and calibration 758
2.12.1. Benchmarking 758
2.12.2. Point, interval, and pHDR forecast error measures 759
2.12.3. Scoring expert forecasts 760
2.12.4. Evaluating probabilistic forecasts 760
2.12.5. Assessing the reliability of probabilistic forecasts 761
2.12.6. Statistical tests of forecast performance 762
2.12.7. Forecasting competitions 762
2.13. The future of forecasting theory 763
3. Practice 764
3.1. Introduction to forecasting practice 764
3.2. Operations and supply chain management 765
3.2.1. Demand management 765
3.2.2. Forecasting in the supply chain 766
3.2.3. Forecasting for inventories 767
3.2.4. Forecasting in retail 767
3.2.5. Promotional forecasting 768
3.2.6. New product forecasting 769
3.2.7. Spare parts forecasting 770
3.2.8. Predictive maintenance 770
3.2.9. Reverse logistics 771
3.3. Economics and finance 771
3.3.1. Macroeconomic survey expectations 771
3.3.2. Forecasting GDP and inflation 772
3.3.3. Forecasting unemployment 773
3.3.4. Forecasting productivity 773
3.3.5. Fiscal forecasting for government budget surveillance 774
3.3.6. Interest rate prediction 775
3.3.7. House price forecasting 776
3.3.8. Exchange rate forecasting 777
3.3.9. Financial time series forecasting with range-based volatility models 777
3.3.10. Copula forecasting with multivariate dependent financial times series 778

3.3.11. Financial forecasting with neural networks 778
3.3.12. Forecasting returns to investment style 779
3.3.13. Forecasting stock returns 780
3.3.14. Forecasting crashes in stock markets 780
3.4. Energy 781
3.4.1. Building energy consumption forecasting and optimisation 781
3.4.2. Electricity price forecasting 782
3.4.3. Load forecasting 785
3.4.4. Crude oil price forecasting 785
3.4.5. Forecasting renewable energy technologies 786
3.4.6. Wind power forecasting 786
3.4.7. Wave forecasting 787
3.4.8. Solar power forecasting 788
3.4.9. Long-term simulation for large electrical power systems 788
3.4.10. Collaborative forecasting in the energy sector 790
3.5. Environmental applications 791
3.5.1. Forecasting two aspects of climate change 791
3.5.2. Weather forecasting 791
3.5.3. Air quality forecasting 793
3.5.4. Forecasting and decision making for floods and water resources management 793
3.6. Social good and demographic forecasting 794
3.6.1. Healthcare 794
3.6.2. Epidemics and pandemics 795
3.6.3. Forecasting mortality 796
3.6.4. Forecasting fertility 797
3.6.5. Forecasting migration 798
3.6.6. Forecasting risk for violence and wars 798
3.7. Systems and humans 799
3.7.1. Support systems 799
3.7.2. Cloud resource capacity forecasting 799
3.7.3. Judgmental forecasting in practice 800
3.7.4. Trust in forecasts 802
3.7.5. Communicating forecast uncertainty 802
3.8. Other applications 803
3.8.1. Tourism demand forecasting 803
3.8.2. Forecasting for aviation 804
3.8.3. Traffic flow forecasting 804
3.8.4. Call arrival forecasting 805
3.8.5. Elections forecasting 807
3.8.6. Sports forecasting 807
3.8.7. Forecasting for megaprojects 808
3.8.8. Competing products 809
3.8.9. Forecasting under data integrity attacks 809
3.8.10. The forecastability of agricultural time series 810
3.8.11. Forecasting in the food and beverage industry 810
3.8.12. Dealing with logistic forecasts in practice 811
3.9. The future of forecasting practice 812
4. Forecasting: benefits, practices, value, and limitations 814
Declaration of competing interest 815
Acknowledgments 815
Appendix A. List of acronyms 815
Appendix B. Software 827
Appendix C. Data sets 827
References 830

1. Introduction[1]

> "In theory, there is no difference between theory and practice. But, in practice, there is."
>
> [Benjamin Brewster (1882)]

Forecasting has come a long way since early humans looked at the sky to see if the weather would be suitable for hunting, and even since hunters could get a forecast such as "a high of 40 with a chance of rain". Now a

[1] This subsection was written by Robert L. Winkler.

hunter can look at a smartphone to instantly get hour-by-hour forecasts of temperatures and probabilities of rain at multiple locations as well as videos of maps showing forecasted weather patterns over the coming hours. Tailored forecasts of increasing sophistication can be generated to inform important decisions of many different types by managers, public officials, investors, and other decision makers.

In the 15 years since the excellent review paper by De Gooijer and Hyndman (2006), the field of forecasting has seen amazing growth in both theory and practice. Thus, this review is both timely and broad, ranging from the highly theoretical to the very applied.

Rapid advances in computing have enabled the analysis of larger and more complex data sets and stimulated interest in analytics and data science. As a result, the forecaster's toolbox of methods has grown in size and sophistication. Computer science has led the way with methods such as neural networks and other types of machine learning, which are getting a great deal of attention from forecasters and decision makers. Other methods, including statistical methods such as Bayesian forecasting and complex regression models, have also benefited from advances in computing. And improvements have not been limited to those based on computing advances. For example, the literature on judgmental forecasting has expanded considerably, driven largely by the "wisdom of crowds" notion.

The combining, or aggregation, of forecasts, which is not a new idea, has received increased attention in the forecasting community recently and has been shown to perform well. For example, the top-performing entries in the M4 Competition run by Spyros Makridakis combined forecasts from multiple methods. Many models have been developed to forecast the number of deaths that will be caused by COVID-19, and combining the forecasts makes sense because it is hard to know which one will be the most accurate. It is consistent with Bayesian ideas since it can be viewed as updating, with each individual forecast added to the combined forecast (also called an ensemble) contributing some new information.

Despite the excitement surrounding these new developments, older methods such as ARIMA and exponential smoothing are still valuable too. Exponential smoothing, along with other simple approaches, are quite robust and not as prone to overfitting as more complex methods. In that sense, they are useful not only on their own merits, but as part of an ensemble that also includes more sophisticated methods. Combined forecasts are more valuable if the forecasts come from methods that are diverse so that their forecast errors are not highly correlated.

The conditions leading to larger, more sophisticated toolboxes for forecasters have also led to larger data sets with denser grids and improved models in areas of application. This has happened with models of the atmosphere, which are important in formulating improved weather forecasts. More detailed information about customers and their preferences allows the development of improved models of customer behaviour for managers. In turn, forecasting methods that can handle all of that information quickly are valuable for decision-making purposes. This process has spurred an explosion in trying to gather information on the internet.

Risk is an important consideration in decision making, and probability forecasts can quantify such risks. Theoretical work in probability forecasting has been active for some time, and decision makers in many areas of practice have embraced the use of probability forecasts. In the Bayesian approach, inferences and forecasts are probabilistic in nature, and probability forecasts can be generated in many other ways too.

The U.S. National Weather Service began issuing probabilities of precipitation to the public in the 1960s. Yet extensive widespread use and dissemination of probabilities has only developed since the turn of the century. Now probability forecasts are increasingly communicated to the public and used as inputs in decision making. Nate Silver's FiveThirtyEight.com report gives probability forecasts for elections, medicine and science, sporting events, economic measures, and many other areas, often looking at multiple forecasting models individually and also combining them.

It is natural for people to desire certainty. When probability forecasts of precipitation were first disseminated widely, many were very sceptical about them, with some accusing the forecasters of hedging and saying "Don't give me a probability. I want to know if it's going to rain or not". Of course, point forecasts often are given along with probability forecasts. The current frequent exposure to probabilities helps the general public better understand, appreciate, and feel more comfortable with them. And the current situation in the world with COVID-19, increases in huge fires, big storms, political polarisation, international conflicts, etc., should help them realise that we are living in an age with huge uncertainties, and forecasts that quantify these uncertainties can be important. Where possible, visualisation can help, as indicated by the saying that a picture is worth a thousand words. Examples are the cones of uncertainty on maps in forecasts of the speed, severity, and future path of hurricanes, and the time line of the probability of a team winning a game, updated quickly after each play.

Put simply, this is an exciting time for the field of forecasting with all of the new theoretical developments and forecasting applications in practice. Forecasting is so ubiquitous that it's not possible to cover all of these developments in a single article. This article manages to cover quite a few, and a good variety. Using short presentations for each one from an expert "close to the ground" on that theoretical topic or field of practice works well to provide a picture of the current state of the art in forecasting theory and practice.

2. Theory

2.1. Introduction to forecasting theory[2]

The theory of forecasting is based on the premise that current and past knowledge can be used to make predictions about the future. In particular for time series,

[2] This subsection was written by Anne B. Koehler.

there is the belief that it is possible to identify patterns in the historical values and successfully implement them in the process of predicting future values. However, the exact prediction of futures values is not expected. Instead, among the many options for a forecast of a single time series at a future time period are an expected value (known as a point forecast), a prediction interval, a percentile and an entire prediction distribution. This set of results collectively could be considered to be "the forecast". There are numerous other potential outcomes of a forecasting process. The objective may be to forecast an event, such as equipment failure, and time series may play only a small role in the forecasting process. Forecasting procedures are best when they relate to a problem to be solved in practice. The theory can then be developed by understanding the essential features of the problem. In turn, the theoretical results can lead to improved practice.

In this introduction, it is assumed that forecasting theories are developed as forecasting methods and models. A forecasting method is defined here to be a predetermined sequence of steps that produces forecasts at future time periods. Many forecasting methods, but definitely not all, have corresponding stochastic models that produce the same point forecasts. A stochastic model provides a data generating process that can be used to produce prediction intervals and entire prediction distributions in addition to point forecasts. Every stochastic model makes assumptions about the process and the associated probability distributions. Even when a forecasting method has an underlying stochastic model, the model is not necessarily unique. For example, the simple exponential smoothing method has multiple stochastic models, including state space models that may or may not be homoscedastic (i.e., possess constant variance). The combining of forecasts from different methods has been shown to be a very successful forecasting method. The combination of the corresponding stochastic models, if they exist, is itself a model. Forecasts can be produced by a process that incorporates new and/or existing forecasting methods/models. Of course, these more complex processes would also be forecasting methods/models.

Consideration of the nature of the variables and their involvement in the forecasting process is essential. In univariate forecasting, the forecasts are developed for a single time series by using the information from the historical values of the time series itself. While in multivariate forecasting, other time series variables are involved in producing the forecasts, as in time series regression. Both univariate and multivariate forecasting may allow for interventions (e.g., special promotions, extreme weather). Relationships among variables and other types of input could be linear or involve nonlinear structures (e.g., market penetration of a new technology). When an explicit functional form is not available, methodologies such as simulation or artificial neural networks might be employed. Theories from fields, such as economics, epidemiology, and meteorology, can be an important part of developing these relationships. Multivariate forecasting could also mean forecasting multiple variables simultaneously (e.g., econometric models).

The data or observed values for time series come in many different forms that may limit or determine the choice of a forecasting method. In fact, there may be no historical observations at all for the item of interest, when judgmental methods must be used (e.g., time taken to complete construction of a new airport). The nature of the data may well require the development of a new forecasting method. The frequency of observations can include all sorts of variations, such as every minute, hourly, weekly, monthly, and yearly (e.g., the electricity industry needs to forecast demand loads at hourly intervals as well as long term demand for ten or more years ahead). The data could be composed of everything from a single important time series to billions of time series. Economic analysis often includes multiple variables, many of which affect one another. Time series for businesses are likely to be important at many different levels (e.g., stock keeping unit, common ingredients, or common size container) and, consequently, form a hierarchy of time series. Some or many of the values might be zero; making the time series intermittent. The list of forms for data is almost endless.

Prior to applying a forecasting method, the data may require pre-processing. There are basic details, such as checking for accuracy and missing values. Other matters might precede the application of the forecasting method or be incorporated into the methods/models themselves. The treatment of seasonality is such a case. Some forecasting method/models require de-seasonalised time series, while others address seasonality within the methods/models. Making it less clear when seasonality is considered relative to a forecasting method/model, some governmental statistical agencies produce forecasts to extend time series into the future in the midst of estimating seasonal factors (i.e., X-12 ARIMA).

Finally, it is extremely important to evaluate the effectiveness of a forecasting method. The ultimate application of the forecasts provides guidance in how to measure their accuracy. The focus is frequently on the difference between the actual value and a point forecast for the value. Many loss functions have been proposed to capture the "average" of these differences. Prediction intervals and percentiles can be used to judge the value of a point forecast as part of the forecast. On the other hand, the quality of prediction intervals and prediction distributions can themselves be evaluated by procedures and formulas that have been developed (e.g., ones based on scoring rules). Another assessment tool is judging the forecasts by metrics relevant to their usage (e.g., total costs or service levels).

In the remaining subsections of Section 2, forecasting theory encompasses both stochastic modelling and forecasting methods along with related aspects.

2.2. Pre-processing data

2.2.1. Box–Cox transformations[3]

A common practice in forecasting models is to transform the variable of interest y using the transformation initially proposed by Box and Cox (1964) as

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}.$$

[3] This subsection was written by Anastasios Panagiotelis.

The range of the transformation will be restricted in a way that depends on the sign of λ, therefore Bickel and Doksum (1981) propose the following modification

$$y^{(\lambda)} = \begin{cases} (|y|^{\lambda} sign(y_i) - 1)/\lambda & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases},$$

which has a range from $(-\infty, \infty)$ for any value of λ. For a recent review of the Box–Cox (and other similar) transformations see Atkinson, Riani, and Corbellini (2021).

The initial motivation for the Box–Cox transformation was to ensure data conformed to assumptions of normality and constant error variance that are required for inference in many statistical models. The transformation nests the log transformation when $\lambda = 0$ and the case of no transformation (up to an additive constant) when $\lambda = 1$. Additive models for $\log(y)$ correspond to multiplicative models on the original scale of y. Choices of λ between 0 and 1 therefore provide a natural continuum between multiplicative and additive models. For examples of forecasting models that use either a log or Box–Cox transformation see Sections 2.3.5 and 2.3.6 and for applications see Sections 3.2.5, 3.6.2 and 3.8.4.

The literature on choosing λ is extensive and dates back to the original Box and Cox (1964) paper - for a review see Sakia (1992). In a forecasting context, a popular method for finding λ is given by Guerrero (1993). The method splits the data into blocks, computes the coefficient of variation within each block and then computes the coefficent of variation again between these blocks. The λ that minimises this quantity is chosen.

Since the transformations considered here are monotonic, the forecast quantiles of the transformed data will, when back-transformed, result in the correct forecast quantiles in terms of the original data. As a result finding prediction intervals in terms of the original data only requires inverting the transformation. It should be noted though, that prediction intervals that are symmetric in terms of the transformed data will not be symmetric in terms of the original data. In a similar vein, back-transformation of the forecast median of the transformed data returns the forecast median in terms of the original data. For more on using the median forecast see Section 2.12.2 and references therein.

The convenient properties that apply to forecast quantiles, do not apply to the forecast mean, something recognised at least since the work of Granger and Newbold (1976). Back-transformation of the forecast mean of the transformed data does not yield the forecast mean of the original data, due to the non-linearity of the transformation. Consequently forecasts on the original scale of the data will be biased unless a correction is used. For some examples of bias correction methods see Granger and Newbold (1976), Guerrero (1993), Pankratz and Dudley (1987) and Taylor (1986b) and references therein.

The issues of choosing λ and bias correcting are accounted for in popular forecasting software packages. Notably, the method of Guerrero (1993) both for finding λ and bias correcting is implemented in the R packages *forecast* and *fable* (see Appendix B).

2.2.2. Time series decomposition[4]

Time series decomposition is an important building block for various forecasting approaches (see, for example, Sections 2.3.3, 2.7.6 and 3.8.3) and a crucial tools for statistical agencies. Seasonal decomposition is a way to present a time series as a function of other time series, called components. Commonly used decompositions are additive and multiplicative, where such functions are summation and multiplication correspondingly. If logs can be applied to time series, any additive decomposition method can serve as multiplicative after applying log transformation to the data.

The simplest additive decomposition of a time series with single seasonality comprises three components: trend, seasonal component, and the "remainder". It is assumed that the seasonal component has a repeating pattern (thus sub-series corresponding to every season are smooth or even constant), the trend component describes the smooth underlying mean and the remainder component is small and contains noise.

The first attempt to decompose time series into trend and seasonality is dated to 1847 when Buys-Ballot (1847) performed decomposition between trend and seasonality, modelling the trend by a polynomial and the seasonality by dummy variables. Then, in 1884 Poynting (1884) proposed price averaging as a tool for eliminating trend and seasonal fluctuations. Later, his approach was extended by Anderson and Nochmals (1914), Hooker (1901) and Spencer (1904). Copeland (1915) was the first who attempted to extract the seasonal component, and Macaulay (1931) proposed a method which is currently considered "classical".

The main idea of this method comes from the observation that averaging a time series with window size of the time series seasonal period leaves the trend almost intact, while effectively removes seasonal and random components. At the next step, subtracting the estimated trend from the data and averaging the result for every season gives the seasonal component. The rest becomes the remainder.

Classical decomposition led to a series of more complex decomposition methods such as X-11 (Shishkin, Young, & Musgrave, 1967), X-11-ARIMA (Dagum, 1988; Ladiray & Quenneville, 2001), X-12-ARIMA (Findley, Monsell, Bell, Otto, & Chen, 1998), and X-13-ARIMA-SEATS (Findley, 2005); see also Section 2.3.4.

Seasonal trend decomposition using Loess (STL: Cleveland, Cleveland, McRae, & Terpenning, 1990) takes iterative approach and uses smoothing to obtain a better estimate of the trend and seasonal component at every iteration. Thus, starting with an estimate of the trend component, the trend component is subtracted from the data, the result is smoothed along sub-series corresponding to every season to obtain a "rough" estimate of the seasonal component. Since it might contain some trend, it is averaged to extract this remaining trend, which is then subtracted to get a detrended seasonal component. This detrended seasonal component is subtracted from the data and the result is smoothed again to obtain a

[4] This subsection was written by Alexander Dokumentov.

better estimate of the trend. This cycle repeats a certain number of times.

Another big set of methods use a single underlining statistical model to perform decomposition. The model allows computation of confidence and prediction intervals naturally, which is not common for iterative and methods involving multiple models. The list of such methods includes TRAMO/SEATS procedure (Monsell, Aston, & Koopman, 2003), the BATS and TBATS models (De Livera, Hyndman, & Snyder, 2011), various structural time series model approaches (Commandeur, Koopman, & Ooms, 2011; Harvey, 1990), and the recently developed seasonal-trend decomposition based on regression (STR: Dokumentov, 2017; Dokumentov & Hyndman, 2018); see also Section 2.3.2. The last mentioned is one of the most generic decomposition methods allowing presence of missing values and outliers, multiple seasonal and cyclic components, exogenous variables with constant, varying, seasonal or cyclic influences, arbitrary complex seasonal schedules. By extending time series with a sequence of missing values the method allows forecasting.

2.2.3. Anomaly detection and time series forecasting[5]

Temporal data are often subject to uncontrolled, unexpected interventions, from which various types of anomalous observations are produced. Owing to the complex nature of domain specific problems, it is difficult to find a unified definition for an anomaly and mostly application-specific (Unwin, 2019). In time series and forecasting literature, an anomaly is mostly defined with respect to a specific context or its relation to past behaviours. The idea of a context is induced by the structure of the input data and the problem formulation (Chandola, Banerjee, & Kumar, 2007, 2009; Hand, 2009). Further, anomaly detection in forecasting literature has two main focuses, which are conflicting in nature: one demands special attention be paid to anomalies as they can be the main carriers of significant and often critical information such as fraud activities, disease outbreak, natural disasters, while the other down-grades the value of anomalies as it reflects data quality issues such as missing values, corrupted data, data entry errors, extremes, duplicates and unreliable values (Talagala, Hyndman, & Smith-Miles, 2020).

In the time series forecasting context, anomaly detection problems can be identified under three major umbrella themes: detection of (*i*) contextual anomalies (point anomalies, additive anomalies) within a given series, (*i*) anomalous sub-sequences within a given series, and (*iii*) anomalous series within a collection of series (Gupta, Gao, Aggarwal, & Han, 2013; Talagala, Hyndman, Smith-Miles, Sevvandi & Muñoz, 2020). According to previous studies forecast intervals are quite sensitive to contextual anomalies and the greatest impact on forecast are from anomalies occurring at the forecast origin (Chen & Liu, 1993a).

The anomaly detection methods in forecasting applications can be categorised into two groups: (*i*) model-based approaches and (*ii*) feature-based approaches. Model-based approaches compare the predicted values with the original data. If the deviations are beyond a certain threshold, the corresponding observations are treated as anomalies (Luo, Hong, & Fang, 2018a; Luo, Hong & Yue, 2018; Sobhani, Hong, & Martin, 2020). Contextual anomalies and anomalous sub-sequences are vastly covered by model-based approaches. Limitations in the detectability of anomalous events depend on the input effects of external time series. Examples of such effects are included in SARIMAX models for polynomial approaches (see also Section 2.3.4). In nonlinear contexts an example is the generalised Bass model (Bass, Krishnan, & Jain, 1994) for special life cycle time series with external control processes (see Section 2.3.18). SARMAX with nonlinear perturbed mean trajectory as input variable may help separating the mean process under control effects from anomalies in the residual process. Feature-based approaches, on the other hand, do not rely on predictive models. Instead, they are based on the time series features measured using different statistical operations (see Section 2.7.4) that differentiate anomalous instances from typical behaviours (Fulcher & Jones, 2014). Feature-based approaches are commonly used for detecting anomalous time series within a large collection of time series. Under this approach, it first forecasts an anomalous threshold for the systems typical behaviour and new observations are identified as anomalies when they fall outside the bounds of the established anomalous threshold (Talagala, Hyndman, Leigh, Mengersen, & Smith-Miles, 2019; Talagala, Hyndman, Smith-Miles & Sevvandi et al., 2020). Most of the existing algorithms involve a manual anomalous threshold. In contrast, Burridge and Robert Taylor (2006) and Talagala, Hyndman, Smith-Miles and Sevvandi et al. (2020) use extreme value theory based data-driven anomalous thresholds. Approaches to the problem of anomaly detection for temporal data can also be divided into two main scenarios: (*i*) batch processing and (*ii*) data streams. The data stream scenario poses many additional challenges, due to nonstationarity, large volume, high velocity, noisy signals, incomplete events and online support (Luo, Hong, Yue, 2018; Talagala, Hyndman, Smith-Miles & Sevvandi et al., 2020).

The performance evaluation of the anomaly detection frameworks is typically done using confusion matrices (Luo, Hong, Yue, 2018; Sobhani et al., 2020). However, these measures are not enough to evaluate the performance of the classifiers in the presence of imbalanced data (Hossin & Sulaiman, 2015). Following Ranawana and Palade (2006) and Talagala et al. (2019), Leigh et al. (2019) have used some additional measures such as negative predictive value, positive predictive value and optimised precision to evaluate the performance of their detection algorithms.

2.2.4. Robust handling of outliers in time series forecasting[6]

Estimators of time series processes can be dramatically affected by the presence of few aberrant observations which are called differently in the time series literature: outliers, spikes, jumps, extreme observations

[5] This subsection was written by Priyanga Dilini Talagala.

[6] This subsection was written by Luigi Grossi.

(see Section 2.2.3). If their presence is neglected, coefficients could be biasedly estimated. Biased estimates of ARIMA processes will decrease the efficiency of predictions (Bianco, García Ben, Martínez, & Yohai, 2001). Moreover, as the optimal predictor of ARIMA models (see Section 2.3.4) is a linear combination of observed units, the largest coefficients correspond to observations near the forecast origin and the presence of outliers among these units can severely affect the forecasts. Proper preliminary analysis of possible extreme observations is an unavoidable step, which should be carried out before any time series modelling and forecasting exercise (see Section 2.3.9). The issue was first raised in the seminal paper by Fox (1972), who suggests a classification of outliers in time series, separating additive outliers (AO) from innovation outliers (IO). The influence of different types of outliers on the prediction errors in conditional mean models (ARIMA models) is studied by Chen and Liu (1993a, 1993b) and Ledolter (1989, 1991), while the GARCH context (see also Section 2.3.11) is explored by Catalán and Trívez (2007) and Franses and Ghijsels (1999). Abraham and Box (1979) propose a Bayesian model which reflects the presence of outliers in time series and allows to mitigate their effects on estimated parameters and, consequently, improve the prediction ability. The main idea is to use a probabilistic framework allowing for the presence of a small group of discrepant units.

A procedure for the correct specification of models, accounting for the presence of outliers, is introduced by Tsay (1986) relying on iterative identification-detection-removal of cycles in the observed time series contaminated by outliers. The same issue is tackled by Abraham and Chuang (1989): in this work non-influential outliers are separated from influential outliers which are observations with high residuals affecting parameter estimation. Tsay's procedure has been later modified (Balke, 1993) to effectively detect time series level shifts. The impulse- and step-indicator saturation approach is used by Marczak and Proietti (2016) for detecting additive outliers and level shifts estimating structural models in the framework of nonstationary seasonal series. They find that timely detection of level shifts located towards the end of the series can improve the prediction accuracy.

All these works are important because outlier and influential observations detection is crucial for improving the forecasting performance of models. The robust estimation of model parameters is another way to improve predictive accuracy without correcting or removing outliers (see Section 3.4.2, for the application on energy data). Sakata and White (1998) introduce a new two-stage estimation strategy for the conditional variance based on Hampel estimators and S-estimators. Park (2002) proposes a robust GARCH model, called RGARCH exploiting the idea of least absolute deviation estimation. The robust approach is also followed for conditional mean models by Gelper, Fried, and Croux (2009) who introduce a robust version of the exponential and Holt-Winters smoothing technique for prediction purposes and by Cheng and Yang (2015) who propose an outlier resistant algorithm developed starting from a new synthetic loss function. Very recently, Beyaztas and Shang (2019) have introduced a robust forecasting procedure based on weighted likelihood estimators to improve point and interval forecasts in functional time series contaminated by the presence of outliers.

2.2.5. Exogenous variables and feature engineering[7]

Exogenous variables are those included in a forecasting system because they add value but are not being predicted themselves, and are sometimes called 'features' (see Section 2.7.4). For example, a forecast of county's energy demand may be based on the recent history of demand (an *endogenous* variable), but also weather forecasts, which are exogenous variables. Many time series methods have extensions that facilitate exogenous variables, such as autoregression with exogenous variables (ARX). However, it is often necessary to prepare exogenous data before use, for example so that it matches the temporal resolution of the variable being forecast (hourly, daily, and so on).

Exogenous variables may be numeric or categorical, and may be numerous. Different types of predictor present different issues depending on the predictive model being used. For instance, models based on the variable's absolute value can be sensitive to extreme values or skewness, whereas models based on the variable value's rank, such as tree-based models, are not. Exogenous variables that are correlated with one another also poses a challenge for some models, and techniques such as regularisation and partial leased squares have been developed to mitigate this.

Interactions between exogenous variables my also be important when making predictions. For example, crop yields depend on both rainfall and sunlight: one without the other or both in excess will result in low yields, but the right combination will result in high yields. Interactions may be included in linear models by including product of the two interacting exogenous as a feature in the model. This is an example of feature engineering, the process of creating new features based on domain knowledge or exploratory analysis of available data. In machine learning (see Section 2.7.10), many features may be created by combining exogenous variables speculatively and passed to a selection algorithm to identify those with predictive power. Combinations are not limited to products, or only two interacting variables, and where many exogenous variables are available, could include summary statistics (mean, standard deviation, range, quantiles...) of groups of variables.

Where exogenous variables are numerous dimension reduction may be applied to reduce the number of features in a forecasting model (see also Section 2.5.3). Dimension reduction transforms multivariate data into a lower dimensional representation while retaining meaningful information about the original data. Principal component analysis (PCA) is a widely used method for linear dimension reduction, and non-linear alternatives are also available. PCA is useful when the number of candidate predictors is greater than the number of time series observations, as is often the case in macroeconomic forecasting (Stock & Watson, 2002). It is routinely applied in

[7] This subsection was written by Jethro Browell.

applications from weather to sales forecasting. In retail forecasting, for example, past sales of thousands of products may be recorded but including them all as exogenous variables in the forecasting model for an individual product may be impractical. Dimension reduction offers an alternative to only using a subset of the available features.

Preparation of data for forecasting tasks is increasingly important as the volume of available data is increasing in many application areas. Further details and practical examples can be found in Albon (2018) and Kuhn and Johnson (2019) among other texts in this area. For deeper technical discussion of a range of non-linear dimension reduction algorithms, see Hastie, Tibshirani, and Friedman (2009).

2.3. Statistical and econometric models

2.3.1. Exponential smoothing models[8]

Exponential smoothing is one of the workhorses of business forecasting. Despite the many advances in the field, it is always a tough benchmark to bear in mind. The development of exponential smoothing dates back to 1944, where Robert G. Brown through a mechanical computing device estimated key variables for fire-control on the location of submarines (Gardner, 1985). More details about the state of the art of exponential smoothing can be found in Gardner (2006).

The idea behind exponential smoothing relies on the weighted average of past observations, where that weight decreases exponentially as one moves away from the present observations. The appropriate exponential smoothing method depends on the components that appear in the time series. For instance, in case that no clear trend or seasonal pattern is present, the simplest form of exponential smoothing methods known as Simple (or Single) Exponential Smoothing (SES) is adequate, such as:

$$f_{t+1} = \alpha y_t + (1 - \alpha) f_t$$

In some references, is also known as Exponentially Weighted Moving Average (Harvey, 1990). The formula for SES can be obtained from minimising the discounted least squares error function and expressing the resulting equation in a recursive form (Harvey, 1990). If observations do not have the same weight, the ordinary least squares cannot be applied. On the other hand, the recursive form is very well-suited for saving data storage.

In order to use SES, we need to estimate the initial forecast (f_1) and the exponential smoothing parameter (α). Traditionally, the initialisation was done by using either ad hoc values or a heuristic scheme (Hyndman, Koehler, Ord, & Snyder, 2008), however nowadays it is standard to estimate both the initial forecast and the optimal smoothing parameter by minimising the sum of squares of the one-step ahead forecast errors. The estimation of the smoothing parameter usually is restricted to values between 0 and 1. Once SES is defined, the method only provides point forecasts, i.e., forecasts of the mean. Nonetheless, it is of vital importance for many applications to provide density (probabilistic) forecasts. To that end, Hyndman, Koehler, Snyder, and Grose (2002) extended exponential smoothing methods under State Space models using a single source of error (see Section 2.3.6) to equip them with a statistical framework capable of providing future probability distributions. For example, SES can be expressed in the State Space as a local level model:

$$\begin{aligned} y_t &= \ell_{t-1} + \epsilon_t, \\ \ell_t &= \ell_{t-1} + \alpha\epsilon_t. \end{aligned}$$

where the state is the level (ℓ) and ϵ is the Gaussian noise. Note the difference between traditional exponential smoothing methods and exponential smoothing models (under the state space approach). The former only provide point forecasts, meanwhile the latter also offers probabilistic forecasts, which obviously includes prediction intervals. In addition, some exponential smoothing models can be expressed an ARIMA models (see also Section 2.3.4).

So far, we have introduced the main exponential smoothing using SES, however real time series can include other components as trends, seasonal patterns, cycles, and the irregular (error) component. In this sense, the exponential smoothing version capable of handling local trends is commonly known as Holt's method (Holt, 2004, originally published in 1957) and, if it also models a seasonality component, which can be incorporated in an additive or multiplicative fashion, it is called Holt-Winters method (Winters, 1960). Exponential smoothing models have been also extended to handle multiple seasonal cycles; see Section 2.3.5.

Fortunately, for various combinations of time series patterns (level only, trended, seasonal, trended and seasonal) a particular exponential smoothing can be chosen. Pegels (1969) proposed a first classification of exponential smoothing methods, later extended by Gardner (1985) and Taylor (2003a). The state space framework mentioned above, developed by Hyndman et al. (2002), allowed to compute the likelihood for each exponential smoothing model and, thus, model selection criteria such as AIC could be used to automatically identify the appropriate exponential smoothing model. Note that the equivalent state space formulation was derived by using a single source of error instead of a multiple source of error (Harvey, 1990). Hyndman et al. (2008) utilised the notation (E,T,S) to classify the exponential smoothing models, where those letters refer to the following components: Error, Trend, and Seasonality. This notation has gained popularity because the widely-used *forecast* package (Hyndman et al., 2020), recently updated to the *fable* package, for R statistical software, and nowadays exponential smoothing is frequently called ETS.

2.3.2. Time-series regression models[9]

The key idea of linear regression models is that a target (or dependent, forecast, explained, regress) variable, y, i.e., a time series of interest, can be forecast through other regressor (or independent, predictor, explanatory)

[8] This subsection was written by Juan Ramón Trapero Arenas.

[9] This subsection was written by Vassilios Assimakopoulos.

variables, x, i.e., time series or features (see Section 2.2.5), assuming that a linear relationship exists between them, as follows

$$y_t = \beta_0 + \beta_1 x_{1_t} + \beta_2 x_{2_t} + \cdots + \beta_k x_{k_t} + e_t,$$

where e_t is the residual error of the model at time t, β_0 is a constant, and coefficient β_i is the effect of regressor x_i after taking into account the effects of all k regressors involved in the model. For example, daily product sales may be forecast using information related with past sales, prices, advertising, promotions, special days, and holidays (see also Section 3.2.4).

In order to estimate the model, forecasters typically minimise the sum of the squared errors (ordinary least squares estimation, OLS), $\text{SSE} = \sum_{t=1}^{n} e_t^2$, using the observations available for fitting the model to the data (Ord, Fildes, & Kourentzes, 2017) and setting the gradient $\frac{\partial \text{SSE}}{\partial \beta_i}$ equal to zero. If the model is simple, consisting of a single regressor, then two coefficients are computed, which are the slope (coefficient of the regressor) and the intercept (constant). When more regressor variables are considered, the model is characterised as a multiple regression one and additional coefficients are estimated.

A common way to evaluate how well a linear regression model fits the target series, reporting an average value of $\bar{y}$, is through the coefficient of determination, $R^2 = \frac{\sum_{t=1}^{n}(f_t-\bar{y})^2}{\sum_{t=1}^{n}(y_t-\bar{y})^2}$, indicating the proportion of variation in the dependent variable explained by the model. Values close to one indicate sufficient goodness-of-fit, while values close to zero insufficient fitting. However, goodness-of-fit should not be confused with forecastability (Harrell, 2015). When the complexity of the model is increased, i.e., more regressors are considered, the value of the coefficient will also rise, even if such additions lead to overfitting (see Section 2.5.2). Thus, regression models should be evaluated using cross-validation approaches (see Section 2.5.5), approximating the post-sample accuracy of the model, or measures that account for model complexity, such as information criteria (e.g., AIC, AICc, and BIC) and the adjusted coefficient of determination, $\bar{R}^2 = 1-(1-R^2)\frac{n-1}{n-k-1}$ (James, Witten, Hastie, & Tibshirani, 2013). Other diagnostics are the standard deviation of the residuals and the t-values of the regressors. Residual standard error, $\sigma_e = \sqrt{\frac{\sum_{t=1}^{n}(y_t-f_t)^2}{n-k-1}}$, summarises the average error produced by the model given the number of regressors used, thus accounting for overfitting. The t-values measure the impact of excluding regressors from the model in terms of error, given the variation in the data, thus highlighting the importance of the regressors.

To make sure that the produced forecasts are reliable, the correlation between the residuals and the observations of the regressors must be zero, with the former displaying also insignificant autocorrelation. Other assumptions suggest that the residuals should be normally distributed with an average value of zero and that their variability should be equal across time (no heteroscedasticity present). Nevertheless, in practice, it is rarely necessary for residuals to be normally distributed in order for the model to produce accurate results, while the homoscedasticity assumption becomes relevant mostly when computing prediction intervals. If these assumptions are violated, that may mean that part of the variance of the target variable has not been explained by the model and, therefore, that other or more regressors are needed. In case of non-linear dependencies between the target and the regressor variables, data power transformations (see Section 2.2.1) or machine learning approaches can be considered (see Section 2.7.10).

Apart from time series regressors, regression models can also exploit categorical (dummy or indicator) variables (Hyndman & Athanasopoulos, 2018) which may e.g., inform the model about promotions, special events, and holidays (binary variables), the day of the week or month of the year (seasonal dummy variables provided as one-hot encoded vectors), trends and structural changes, and the number of trading/working days included in the examined period. In cases where the target series is long and displays complex seasonal patterns, additional regressors such as Fourier series and lagged values of both the target and the regressor variables may become useful. Moreover, when the number of the potential regressor variables is significant compared to the observations available for estimating the respective coefficients (see Section 2.7.1), step-wise regression (James et al., 2013) or dimension reduction and shrinkage estimation methods (see Section 2.5.3) can be considered to facilitate training and avoid overfitting. Finally, mixed data sampling (MIDAS) regression models are a way of allowing different degrees of temporal aggregation for the regressors and predictand (see also Section 2.10.2 for further discussions on forecasting with temporal aggregation).

2.3.3. Theta method and models[10]

In the age of vast computing power and computational intelligence, the contribution of simple forecasting methods is possibly not *en vogue*; the implementation of complicated forecasting systems becomes not only expedient but possibly desirable. Nevertheless forecasting, being a tricky business, does not always favour the complicated or the computationally intensive. Enter the theta method. From its beginnings 20 years back in Assimakopoulos and Nikolopoulos (2000) to recent advances in the monograph of Nikolopoulos and Thomakos (2019), to other work in-between and recently too, the theta method has emerged as not only a powerfully simple but also enduring method in modern time series forecasting. The reader will benefit by reviewing Sections 2.3.1, 2.3.4 and 2.3.9 for useful background information.

The original idea has been now fully explained and understood and, as Nikolopoulos and Thomakos (2019) have shown, even the revered AR(1) model forecast is indeed a theta forecast – and it has already been shown by Hyndman and Billah (2003) that the theta method can represent SES (with a drift) forecasts as well. In its simplest form the method generates a forecast from a linear combination of the last observation and some form of "trend" function, be that a constant, a linear trend, a non-parametric trend or a non-linear trend. In summary, and

[10] This subsection was written by Dimitrios Thomakos.

under the conditions outlined extensively in Nikolopoulos and Thomakos (2019), the theta forecasts can be expressed as functions of the "theta line":

$$Q_t(\theta) = \theta y_t + (1-\theta)T_{t+1}$$

where T_{t+1} is the trend function, variously defined depending on the modelling approach and type of trend one is considering in applications. It can be shown that the, univariate, theta forecasts can given either as

$$f_{t+1|t} = y_t + \Delta Q_t(\theta)$$

when the trend function is defined as $T_{t+1} = \mu t$ and as

$$f_{t+1|t} = Q_t(\theta) + \theta \Delta \mathbb{E}(T_{t+1})$$

when the trend function is left otherwise unspecified. The choice of the weight parameter θ on the linear combination of the theta line, the choice and number of trend functions and their nature and other aspects on expanding the method have been recently researched extensively.

The main literature has two strands. The first one details the probabilistic background of the method and derives certain theoretical properties, as in Hyndman and Billah (2003), Thomakos and Nikolopoulos (2012, 2015) and a number of new theoretical results in Nikolopoulos and Thomakos (2019). The work of Thomakos and Nikolopoulos provided a complete analysis of the theta method under the unit root data generating process, explained its success in the M3 competition (Makridakis & Hibon, 2000), introduced the multivariate theta method and related it to cointegration and provided a number of other analytical results for different trend functions and multivariate forecasting. The second strand of the literature expands and details various implementation (including hybrid approaches) of the method, as in the theta approach in supply chain planning of Nikolopoulos, Assimakopoulos, Bougioukos, Litsa, and Petropoulos (2012), the optimised theta models and their relationship with state space models in Fioruci, Pellegrini, Louzada, and Petropoulos (2015) and Fiorucci, Pellegrini, Louzada, Petropoulos, and Koehler (2016), hybrid approaches as in Theodosiou (2011) and Spiliotis, Assimakopoulos and Nikolopoulos (2019), to the very latest generalised theta method of Spiliotis, Assimakopoulos and Makridakis (2020). These are major methodological references in the field, in addition to many others of pure application of the method.

The theta method is also part of the family of adaptive models/methods, and a simple example illustrates the point: the AR(1) forecast or the SES forecast are both theta forecasts but they are also both adaptive learning forecasts, as in the definitions of the recent work by Kyriazi, Thomakos, and Guerard (2019). As such, the theta forecasts contain the basic building blocks of successful forecasts: simplicity, theoretical foundations, adaptability and performance enhancements. Further research on the usage of the theta method within the context of adaptive learning appears to be a natural next step. In the context of this section, see also Section 2.3.16 on equilibrium correcting models and forecasts.

Given the simplicity of its application, the freely available libraries of its computation, its scalability and performance, the theta method should be considered as a critical benchmark henceforth in the literature – no amount of complexity is worth its weight if it cannot beat a single Greek letter!

2.3.4. Autoregressive integrated moving average (ARIMA) models[11]

Time series models that are often used for forecasting are of the autoregressive integrated moving average class (ARIMA: Box, George, Jenkins, & Gwilym, 1976). The notation of an ARIMA(p, d, q) model for a time series y is

$$(1-\phi_1 L - \cdots - \phi_p L^p)(1-L)^d y_t = c + (1+\theta_1 L + \cdots + \theta_q L^q) + \epsilon_t,$$

where the lag operator L is defined by $L^k y_t = y_{t-k}$. The ϵ_t is a zero-mean uncorrelated process with common variance σ_ϵ^2. Some exponential smoothing models (see Section 2.3.1) can also be written in ARIMA format, where some ETS models assume that $d = 1$ or $d = 2$. For example, SES is equivalent to ARIMA(0,1,1) when $\theta_1 = \alpha - 1$.

The parameters in the ARIMA model can be estimated using Maximum Likelihood, whereas for the ARIMA(p, d, 0) Ordinary Least Squares can be used. The iterative model-building process (Franses, van Dijk, & Opschoor, 2014) requires the determination of the values of p, d, and q. Data features as the empirical autocorrelation function and the empirical partial autocorrelation function can be used to identify the values of p and q, in case of low values of p and q. Otherwise, in practice one relies on the well-known information criteria like AIC and BIC (see Section 2.5.4). The function `auto.arima` of the *forecast* package (Hyndman et al., 2020) for R statistical software compares models using information criteria, and has been found to be very effective and increasingly being used in ARIMA modelling.

Forecasts from ARIMA models are easy to make. And, at the same time, prediction intervals can be easily computed. Take for example the ARIMA(1,0,1) model:

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}.$$

The one-step-ahead forecast from forecast origin n is $f_{n+1|n} = c + \phi_1 y_n + \theta_1 \epsilon_n$ as the expected value $E(\epsilon_{n+1}) = 0$. The forecast error is $y_{n+1} - f_{n+1|n} = \epsilon_{n+1}$ and, hence, the forecast error variance is σ_ϵ^2. The two-steps-ahead forecast from n is $f_{n+2|n} = c + \phi_1 f_{n+1|n}$ with the forecast error equal to $\epsilon_{n+2} + \phi_1 \epsilon_{n+1}$ and the forecast error variance $(1+\phi_1^2)\sigma_\epsilon^2$. These expressions show that the creation of forecasts and forecast errors straightforwardly follow from the model expressions, and hence can be automated if necessary.

An important decision when using an ARIMA model is the choice for the value of d. When $d = 0$, the model is created for the levels of the time series, that is, y_t. When $d = 1$, there is a model for $(1-L)y_t$, and the data need to be differenced prior to fitting an ARMA model. In some specific but rare cases, $d = 2$. The decision on the value of

[11] This subsection was written by Philip Hans Franses & Sheik Meeran.

d is usually based on so-called tests for unit roots (Dickey & Fuller, 1979; Dickey & Pantula, 1987). Under the null hypothesis that $d = 1$, the data are non-stationary, and the test involves non-standard statistical theory (Phillips, 1987). One can also choose to make $d = 0$ as the null hypothesis (Hobijn, Franses, & Ooms, 2004; Kwiatkowski, Phillips, Schmidt, & Shin, 1992). The power of unit root tests is not large, and in practice one often finds signals to consider $d = 1$ (Nelson & Plosser, 1982).

For seasonal data, like quarterly and monthly time series, the ARIMA model can be extended to Seasonal ARIMA (SARIMA) models represented by ARIMA$(p, d, q)(P, D, Q)_s$, where P, D, and Q are the seasonal parameters and the s is the periodicity. When $D = 1$, the data are transformed as $(1 - L^s)y_t$. It can also be that $D = 0$ and $d = 1$, and then one can replace c by $c_1D_{1,t} + c_2D_{2,t} + \cdots + c_sD_{s,t}$ where the $D_{i,t}$ with $i = 1, 2, \ldots, s$ are seasonal dummies. The choice of D is based on tests for so-called seasonal unit roots (Franses, 1991; Ghysels, Lee, & Noh, 1994; Hylleberg, Engle, Granger, & Yoo, 1990).

Another popular extension to ARIMA models is called ARIMAX, implemented by incorporating additional exogenous variables (regressors) that are external to and different from the forecast variable. An alternative to ARIMAX is the use of regression models (see Section 2.3.2) with ARMA errors.

2.3.5. Forecasting for multiple seasonal cycles[12]

With the advances in digital data technologies, data is recorded more frequently in many sectors such as energy (Wang, Liu, & Hong, 2016, and Section 3.4), healthcare (Whitt & Zhang, 2019, and 3.6.1), transportation (Gould, Koehler, Ord, Snyder, Hyndman, & Vahid-Araghi, 2008), and telecommunication (Meade & Islam, 2015a). This often results in time series that exhibit multiple seasonal cycles (MSC) of different lengths. Forecasting problems involving such series have been increasingly drawing the attention of both researchers and practitioners leading to the development of several approaches.

Multiple Linear Regression (MLR) is a common approach to model series with MSC (Kamisan, Lee, Suhartono, Hussin, & Zubairi, 2018; Rostami-Tabar & Ziel, 2020); for an introduction on time-series regression models, see Section 2.3.2. While MLR is fast, flexible, and uses exogenous regressors, it does not allow to decompose components and change them over time. Building on the foundation of the regression, Facebook introduced Prophet (Taylor & Letham, 2018), an automated approach that utilises the Generalised Additive Model (Hastie & Tibshirani, 1990). Although the implementation of Prophet may be less flexible, it is easy to use, robust to missing values and structural changes, and can handles outliers well.

Some studies have extended the classical ARIMA (see Section 2.3.4) and Exponential Smoothing (ETS; see Section 2.3.1) methods to accommodate MSC. Multiple/multiplicative Seasonal ARIMA (MSARIMA) model is an extension of ARIMA for the case of MSC (Taylor, 2003b). MSARIMA allows for exogenous regressors and terms can evolve over time, however, it is not flexible, and the computational time is high. Svetunkov and Boylan (2020) introduced the Several Seasonalities ARIMA (SSARIMA) model which constructs ARIMA in a state-space form with several seasonalities. While SSARIMA is flexible and allows for exogenous regressors, it is computationally expensive, especially for high frequency series.

Taylor (2003b) introduced Double Seasonal Holt-Winters (DSHW) to extend ETS for modelling daily and weekly seasonal cycles. Following that, Taylor (2010) proposed a triple seasonal model to consider the intraday, intraweek and intrayear seasonalities. Gould et al. (2008) and Taylor and Snyder (2012) instead proposed an approach that combines a parsimonious representation of the seasonal states up to a weekly period in an innovation state space model. With these models, components can change, and decomposition is possible. However, the implementation is not flexible, the use of exogenous regressors is not supported, and the computational time could be high.

An alternative approach for forecasting series with MSC is TBATS (De Livera et al., 2011, see also Section 2.2.2). TBATS uses a combination of Fourier terms with an exponential smoothing state space model and a Box–Cox transformation (see Section 2.2.1), in an entirely automated manner. It allows for terms to evolve over time and produce accurate forecasts. Some drawbacks of TBATS, however, are that it is not flexible, can be slow, and does not allow for covariates.

In response to shortcomings in current models, Forecasting with Additive Switching of Seasonality, Trend and Exogenous Regressors (FASSTER) has been proposed by O'Hara-Wild and Hyndman (2020). FASSTER is fast, flexible and support the use of exogenous regressors into a state space model. It extends state space models such as TBATS by introducing a switching component to the measurement equation which captures groups of irregular multiple seasonality by switching between states.

In recent years, Machine Learning (ML; see Section 2.7.10) approaches have also been recommended for forecasting time series with MSC. MultiLayer Perceptron (MLP: Dudek, 2013; Zhang & Qi, 2005), Recurrent Neural Networks (RNN: Lai, Chang, Yang, & Liu, 2018), Generalised Regression Neural Network (GRNN: Dudek, 2015), and Long Short-Term Memory Networks (LSTM: Zheng, Xu, Zhang, & Li, 2017) have been applied on real data (Bandara, Bergmeir, & Hewamalage, 2020a; Xie & Ding, 2020) with promising results. These approaches are flexible, allow for any exogenous regressor and suitable when non-linearity exists in series, however interpretability might be an issue for users (Makridakis, Spiliotis, & Assimakopoulos, 2018).

2.3.6. State-space models[13]

State Space (SS) systems are a very powerful and useful framework for time series and econometric modelling and forecasting. Such systems were initially developed by engineers, but have been widely adopted and developed in Economics as well (Durbin & Koopman, 2012;

[12] This subsection was written by Bahman Rostami-Tabar.

[13] This subsection was written by Diego J. Pedregal.

Harvey, 1990). The main distinguishing feature of SS systems is that the model is formulated in terms of *states* ($\boldsymbol{\alpha}_t$), which are a set of variables usually unobserved, but which have some meaning. Typical examples are trends, seasonal components or time varying parameters.

A SS system is built as the combination of two sets of equations: (*i*) *state* or *transition* equations which describe the dynamic law governing the states between two adjacent points in time; and (*ii*) *observation* equations which specify the relation between observed data (both inputs and outputs) and the unobserved states. A linear version of such a system is shown in Eq. (1).

$$\begin{aligned} \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t\boldsymbol{\alpha}_t + \boldsymbol{\Gamma}_t + \mathbf{R}_t\boldsymbol{\eta}_t, \quad & \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}_t) \\ \mathbf{y}_t &= \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{D}_t + \mathbf{C}_t\boldsymbol{\epsilon}_t, \quad & \boldsymbol{\epsilon}_t \sim N(0, \mathbf{H}_t) \\ \boldsymbol{\alpha}_1 &\sim N(\boldsymbol{a}_1, \mathbf{P}_1) \end{aligned} \tag{1}$$

In this equations $\boldsymbol{\eta}_t$ and $\boldsymbol{\epsilon}_t$ are the state and observational vectors of zero mean Gaussian noises with covariance $\mathbf{S}_t$. $\mathbf{T}_t$, $\boldsymbol{\Gamma}_t$, $\mathbf{R}_t$, $\mathbf{Q}_t$, $\mathbf{Z}_t$, $\mathbf{D}_t$, $\mathbf{C}_t$, $\mathbf{H}_t$ and $\mathbf{S}_t$ are the so-called (time-varying) system matrices, and $\boldsymbol{a}_1$ and $\mathbf{P}_1$ are the initial state and state covariance matrix, respectively. Note that $\mathbf{D}_t$ and $\boldsymbol{\Gamma}_t$ may be parameterised to include some input variables as linear or non-linear relations to the output variables $\mathbf{y}_t$.

The model in Eq. (1) is a *multiple error SS model*. A different formulation is the *single error SS model* or the *innovations SS model*. This latter is similar to (1), but replacing $\mathbf{R}_t\boldsymbol{\eta}_t$ and $\mathbf{C}_t\boldsymbol{\epsilon}_t$ by $\mathbf{K}_t\mathbf{e}_t$ and $\mathbf{e}_t$, respectively. Then, naturally, the innovations form may be seen as a restricted version of model (1), but, conversely, under weak assumptions, (1) may also be written as an observationally equivalent *innovations form* (see, for example, Casals, Garcia-Hiernaux, Jerez, Sotoca, & Trindade, 2016, pp. 12–17).

Once a SS system is fully specified, the core problem is to provide optimal estimates of states and their covariance matrix over time. This can be done in two ways, either by looking back in time using the well-known *Kalman filter* (useful for online applications) or taking into account the whole sample provided by smoothing algorithms (typical of offline applications) (Anderson & Moore, 1979).

Given any set of data and a specific model, the system is not fully specified in most cases because it usually depends on unknown parameters scattered throughout the system matrices that define the SS equations. Estimation of such parameters is normally carried out by Maximum Likelihood defined by prediction error decomposition (Harvey, 1990).

Non-linear and non-Gaussian models are also possible, but at the cost of a higher computational burden because more sophisticated recursive algorithms have to be applied, like the extended Kalman filters and smoothers of different orders, particle filters (Doucet & Gordon, 2001), Unscented Kalman filter and smoother (Julier & Uhlmann, 1997), or simulation of many kinds, like Monte Carlo, bootstrapping or importance sampling (Durbin & Koopman, 2012).

The paramount advantage of SS systems is that they are not a particular model or family of models strictly speaking, but a container in which many very different model families may be implemented, indeed many treated in other sections of this paper. The following is a list of possibilities, not at all exhaustive:

- Univariate models with or without inputs: regression (Section 2.3.2), ARIMAx (Section 2.3.4), transfer functions, exponential smoothing (Section 2.3.1), structural unobserved components, Hodrick–Prescott filter, spline smoothing.
- Fully multivariate: natural extensions of the previous ones plus echelon-form VARIMAx, Structural VAR, VECM, Dynamic Factor models, panel data (Section 2.3.9).
- Non-linear and non-Gaussian: TAR, ARCH, GARCH (Section 2.3.11), Stochastic Volatility (Durbin & Koopman, 2012), Dynamic Conditional Score (Harvey, 2013), Generalised Autoregressive Score (Creal, Koopman, & Lucas, 2013), multiplicative unobserved components.
- Other: periodic cubic splines, periodic unobserved components models, state dependent models, Gegenbauer long memory processes (Dissanayake, Peiris, & Proietti, 2018).

Once any researcher or practitioner becomes acquainted to a certain degree with the SS technology, some important advanced issues in time series forecasting may be comfortably addressed (Casals et al., 2016). It is the case, for example, of systems block concatenation, systems nesting in errors or in variables, treating errors in variables, continuous time models, time irregularly spaced data, mixed frequency models, time varying parameters, time aggregation, hierarchical and group forecasting (Villegas & Pedregal, 2018) (time, longitudinal or both), homogeneity of multivariate models (proportional covariance structure among perturbations), etc.

All in all, the SS systems offer a framework capable of handling many modelling and forecasting techniques available nowadays in a single environment. Once the initial barriers are overcome, a wide panorama of modelling opportunities opens up.

2.3.7. Models for population processes[14]

Over the past two centuries, formal demography has established its own, discipline-specific body of methods for predicting (or *projecting*[15]) populations. Population sciences, since their 17th century beginnings, have been traditionally very empirically focused, with strong links with probability theory (Courgeau, 2012). Given the observed regularities in population dynamics, and that populations are somewhat better predictable than many other socio-economic processes, with reasonable horizons possibly up to one generation ahead (Keyfitz, 1972, 1981), demographic forecasts have become a best-selling product of the discipline (Xie, 2000). Since the 20th

[14] This subsection was written by Jakub Bijak.

[15] The demographic literature sometimes makes a distinction between unconditional *forecasts* (or *predictions*) and *projections*, conditional on their underlying assumptions. In this section, we use the former term to refer to statements about the future, and the latter to the result of a numerical exercise of combining assumptions on fertility, mortality and migration in a deterministic model of population renewal.

century, methodological developments in human demography have been augmented by the work carried out in mathematical biology and population ecology (Caswell, 2019a).

The theoretical strength of demography also lies almost exclusively in the formal mathematical description of population processes (Burch, 2018), typically growth functions and structural changes. Historically, such attempts started from formulating the logistic model of population dynamics, inspired by the Malthusian theory (Pearl & Reed, 1920; Verhulst, 1845). Lotka's (1907) work laid the foundations of the stable population theory with asymptotic stability under constant vital rates, subsequently extended to modelling of interacting populations by using differential equations (Lotka, 1925; Volterra, 1926b). By the middle of the 20th century, the potential and limitations of demographic forecasting methods were already well recognised in the literature (Brass, 1974; Hajnal, 1955).

In the state-of-the-art demographic forecasting, the core engine is provided by matrix algebra. The most common approach relies on the cohort-component models, which combine the assumptions on fertility, mortality and migration, in order to produce future population by age, sex, and other characteristics. In such models, the deterministic mechanism of population renewal is known, and results from the following demographic accounting identity (population balancing equation, see Bryant & Zhang, 2018; Rees & Wilson, 1973):

$$P[x+1,t+1] = P[x,t] - D[(x,x+1),(t,t+1)] + I[(x,x+1),(t,t+1)] - E[(x,x+1),(t,t+1)]$$

where $P[x,t]$ denotes population aged x at time t, $D[(x,x+1),(t,t+1)]$ refer to deaths between ages x and $x+1$ in the time interval t to $t+1$, with I and E respectively denoting immigration (and other entries) and emigration (and other exits). In addition, for the youngest age group, births $B[(t,t+1)]$ need to be added. The equation above can be written up in the general algebraic form: $\mathbf{P}_{t+1} = \mathbf{G}\mathbf{P}_t$, where $\mathbf{P}_t$ is the population vector structured by age (and other characteristics), and $\mathbf{G}$ is an appropriately chosen growth matrix (Leslie matrix), closely linked with the life table while reflecting the relationship above, expressed in terms of rates rather than events (Caswell, 2019a; Leslie, 1945, 1948; Preston, Heuveline, & Guillot, 2000).

In the cohort-component approach, even though the mechanism of population change is known, the individual components still need forecasting. The three main drivers of population dynamics — fertility, mortality, and migration — differ in terms of their predictability (National Research Council, 2000): mortality, which is mainly a biological process moderated by medical technology, is the most predictable; migration, which is purely a social and behavioural process is the least; while the predictability of fertility — part-biological, part-behavioural – is in the middle (for component forecasting methods, see Sections 3.6.3–3.6.5). In practical applications, the components can be either projected deterministically, following judgment-based or expert assumptions (for example, Lutz, Butz, & Samir, 2017), or extrapolated by using probabilistic methods, either for the components or for past errors of prediction (Alho & Spencer, 1985, 2005; De Beer, 2008). An impetus to the use of stochastic methods has been given by the developments in the UN World Population Prospects (Azose, Ševčíková, & Raftery, 2016; Gerland et al., 2014). Parallel, theoretical advancements included a stochastic version of the stable population theory (Keiding & Hoem, 1976), as well as coupling of demographic uncertainty with economic models (Alho, Hougaard Jensen, & Lassila, 2008).

Since its original formulation, the cohort-component model has been subject to several extensions (see, for example, Stillwell & Clarke, 2011). The multiregional model (Rogers, 1975) describes the dynamics of multiple regional populations at the same time, with regions linked through migration. The multistate model (Schoen, 1987) generalises the multiregional analysis to any arbitrary set of states (such as educational, marital, health, or employment statuses, and so on; see also state-space models in Section 2.3.6). The multiregional model can be in turn generalised to include multiple geographic levels of analysis in a coherent way (Kupiszewski & Kupiszewska, 2011). Recent developments include multifocal analysis, with an algebraic description of kinship networks (Caswell, 2019b, 2020). For all these extensions, however, data requirements are very high: such models require detailed information on transitions between regions or states in a range of different breakdowns. For pragmatic reasons, microsimulation-based methods offer an appealing alternative, typically including large-sample Monte Carlo simulations of population trajectories based on available transition rates (Bélanger & Sabourin, 2017; Zaidi, Harding, & Williamson, 2009).

Aside of a few extensions listed above, the current methodological developments in the forecasting of human populations are mainly concentrated on the approaches for predicting individual demographic components (see Sections 3.6.3–3.6.5), rather than the description of the population renewal mechanism. Still, the continuing developments in population ecology, for example on the algebraic description of transient and asymptotic population growth (Nicol-Harper, Dooley, Packman, Mueller, Bijak, Hodgson, Townley, & Ezard, 2018), bear substantial promise of further advancements in this area, which can be additionally helped by strengthened collaboration between modellers and forecasters working across the disciplinary boundaries on the formal descriptions of the dynamics of human, as well as other populations.

2.3.8. *Forecasting count time series*[16]

Probabilistic forecasts based on predictive mass functions are the most natural way of framing predictions of a variable that enumerates the occurrences of an event over time; i.e. the most natural way of predicting a time series of *counts*. Such forecasts are both *coherent*, in the sense of being consistent with the discrete support of the variable, and capture all distributional – including tail –

[16] This subsection was written by Gael M. Martin.

information. In contrast, point forecasts based on summary measures of central location (e.g., a (conditional) mean, median or mode), convey no such distributional information, and potentially also lack coherence as, for example, when the mean forecast of the integer-valued count variable assumes non-integer values. These comments are even more pertinent for *low count* time series, in which the number of *rare* events is recorded, and for which the cardinality of the support is small. In this case, point forecasts of any sort can be misleading, and continuous (e.g., Gaussian) approximations (sometimes adopted for high count time series) are particularly inappropriate.

These points were first elucidated in Freeland and McCabe (2004), and their subsequent acceptance in the literature is evidenced by the numerous count data types for which discrete predictive distributions are now produced; including counts of: insurance claims (McCabe & Martin, 2005), medical injury deaths (Bu & McCabe, 2008), website visits (Bisaglia & Canale, 2016), disease cases (Bisaglia & Gerolimetto, 2019; Mukhopadhyay & Sathish, 2019; Rao & McCabe, 2016), banking crises (Dungey, Martin, Tang, & Tremayne, 2020), company liquidations (Homburg, Weiß, Alwan, Frahm, & Göb, 2020), hospital emergency admissions (Sun, Sun, Zhang, & McCabe, 2021), work stoppages (Weiß, Homburg, Alwan, Frahm, & Göb, 2021), and the intermittent product demand described in Section 2.8 (Berry & West, 2020; Kolassa, 2016; Snyder, Ord, & Beaumont, 2012).

The nature of the predictive model for the count variable, together with the paradigm adopted (Bayesian or frequentist), determine the form of the probabilistic forecast, including the way in which it does, or does not, accommodate parameter and model uncertainty. As highlighted in Sections 2.4.1 and 2.4.2, the Bayesian approach to forecasting is *automatically* probabilistic, no matter what the data type. It also factors parameter uncertainty into the predictive distribution, plus model uncertainty if Bayesian model averaging is adopted, producing a distribution whose location, shape and degree of dispersion reflect all such uncertainty as a consequence. See Berry and West (2020), Bisaglia and Canale (2016), Frazier, Maneesoonthorn, Martin, and McCabe (2019), Lu (2021), McCabe and Martin (2005) and Neal and Kypraios (2015), for examples of Bayesian probabilistic forecasts of counts.

In contrast, frequentist probabilistic forecasts of counts typically adopt a 'plug-in' approach, with the predictive distribution conditioned on estimates of the unknown parameters of a given count model. Sampling variation in the estimated predictive (if acknowledged) is quantified in a variety of ways. Freeland and McCabe (2004), for instance, produce confidence intervals for the true (point-wise) predictive probabilities, exploiting the asymptotic distribution of the (MLE-based) estimates of those probabilities. Bu and McCabe (2008) extend this idea to (correlated) estimates of sequential probabilities, whilst Jung and Tremayne (2006) and Weißet al. (2021) exploit bootstrap techniques to capture point-wise sampling variation in the forecast distribution. McCabe, Martin, and Harris (2011), on the other hand, use subsampling methods to capture sampling fluctuations in the *full* predictive distribution, retaining the non-negativity and summation to unity properties of the probabilities (see also Harris, Martin, Perera, & Poskitt, 2019, for related, albeit non-count data work). Model uncertainty is catered for in a variety of ways: via nonparametric (McCabe et al., 2011) or bootstrapping (Bisaglia & Gerolimetto, 2019) methods; via (frequentist) model averaging (Sun et al., 2021); or via an informal comparison of predictive results across alternative models (Jung & Tremayne, 2006). Methods designed explicitly for calibrating predictive mass functions to observed count data – whether those functions be produced using frequentist *or* Bayesian methods – can be found in Czado, Gneiting, and Held (2009) and Wei and Held (2014); see also Sections 2.12.4 and 2.12.5.

Finally, whilst full probabilistic forecasts are increasingly common, point, interval and quantile forecasts are certainly still used. The need for such summaries to be coherent with the discrete nature of the count variable appears to be now well-accepted, with recent work emphasising the importance of this property (Bu & McCabe, 2008; Homburg, Weiß, Alwan, Frahm, & Göb, 2019; Homburg et al., 2020; Mukhopadhyay & Sathish, 2019).

2.3.9. Forecasting with many variables[17]

Multivariate models – regression models with multiple explanatory variables – are often based on available theories regarding the determination of the variable to be forecast, and are often referred to as *structural models*. In a stationary world without structural change, then it would be anticipated that the best structural model would provide the best forecasts, since it would provide the conditional mean of the data process (see, for example, Clements & Hendry, 1998). In a non-stationary world of unit roots and structural breaks, however, this need not be the case. In such situations, often simple forecast models can outperform structural models, especially at short forecast horizons (see, for example, Hendry & Clements, 2001). Multivariate forecast models require that explanatory variables also be forecast – or at least, scenarios be set out for them. These may be simplistic scenarios, for example all explanatory variables take their mean values. Such scenarios can play a useful role in formulating policy making since they illustrate in some sense the outcomes of different policy choices.

Since the 1980s and Sims (1980), vector autoregressive (VAR) models have become ubiquitous in macroeconomics, and common in finance (see, for example, Hasbrouck, 1995). A VAR model is a set of linear regression equations (see also Section 2.3.2) describing the evolution of a set of endogenous variables. Each equation casts each variable as a function of lagged values of all the variables in the system. Contemporaneous values of system variables are not included in VAR models for identification purposes; some set of identifying restrictions are required, usually based on economic theory, and when imposed the resulting model is known as a structural VAR model. VAR models introduce significantly greater levels of parameterisation of relationships, which increases the level of estimation uncertainty. At the same time VAR models afford the forecaster a straightforward way to

[17] This subsection was written by J. James Reade.

generate forecasts of a range of variables, a problem when forecasting with many variables. As with autoregressive methods, VAR models can capture a significant amount of variation in data series that are autocorrelated, and hence VAR methods can be useful as baseline forecasting devices. VAR-based forecasts are often used as a benchmark for complex models in macroeconomics like DSGE models (see, for example, Del Negro & Schorfheide, 2006). The curse of dimensionality in VAR models is particularly important and has led to developments in factor-augmented VAR models, with practitioners often reducing down hundreds of variables into factors using principal component analysis (see, for example, Bernanke, Boivin, & Eliasz, 2005). Bayesian estimation is often combined with factor-augmented VAR models.

Often, significant numbers of outliers and structural breaks require many indicator variables to be used to model series (see also Sections 2.2.3 and 2.2.4). Indicator saturation is a method of detecting outliers and structural breaks by saturating a model with different types of indicator, or deterministic variables (Castle, Doornik, Hendry, & Pretis, 2015c; Johansen & Nielsen, 2009). The flexibility of the approach is such that it has been applied in a wide variety of contexts, from volcanic eruptions (Pretis, Schneider, & Smerdon, 2016) to prediction markets and social media trends (Vaughan Williams & Reade, 2016).

A particularly important and ever-expanding area of empirical analysis involves the use of panel data sets with long time dimensions: panel time series (Eberhardt, 2012). The many variables are then extended across many cross sectional units, and a central concern is the dependence between these units, be they countries, firms, or individuals. At the country level one approach to modelling this dependence has been the Global VAR approach of, for example, Dees, Mauro, Pesaran, and Smith (2007). In more general panels, the mean groups estimator has been proposed to account for cross-section dependence (Pesaran, Shin, & Smith, 1999).

Outliers, structural breaks, and split trends undoubtedly also exist in panel time series. The potential to test for common outliers and structural changes across cross sectional units would be useful, as would the ability to allow individual units to vary individually, e.g., time-varying fixed effects. Nymoen and Sparrman (2015) is the first application of indicator saturation methods in a panel context, looking at equilibrium unemployment dynamics in a panel of OECD countries, but applications into the panel context are somewhat constrained by computer software packages designed for indicator saturation (Section 3.3.3 discusses further the case of forecasting unemployment). The *gets* R package of Pretis, Reade, and Sucarrat (2017, 2018) can be used with panel data.

2.3.10. Functional time series models[18]

Functional time series consist of random functions observed at regular time intervals. Functional time series can be classified into two categories depending on if the continuum is also a time variable. On the one hand, functional time series can arise from measurements obtained by separating an almost continuous time record into consecutive intervals (e.g., days or years, see Horváth & Kokoszka, 2012). We refer to such data structure as sliced functional time series, examples of which include daily precipitation data (Gromenko, Kokoszka, & Reimherr, 2017). On the other hand, when the continuum is not a time variable, functional time series can also arise when observations over a period are considered as finite-dimensional realisations of an underlying continuous function (e.g., yearly age-specific mortality rates, see Li, Robinson, & Shang, 2020e).

Thanks to recent advances in computing storage, functional time series in the form of curves, images or shapes is common. As a result, functional time series analysis has received increasing attention. For instance, Bosq (2000) and Bosq and Blanke (2007) proposed the functional autoregressive of order 1 (FAR(1)) and derived one-step-ahead forecasts that are based on a regularised Yule–Walker equations. FAR(1) was later extended to FAR(p), under which the order p can be determined via Kokoszka and Reimherr's (2013) hypothesis testing. Horváth, Liu, Rice, and Wang (2020) compared the forecasting performance between FAR(1), FAR(p), and functional seasonal autoregressive models of Chen, Marron, and Zhang (2019b).

To overcome the curse of dimensionality (see also Sections 2.2.5, 2.5.2 and 2.5.3), a dimension reduction technique, such as functional principal component analysis (FPCA), is often used. Aue, Norinho, and Hörmann (2015) showed asymptotic equivalence between a FAR and a VAR model (for a discussion of VAR models, see Section 2.3.9). Via an FPCA, Aue et al. (2015) proposed a forecasting method based on the VAR forecasts of principal component scores. This approach can be viewed as an extension of Hyndman and Shang (2009), in which principal component scores are forecast via a univariate time series forecasting method. With the purpose of forecasting, Kargin and Onatski (2008) proposed to estimate the FAR(1) model by using the method of predictive factors. Klepsch and Klüppelberg (2017) proposed a functional moving average process and introduced an innovations algorithm to obtain the best linear predictor. Klepsch, Klüppelberg, and Wei (2017) extended the VAR model to the vector autoregressive moving average model and proposed the functional autoregressive moving average model. The functional autoregressive moving average model can be seen as an extension of autoregressive integrated moving average model in the univariate time series literature (see Section 2.3.4).

Extending short-memory to long-memory functional time series analysis, Li et al. (2020e) and Li, Robinson, and Shang (2021) considered local Whittle and rescale-range estimators in a functional autoregressive fractionally integrated moving average model. The models mentioned above require stationarity, which is often rejected. Horváth, Kokoszka, and Rice (2014) proposed a functional KPSS test for stationarity. Chang, Kim, and Park (2016) studied nonstationarity of the time series of state densities, while Beare, Seo, and Seo (2017) considered a cointegrated linear process in Hilbert space. Nielsen, Seo, and

18 This subsection was written by Han Lin Shang.

Seong (2019) proposed a variance ratio-type test to determine the dimension of the nonstationary subspace in a functional time series. Li, Robinson, and Shang (2020d) studied the estimation of the long-memory parameter in a functional fractionally integrated time series, covering the functional unit root.

From a nonparametric perspective, Besse, Cardot, and Stephenson (2000) proposed a functional kernel regression method to model temporal dependence via a similarity measure characterised by semi-metric, bandwidth and kernel function. Aneiros-Pérez and Vieu (2008) introduced a semi-functional partial linear model that combines linear and nonlinear covariates. Apart from conditional mean estimation, Hörmann, Horváth, and Reeder (2013) considered a functional autoregressive conditional heteroscedasticity model for estimating conditional variance. Rice, Wirjanto, and Zhao (2020) proposed a conditional heteroscedasticity test for functional data. Kokoszka, Rice, and Shang (2017) proposed a portmanteau test for testing autocorrelation under a functional generalised autoregressive conditional heteroscedasticity model.

2.3.11. ARCH/GARCH models[19]

Volatility has been recognised as a primary measure of risks and uncertainties (Gneiting, 2011a; Markowitz, 1952; Sharpe, 1964; Taylor, McSharry, & Buizza, 2009); for further discussion on uncertainty estimation, see Section 2.3.21. Estimating future volatility for measuring the uncertainty of forecasts is imperative for probabilistic forecasting. Yet, the right period in which to estimate future volatility has been controversial as volatility based on too long a period will make irrelevant the forecast horizon of our interests, whereas too short a period results in too much noise (Engle, 2004). An alternative to this issue is the dynamic volatility estimated through the autoregressive conditional heteroscedasticity (ARCH) proposed by Engle (1982), and the generalised autoregressive conditional heteroscedasticity (GARCH) model proposed by Bollerslev (1987). The ARCH model uses the weighted average of the past squared forecast error whereas the GARCH model generalises the ARCH model by further adopting past squared conditional volatilities. The GARCH model is the combination of (*i*) a constant volatility, which estimates the long-run average, (*ii*) the volatility forecast(s) in the last steps, and (*iii*) the new information collected in the last steps. The weightings on these components are typically estimated with maximum likelihood. The models assume a residual distribution allowing for producing density forecasts. One of the benefits of the GARCH model is that it can model heteroscedasticity, the volatility clustering characteristics of time series (Mandelbrot, 1963), a phenomenon common to many time series where uncertainties are predominant. Volatility clustering comes about as new information tends to arrive time clustered and a certain time interval is required for the time series to be stabilised as the new information is initially recognised as a shock.

The GARCH model has been extended in the diverse aspects of non-linearity, asymmetry and long memory. Among many such extensions, the Exponential GARCH (EGARCH) model by Nelson (1991) uses log transformation to prevent negative variance; the Threshold GARCH (TGARCH) model by Zakoian (1994) allows for different responses on positive and negative shocks. A small piece of information can have more impact when the time series is under stress than under a stable time series (Engle, 2004). Another pattern often observed in the volatility time series is slowly decaying autocorrelation, also known as a long memory pattern, which Baillie, Bollerslev, and Mikkelsen (1996) capture using a slow hyperbolic rate of decay for the ARCH terms in the fractionally integrated GARCH (FIGARCH) model. Separately, in a further approach to directly estimating long term volatility, the GARCH-MIDAS (Mixed Data Sampling) model proposed by Engle, Ghysels, and Sohn (2013) decomposes the conditional volatility into the short-term volatility, as captured by the traditional GARCH, and the long-term volatility represented by the realised volatilities. The Heterogeneous Autoregressive (HAR) model by Corsi (2009) considers the log-realised volatility as a linear function of the log-realised volatility of yesterday, last week and last month to reflect traders' preferences on different horizons in the past. This model is extended by Wilms, Rombouts, and Croux (2021) to incorporate information about future stock market volatility by further including option-implied volatility. A different approach to volatility modelling, discussed in Section 2.3.14, is the use of low and high prices in the range-based volatility models.

The univariate GARCH models surveyed so far have been exended to multivariate versions, in order to model changes in the conditional covariance in multiple time series, resulting in such examples as the VEC (Bollerslev, 1987) and BEKK (Engle & Kroner, 1995), an acronym derived from Baba, Engle, Kraft, and Kroner. The VEC model, a direct generalisation of the univariate GARCH, requires more parameters in the covariane matrices and provides better fitness at the expense of higher estimation costs than the BEKK. The VEC model has to ensure the positivity of the covariance matrix with further constraints, whereas the BEKK model and its specific forms, e.g., factor models, avoid this positivity issue directly at the model specification stage. In an effort to further reduce the number of parameters to be estimated, the linear and nonlinear combinations of the univariate GARCH models, such as the constant conditional correlation model of Bollerslev (1990) and the dynamic conditional correlation models of Tse and Tsui (2002) and of Engle (2002), were investigated.

2.3.12. Markov switching models[20]

Since the late 1980s, especially in macroeconomics and finance, the applications of dynamic econometric modelling and forecasting techniques have increasingly relied on a special class of models that accommodate regime shifts, Markov switching (MS) models. The idea of MS is to relate the parameters of otherwise standard dynamic

[19] This subsection was written by Jooyoung Jeon.

[20] This subsection was written by Massimo Guidolin.

econometric frameworks (such as systems of regressions, vector autoregressions, and vector error corrections) to one or more unobserved state variables (see Section 2.3.6 for a definition), say, S_t, that can take K values and capture the notion of systems going through phases or "regimes", which follow a simple, discrete stochastic process and are independent of the shocks of the model.

For instance, an otherwise standard AR(1) model can be extended to $y_t = \phi_{0,S_t} + \phi_{1,S_t} y_{(t-1)} + \sigma_{S_t}\epsilon_t$, where all the parameters in the conditional mean as well as the variance of the shocks may be assumed to take different, estimable values as a function of S_t. Similarly, in a K-regime MS VAR(p), the vector of intercepts and the p autoregressive matrices may be assumed to depend on S_t. Moreover, the covariance matrix of the system shocks may be assumed to depend on some state variable, either the same as the mean parameters (S_t) or an additional, specific one (V_t), which may depend on lags of S_t. When a MS VAR model is extended to include exogenous regressors, we face a MS VARX, of which MS regressions are just a special case.

Even though multivariate MS models may suffer from issues of over-parameterisations that must be kept in check, their power of fitting complex non-linearities is unquestioned because, as discussed by Marron and Wand (1992), mixtures of normal distributions provide a flexible family that can be used to approximate many distributions. Moreover, MS models are known (Timmermann, 2000) to capture key features of many time series. For instance, differences in conditional means across regimes enter the higher moments such as variance, skewness, and kurtosis; differences in means in addition to differences in variances across regimes may generate persistence in levels and squared values of the series.

The mainstream literature (see, e.g., Hamilton 1990, or the textbook treatments by Kim, Charles, and Nelson 1999, and Guidolin and Pedio 2018) initially focused on time-homogeneous Markov chains (where the probabilities of the state transitions are constant). However, the finance and business cycles literatures (Gray, 1996) has moved towards time-heterogeneous MS models, in which the transition matrix of the regimes may change over time, reacting to lagged values of the endogenous variables, to lagged exogenous variables, or to the lagged values of the state (in a self-exciting fashion).

MS models may be estimated by maximum likelihood, although other estimation methods cannot be ruled out, like GMM (Lux, 2008). Typically, estimation and inference are based on the Expectation-Maximisation algorithm proposed by Dempster, Laird, and Rubin (1977), a filter that allows the iterative calculation of the one-step ahead forecast of the state vector given the information set and a simplified construction of the log-likelihood of the data. However, there is significant evidence of considerable advantages offered by Bayesian approaches based on Monte Carlo Markov chain techniques to estimating multivariate MS models (see, for example, Hahn, Frühwirth-Schnatter, & Sass, 2010).

Notably, MS models have been recently generalised in a number of directions, such as including regimes in conditional variance functions, for example of a GARCH or DCC type (see Pelletier, 2006, and Section 2.3.11).

2.3.13. Threshold models[21]

It is a well-known fact that financial and economic time series often display non-linear patterns, such as structural instability, which may appear in the form of recurrent regimes in model parameters. In the latter case, such instability is stochastic, it displays structure, and as such, it can be predicted. Accordingly, modelling economic and financial instability has become an essential goal for econometricians since the 1970s.

One of the first and most popular models is the threshold autoregressive (TAR) model developed by Tong (1978). A TAR model is an autoregressive model for the time series y_t in which the parameters are driven by a state variable S_t (see Section 2.3.6 for a definition), which is itself a random variable taking K distinct integer values (i.e., $S_t = k$, $k = 1, \ldots, K$). In turn, the value assumed by S_t depends on the value of the threshold variable q_t when compared to $K-1$ threshold levels, q_k^*. For instance, if only two regimes exists, it is $S_t = 1$ if $q_t \leq q_1^*$ and $S_t = 2$ otherwise. The threshold variable q_t can be exogenous or can be a lagged value of y_t. In the latter case, we speak of self-exciting threshold autoregressive (SETAR) models. Other choices of q_t include linear (Chen, Chiang, & So, 2003; Chen & So, 2006; Gerlach, Chen, Lin, & Huang, 2006) or non-linear (Chen, 1995; Wu & Chen, 2007) combinations of the lagged dependent variable or of exogenous variables.

The TAR model has also been extended to account for different specifications of the conditional mean function leading to the development of the threshold moving average (TMA – see, for example, De Gooijer, 1998; Ling, Tong, & Li, 2007; Tong, 1990) and the threshold autoregressive moving average (TARMA – see, for example, Amendola, Niglio, & Vitale, 2006; Ling, 1999) models. Those models are similar to the ones described in Section 2.3.4, but their parameters depend on the regime K.

A criticism of TAR models is that they imply a conditional moment function that fails to be continuous. To address this issue, Chan and Tong (1986) proposed the smooth transition autoregressive (STAR) model. The main difference between TAR and STAR models is that, while a TAR imposes an abrupt shift from one regime to the others at any time that the threshold variable crosses above (or below) a certain level, a STAR model allows for gradual changes among regimes.

In its simplest form, a STAR is a two-regime model where the dependent variable y_t is determined as the weighted average of two autoregressive (AR) models, i.e.,

$$y_t = \sum +j = 1^p \phi_{j,1} y_{t-j} P(S_t = 1; g(x_t)) + \sum_{j=1}^{p} \phi_{j,2} y_{t-j} P(S_t = 2; g(x_t)) + \epsilon_t,$$

where x_t is the transition variable and g is some transformation of the transition variable x_t. Regime probabilities are assigned through the transition function $F(k; g(x_t))$, with F being a cumulative density function of choice. The transition variable x_t can be the lagged

[21] This subsection was written by Manuela Pedio.

endogenous variable, y_{t-d} for $d \geq 1$ (Teräsvirta, 1994), a (possibly non-linear) function of it, or an exogenous variable. The transition variable can also be a linear time trend ($x_t = t$), which generates a model with smoothly changing parameters (Lin & Teräsvirta, 1994). Popular choices for the transition function F are the logistic function (which gives rise to the LSTAR model) and the exponential function (ESTAR). Notably, the simple STAR model we have described can be generalised to have multiple regimes (Van Dijk, Franses, & Lucas, 1999).

Threshold models are also applied to modelling and forecasting volatility; for instance, the GJR-GARCH model of Glosten, Jagannathan, and Runkle (1993) can be interpreted as a special case of a threshold model. A few multivariate extensions of threshold models also exist, such as vector autoregressive threshold models, threshold error correction models (Balke & Fomby, 1997), and smooth transition vector error correction models (Granger & Swanson, 1996).

2.3.14. Low and high prices in volatility models[22]

Volatility models of financial instruments are largely based solely on closing prices (see Section 2.3.11); meanwhile, daily low and high (LH) prices significantly increase the amount of information about the variation of returns during a day. LH prices were used for the construction of highly efficient estimators of variance, so called the range-based (RB) estimators (e.g., Fiszeder & Perczak, 2013; Garman & Klass, 1980; Magdon-Ismail & Atiya, 2003; Parkinson, 1980; Rogers & Satchell, 1991; Yang & Zhang, 2000). Recently, Riedel (2021) analysed how much additional information about LH reduces the time averaged variance in comparison to knowing only open and close. RB variance estimators, however, have a fundamental drawback, as they neglect the temporal dependence of returns (like conditional heteroscedasticity) and do not allow for the calculation of multi-period dynamic volatility forecasts.

In the last dozen or so years, numerous univariate dynamic volatility models have been constructed based on LH prices. Some of them were presented in the review paper of Chou, Chou, and Liu (2015). These models can be divided into four groups. The first one comprises simple models, used traditionally to describe returns, but they are based on the price range or on the mentioned earlier RB variance estimators. They include such models as random walk, moving average, exponentially weighted moving average (EWMA), autoregressive (AR), autoregressive moving average (ARMA; see Section 2.3.4), and heterogeneous autoregressive (HAR). The second group contains models which describe the conditional variance (or standard deviation) of returns. It comprises models like GARCH-PARK-R (Mapa, 2003), GARCH-TR (Fiszeder, 2005), REGARCH (Brandt & Jones, 2006), RGARCH (Molnár, 2016). The third group includes models which describe the conditional mean of the price range. It means that in order to forecast variance of returns the results have to be scaled. This group contains models like RB SV (Alizadeh, Brandt, & Diebold, 2002), CARR (Chou, 2005), TARR (Chen, Gerlach, & Lin, 2008), CARGPR (Chan, Lam, Yu, Choy, & Chen, 2012), STARR (Lin, Chen, & Gerlach, 2012), and MSRB (Miao, Wu, & Su, 2013). The last group is methodologically different because the estimation of model parameters is based on the sets of three prices, i.e., low, high and closing. This approach comprises the GARCH models (Fiszeder & Perczak, 2016; Lildholdt, 2002; Venter, De Jongh, & Griebenow, 2005) and the SV model (Horst, Rodriguez, Gzyl, & Molina, 2012).

The development of multivariate models with LH prices has taken place in the last few years. They can be divided into three groups. The first one includes models, used traditionally to describe returns or prices, but they are based on the price range or RB variance estimators. They comprise such models like multivariate EWMA, VAR, HVAR, and vector error correction (VEC). It is a simple approach, however most models omit modelling the covariance of returns. The second group is formed by the multivariate RB volatility models like RB-DCC (Chou, Wu, & Liu, 2009), DSTCC-CARR (Chou & Cai, 2009), RR-HGADCC (Asai, 2013), RB-MS-DCC (Su & Wu, 2014), DCC-RGARCH (Fiszeder, Fałdziński, & Molnár, 2019), RB-copula (Chiang & Wang, 2011; Wu & Liang, 2011). The third group includes the multivariate co-range volatility models like multivariate CARR (Fernandes, de Sá Mota, & Rocha, 2005), BEKK-HL (Fiszeder, 2018) and co-range DCC (Fiszeder & Fałdziński, 2019). These models apply LH prices directly not only for the construction of variances of returns but also for covariances. Section 3.3.9 discusses the use of the range-based volatility models in financial time series forecasting.

2.3.15. Forecasting with DSGE models[23]

Dynamic Stochastic General Equilibrium (DSGE) models are the workhorse of modern macroeconomics employed by monetary and fiscal authorities to explain and forecast comovements of aggregate time series over the business cycle and to perform quantitative policy analysis. These models are studied in both academia and policy-making institutions (for details, see: Christiano, Eichenbaum, & Trabandt, 2018; Del Negro & Schorfheide, 2013; Paccagnini, 2017). For example, the European Central Bank uses the New Area-Wide Model introduced by Warne, Coenen, and Christoffel (2010) and the Federal Reserve Board has created the Estimated, Dynamic, Optimisation-based model (FRB/EDO) as discussed in Chung, Kiley, and Laforte (2010). For an application on forecasting GDP and inflation, see Section 3.3.2. Developed as a response to Lucas (1976) critique of structural macroeconometrics models, DSGEs introduced microfundations to describe business cycle fluctuations. Initially calibrated, estimated DSGEs have been employed in shocks identification and forecasting horseraces for the last 15 years. Estimation became possible thanks to computational progress and adoption of Bayesian techniques (for technical details, see: An & Schorfheide, 2007; Fernández-Villaverde & Guerrón-Quintana, 2020; Herbst & Schorfheide, 2016). Bayesian estimation allows for attributing prior distributions, instead of calibrating, and

[22] This subsection was written by Piotr Fiszeder.

[23] This subsection was written by Alessia Paccagnini.

computing the posterior distribution for selected model parameters as well as drawing from predictive density. The Smets and Wouters (2007) DSGE is the most popular framework referred to in both research and policy literature. Proposed for the US economy, this medium-scale model is a closed economy composed of households, labor unions, a productive sector, and a monetary policy authority that sets the short-term interest rate according to a Taylor rule. These ingredients are mathematically represented by a system of linear rational expectation equations. Using a solution algorithm (for example, Blanchard & Kahn, 1980; Sims, 2002), researchers can write the model using the state-space representation composed by the transition equation and the measurement equation. The latter matches the observed data (in the Smets and Wouters: output growth rate, consumption, investment, wages, worked hours, inflation, and short-term interest rate) with the model latent variables. The solved model is employed for quantitative policy analysis and to predict and explain the behavior of macroeconomic and financial indicators.

DSGE models forecasting performance is investigated along two dimensions: point forecast and density forecast (see Sections 2.12.2 and 2.12.4 for discussions on their evaluation).

The point forecast is implemented by conducting both static and dynamic analysis, as described in Cardani, Paccagnini, and Villa (2019). If the static analysis provides a unique forecast value, the dynamic analysis describes the evolution of the prediction along the time dimension to investigate possible time-varying effects. Usually, point predictions are compared using the Diebold and Mariano (1995) and the Clark and West (2006) tests that compare predictions from two competing models The accuracy of the static analysis is based mainly on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). MAE and RMSE are used to provide a relative forecasting evaluation compared to other competitors. Following Clements and Hendry (1998), Kolasa, Rubaszek, and Skrzypczyński (2012) apply the standard forecast unbiased test to assess if DSGEs are good forecasters in the absolute sense. The accuracy of the dynamic analysis is based on the Fluctuation Test (for some DSGE applications, see: Boneva, Fawcett, Masolo, & Waldron, 2019; Cardani et al., 2019; Giacomini & Rossi, 2016). This test is based on the calculation of RMSEs that are assessed to investigate if the forecasting performance can be influenced by instabilities in the model parameters.

The density forecast is based on the uncertainty derived by the Bayesian estimation and it is commonly evaluated using the probability integral transform and the log predictive density scores (as main references, Kolasa & Rubaszek, 2015a; Wolters, 2015). The statistical significance of these predictions is evaluated using the Amisano and Giacomini (2007) test that compares log predictive density scores from two competing models.

2.3.16. Robust equilibrium-correction forecasting devices[24]

The use of equilibrium-correction models is ubiquitous in forecasting. Hendry (2010) notes that this class commonly includes models with explicit equilibrium-correction mechanisms such as vector equilibrium-correction models (VEqCM) as well as models with implicit equilibrium-correction (or long-run mean reversion) mechanisms such as vector auto-regressions (VARs; see Section 2.3.9), dynamic factor models (DFMs), dynamic stochastic general-equilibrium (DSGE) models (see Section 2.3.15), most models of the variance (see Section 2.3.11), and almost all regression equations (see Sections 2.3.2 and 2.3.4). This class of forecast model is prevalent across most disciplines. For example, Pretis (2020) illustrates that there is an equivalence between physical energy balance models, which are used to explain and predict the evolution of climate systems, and VEqCMs.

Despite their wide-spread use in economic modeling and forecasting, equilibrium-correction models often produce forecasts that exhibit large and systematic forecast errors. Clements and Hendry (1998, 1999) showed that forecasts from equilibrium-correction models are not robust to abrupt changes in the equilibrium. These types of regime changes are very common in macroeconomic time series (see Hamilton, 2016, as well as Section 2.3.12) and can cause the forecasts from many models to go off track. Therefore, if for example, there is a change in the equilibrium towards the end of the estimation sample, forecasts from this class of models will continue to converge back to the previous equilibrium.

In general, the forecasts from equilibrium-correction models can be robustified by estimating all model parameters over smaller or more flexible sub-samples. Several studies have proposed general procedures that allow for time-variation in the parameters; see, for example, Giraitis, Kapetanios, and Price (2013), Inoue, Jin, and Rossi (2017) and Pesaran, Pick, and Pranovich (2013). This allows for an optimal or more adaptive selection of model estimation windows in order generate forecasts after structural breaks have occurred.

An alternative approach for robustifying forecasts from equilibrium-correction models is to focus on the formulation and estimation of the equilibrium. Hendry (2006) shows that differencing the equilibrium-correction mechanism can improve the forecasts by removing aspects which are susceptible to shifts. However, differencing the equilibrium also induces additional forecast-error variance. Castle, Fawcett, and Hendry (2010) show that it is beneficial to update the equilibrium or to incorporate the underlying structural break process. Alternatively, Castle, Clements, and Hendry (2015a) show that there can be large forecast gains from smoothing over estimates of the transformed equilibrium. Building on this, Martinez, Castle, and Hendry (2021) show that there are many possible transformations of the equilibrium that can improve the forecasts. Several of these transformations imply that the equilibrium-correction model collapses to different naive forecast devices whose forecasts are often difficult to beat. By replacing the equilibrium with smooth estimates of these transformations, it is possible to outperform the naive forecasts at both short and long forecast horizons while retaining the underlying economic theory embedded within the equilibrium-correction model.

[24] This subsection was written by Andrew B. Martinez.

Thus, it is possible to dramatically improve forecasts from equilibrium-correction models using targeted transformations of the estimated equilibrium so that it is less susceptible to the shifts which are so damaging to the model forecasts.

2.3.17. Forecasting with data subject to revision[25]

When a forecast is made today of the future value of a variable, the forecast is necessarily 'real time' – only information available at the time the forecast is made can be used. The forecasting ability of a model can be evaluated by mimicking this setup – generating forecasts over some past period (so outcomes known) only using data known at each forecast origin. As noted by Clements and Hendry (2005), out-of-sample forecast performance is the gold standard. Sometimes the analysis is pseudo real time. At a given forecast origin t, forecasts are constructed only using data up to the period t, but the data are taken from the latest-available vintage at the time the study is undertaken. Using revised data to estimate the forecasting model – instead of the data available at the time the forecast was made – may exaggerate forecast performance, and present a misleading picture of how well the model might perform in real time. The improved availability of real-time databases has facilitated proper real-time studies.[26] At time t the data are taken from the vintage available at time t. Data revisions are often important, and occur because statistical agencies strive to provide timely estimates which are based on incomplete source data (see, for example, Fixler & Grimm, 2005, 2008; Zwijnenburg, 2015).

There are a number of possible real-time approaches. The conventional approach is to estimate the forecasting model using the latest vintage of data available at time t. Suppose the vintage-t contains data for time periods up to $t-1$, denoted $\ldots, y^t_{t-3}, y^t_{t-2}, y^t_{t-1}$. The observation for time $t-1$ is a first estimate, for $t-2$ a second estimate, and so on, such that data for earlier periods will have been revised many times. Hence the model will be estimated on data of different maturities, much of which will have been revised many times. But the forecast will typically be generated by feeding into the model 'lightly-revised' data for the most recent time periods. The accuracy of the resulting forecasts can be improved upon (in principle) by taking into account data revisions (see, for example, Clements & Galvão, 2013b; Kishor & Koenig, 2012; Koenig, Dolmas, & Piger, 2003). In the following two paragraphs, we consider alternative real-time approaches which solve the problem of estimating the model on mainly revised data, and feeding in mainly unrevised forecast origin data.

Koenig et al. (2003) suggest using real-time-vintage (RTV) data to estimate the model. The idea is to use early estimates of the data to estimate the model, so that the model is estimated on 'lightly-revised' data that matches the maturity of the forecast-origin data that the forecast is conditioned on.

Other approaches seek to model the data revisions process along with the fully-revised true values of the data, as in Cunningham, Eklund, Jeffery, Kapetanios, and Labhard (2009), Jacobs and van Norden (2011) and Kishor and Koenig (2012). Reduced form models that avoid the necessity of estimating unobserved components have adapted the vector autoregression (VAR; see also Section 2.3.9) of Sims (1980) to jointly model different observed vintages of data. Following Patterson (1995), Garratt, Lee, Mise, and Shields (2008) work in terms of the level of the log of the variable, Y^{t+1}_t, and model the vector given by $\mathbf{Z}^{t+1} = \left(Y^{t+1}_t - Y^t_{t-1}, Y^{t+1}_{t-1} - Y^t_{t-1}, Y^{t+1}_{t-2} - Y^t_{t-2}\right)'$. Carriero, Clements, and Galvão (2015) and Clements and Galvão (2012, 2013a) minimise the effects of benchmark revisions and re-basing by modelling 'same-vintage-growth rates', namely $\mathbf{Z}^{t+1} = \left(y^{t+1}_t, y^{t+1}_{t-1}, \ldots, y^{t+1}_{t-q+1}\right)'$, where $y^{t+1}_t = Y^{t+1}_t - Y^{t+1}_{t-1}$, and q denotes the greatest data maturity.

Galvão (2017) shows how forecasts of fully-revised data can be generated for dynamic stochastic general equilibrium (DSGE; Section 2.3.15) models (for example, Del Negro & Schorfheide, 2013), by applying the approach of Kishor and Koenig (2012). Clements (2017) argues that improvements in forecast accuracy might be expected to be greater for interval or density forecasts than point forecasts, and this is further explored by Clements and Galvão (2017).

Surveys on data revisions and real-time analysis, including forecasting, are provided by Clements and Galvão (2019) and Croushore (2006, 2011a, 2011b); see also Section 3.3.1.

2.3.18. Innovation diffusion models[27]

Forecasting the diffusion of innovations is a broad field of research, and influential reviews on the topic have highlighted its importance in many disciplines for strategic or anticipative reasons (Mahajan, Muller, & Bass, 1990; Meade & Islam, 2006; Peres, Muller, & Mahajan, 2010). Large-scale and fast diffusion processes of different nature, ranging from the spread of new epidemics to the adoption of new technologies and products, from the fast diffusion of news to the wide acceptance of new trends and social norms, are demanding a strong effort in terms of forecasting and control, in order to manage their impact into socio-economic, technological and ecological systems.

The formal representation of diffusion processes is often based on epidemic models, under the hypothesis that an innovation spreads in a social system through communication among people just like an epidemics does through contagion. The simplest example is represented by the (cumulative) logistic equation that describes a pure epidemic process in a homogeneously mixing population (Verhulst, 1838). The most famous and employed evolution of the logistic equation is the Bass model (Bass, 1969), developed in the field of quantitative marketing and soon become a major reference, due to its simple and powerful structure.

[25] This subsection was written by Michael P. Clements.

[26] For example, the Federal Reserve Bank of Philadelphia maintain a real-time data set covering a number of US macro-variables, at: https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/, and see Croushore and Stark (2001).

[27] This subsection was written by Mariangela Guidolin.

The Bass model (BM) describes the life-cycle of an innovation, depicting its characterising phases of launch, growth/maturity, and decline, as result of the purchase decisions of a given cohort of potential adopters. Mathematically, the model is represented by a first order differential equation, describing a diffusion process by means of three parameters: the maximum market potential, m, assumed to be constant along the whole diffusion process, and parameters p and q, referring respectively to two distinct categories of consumers, the *innovators*, identified with parameter p, adopting for first, and the *imitators*, adopting at a later stage by imitating others' behaviour and thus responsible for *word-of-mouth* dynamics. In strategic terms, crucial forecasts are referred to the point of maximum growth of the life cycle, the *peak*, and the point of market saturation. For a general description of new product forecasting please refer to Section 3.2.6.

Innovation diffusion models may be also used for *post-hoc* explanations, helping understand the evolution of a specific market and its response to different external factors. Indeed, one of the most appealing characteristics of this class of models is the possibility to give a simple and nice interpretation to all the parameters involved. In this perspective, a valuable generalisation of the BM was proposed in Bass et al. (1994) with the Generalised Bass Model (GBM). The GBM enlarges the BM by multiplying its hazard rate by a very general intervention function $x(t)$, assumed to be non-negative, which may account for exogenous shocks able to change the temporal dynamics of the diffusion process, like marketing strategies, incentive mechanisms, change in prices and policy measures.

Another generalisation of the BM and the GBM, relaxing the assumption of a constant market potential was proposed in Guseo and Guidolin (2009) with the GGM. This model postulates a time-dependent market potential, $m(t)$, which is function of the spread of knowledge about the innovation, and thus assumes that a diffusion process is characterised by two separate phases, information and adoption. The GGM allows a significant improvement in forecasting over the simpler BM, especially through a more efficient description of the first part of the time series, often characterised by a slowdown pattern, as noticed by Guseo and Guidolin (2011).

Other generalisations of innovation diffusion models, considering competition between products, are treated in Section 2.3.20. Applications of innovation diffusion models are presented in Sections 3.2.6 and 3.4.5.

2.3.19. The natural law of growth in competition[28]

As early as in 1925 Alfred J. Lotka demonstrated that manmade products diffuse in society along S-shaped patterns similar to those of the populations of biological organisms (Lotka, 1925). Since then S curve logistic descriptions have made their appearance in a wide range of applications from biology, epidemiology and ecology to industry, competitive substitutions, art, personal achievement and others (Fisher & Pry, 1971; Marchetti, 1983; Meade, 1984; Modis, 1992). The reader is also referred to Sections 2.3.18 and 3.4.5. In fact, logistic growth can be detected whenever there is growth in competition, and competition can be generalised to a high level of abstraction, e.g. diseases competing for victims and all possible accidents competing for the chance to be materialised.

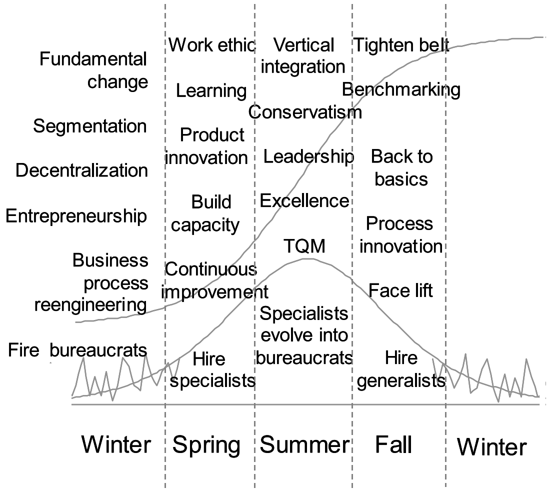

Fig. 1. Typical attributes of a growth cycle's "seasons".
Source: Adopted from Modis (1998) with the permission from the author.

S-curves enter as modular components in many intricate natural patterns. One may find S curves inside S curves because logistics portray a fractal aspect; a large S curve can be decomposed in a cascade of smaller ones (Modis, 1994). One may also find chaos by rendering the logistic equation discrete (Modis & Debecker, 1992). Finally, logistics sit in the heart of the Lotka–Volterra equations, which describe the predator–prey relations and other forms of competition. In its full generality, the logistic equation, in a discrete form, with cross terms to account for all interrelations between competing species, would give a complete picture in which growth in competition, chaos, self-organisation, complex adaptive systems, autopoiesis, and other such academic formulations, all ensue as special cases (Modis, 1997).

Each S curve has its own life cycle, undergoing good and bad "seasons" (see Fig. 1). A large set of behaviours have been tabulated, each one best suited for a particular season (Modis, 1998). Becoming conservative – seeking no change – is appropriate in the summer when things work well. But excellence drops in second place during the difficult times of winter – characterised by chaotic fluctuations – when fundamental change must take place. Learning and investing are appropriate for spring, but teaching, tightening the belt, and sowing the seeds for the next season's crop belong in the fall.

Focusing on *what* to do is appropriate in spring, whereas in fall the emphasis shifts to the *how*. For example, the evolution of classical music followed a large-timeframe S curve beginning in the fifteenth century and reaching a ceiling in the twentieth century; see Fig. 2 (Modis, 2013). In Bach's time composers were concerned

[28] This subsection was written by Theodore Modis.

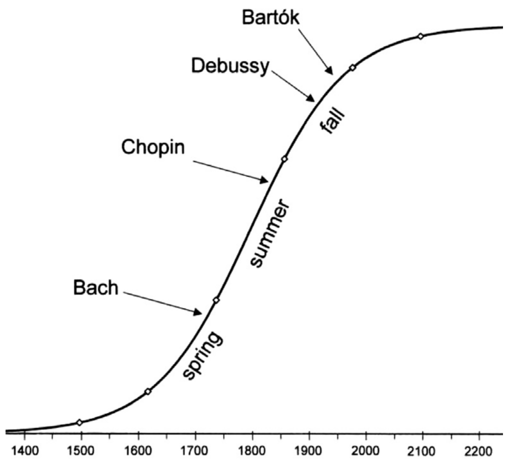

Fig. 2. The evolution of classical music. The vertical axis could be something like "importance", "public appreciation", or "public preoccupation with music" (always cumulative).
Source: Adopted from Modis (2013) with the permission from the author.

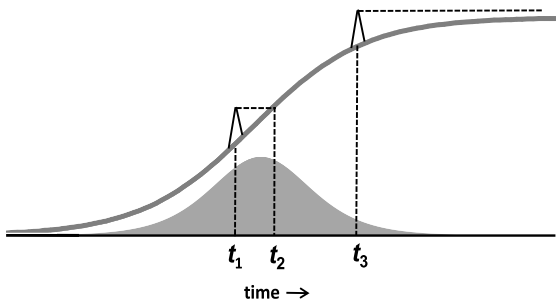

Fig. 3. An upward excursion at t_1 reaches the same level as the logistic curve at t_2 and can be considered as a "natural" deviation. The same-size excursion at time t_3 has no corresponding point on the curve. The grey life cycle delimits the position and size of all "natural" deviations.
Source: Adapted from Modis (2007) with permission from the author.

with *what* to say. The value of their music is in its architecture and as a consequence it can be interpreted by any instrument, even by simple whistling. But two hundred years later composers such as Debussy wrote music that depends crucially on the interpretation, the *how*. Classical music was still "young" in Bach's time but was getting "old" by Debussy's time. No wonder Chopin is more popular than Bartók. Chopin composed during the "summer" of music's S curve when public preoccupation with music grew fastest. Around that time composers were rewarded more handsomely than today. The innovations they made in music – excursions above the curve – were assimilated by the public within a short period of time because the curve rose steeply and would rapidly catch up with each excursion/innovation. But today the curve has flattened and composers are given limited space. If they make an innovation and find themselves above the curve, there won't be any time in the future when the public will appreciate their work; see Fig. 3 (Modis, 2007). On the other hand, if they don't innovate, they will not be saying anything new. In either case today's composers will not be credited with an achievement.

S curves constructed only qualitatively can be accurate, informative, and insightful. Practical challenges of applying S curves are discussed in Section 3.8.12.

2.3.20. Synchronic and diachronic competition[29]

Synchronic and diachronic competition models account for critical life cycle perturbations due to interactions, not captured by univariate innovation diffusion models (see Section 2.3.18) or other time series models, such as ARIMA and VAR. This is important especially in medium and long-term prediction.

Competition in a natural or socio-economic system generally refers to the presence of antagonists that contend for the same resource. This typically occurs in nature, where multiple species struggle with each other to survive, or in socio-economic contexts where products, technologies and ideas concur to be finally accepted within a market and compete for the same market potential. These competition dynamics are reflected in separate time series – one for each concurrent – characterised by a locally oscillatory behaviour with nonlinear trends, unexpected deviations and saturating effects.

The analytic representation of competition has followed different approaches. A first approach has been based on complex systems analysis (Boccara, 2004), which refers to a class of agents (see also Section 2.7.3) that interact, through local transition rules, and produce competition as an emergent behaviour. This approach may be frequently reduced to a system of differential equations, with suitable mean field approximations. A second approach, systems analysis, has been based on systems of ordinary differential equations (ODE).

In this domain, competition may be a synchronic process, if competitors are present in the environment at the same time; for example, two products may enter the market in the same period. Instead, it is diachronic if competitors come at a later stage; for example, one product enters a market in a given period and just subsequently other similar products enter the same market and start to compete. Pioneering contributions of this competition modelling are due to Lotka (1920) and Volterra (1926a), who independently obtained a class of synchronic predator–prey models; see also Section 2.3.19. A generalised version of the Lotka–Volterra (LV) model has been provided by Abramson and Zanette (1998).

Morris and Pratt (2003) proposed an extended LV model for a duopolistic situation, by making explicit the role of carrying capacities or market potentials, and the inhibiting strength of the competitors in accessing the residual resource. LV models typically do not have closed form solutions. In such cases, a staked form of equations allows a first-stage inference based on nonlinear least squares (NLS), with no strong assumptions on the stochastic distributions of error component. Short-term refining may be grounded on a Seasonal Autoregressive Moving Average with exogenous input (SARMAX) representation. Outstanding forecasts may be obtained

[29] This subsection was written by Renato Guseo.

including the estimated first-stage solution as 'exogenous' input (see Sections 2.2.3 and 2.2.5).

A different synchronic model, termed Givon-Bass (GB) model, extending the univariate innovation diffusion models described in Section 2.3.18 (Bass, 1969; Givon, Mahajan, & Müller, 1995), has been presented in Bonaldo (1991), introducing parametric components of global interaction. In this model, the residual market (or carrying capacity) is completely accessible to all competitors, and the rate equations introduce distributed seeding effects. The GB model has a closed form solution that was independently published by Krishnan, Bass, and Kummar (2000). The more general model by Savin and Terwiesch (2005) and related advances by Guseo and Mortarino (2010) were extended to the diachronic case in Guseo and Mortarino (2012), defining a competition and regime change diachronic (CRCD) model with a closed form solution. A relevant improvement of CRCD has been proposed in Guseo and Mortarino (2014), by introducing within-brand and cross-brand word-of-mouth effects, not present in standard LV models. The unrestricted unbalanced diachronic competition (unrestricted UCRCD) model, is defined with these new factors. The model assumes, among other specifications, a constant market potential. In Guseo and Mortarino (2015) this assumption is relaxed, by introducing a dynamic market potential (Guseo & Guidolin, 2009, 2011). Some applications are summarised in Section 3.8.8.

2.3.21. Estimation and representation of uncertainty[30]

Forecasting uncertainty consists in estimating the possible range of forecast errors (or true values) in the future and the most widely adopted representation is a forecast interval (Patel, 1989). The forecast interval indicates a range of values and the respective probability, which is likely to contain the true value (which is yet to be observed) of the response variable. Since for a specific lead-time the forecast interval only encompasses information from a marginal distribution, it can also be named marginal forecast interval (MFI). A MFI can be obtained from: parametric distribution, such as a Gaussian distribution with conditional mean and variance estimated with a Generalised ARCH model (Baillie & Bollerslev, 1992); non-parametric distribution, e.g., obtained with conditional kernel density estimation (Hyndman, Bashtannyk, & Grunwald, 1996); directly estimated with statistical learning methods, such as quantile regression (Taylor & Bunn, 1999) or bootstrapping (Masarotto, 1990), or with machine learning algorithms like quantile random forests (Meinshausen, 2006). For a combination of density forecasts from different models, see Section 2.6.2.

For multi-step ahead forecasting problems (see also Section 2.7.7), in particular when information about forecast uncertainty is integrated in multi-period stochastic optimisation (Dantzig & Infanger, 1993), information about the temporal dependency structure of forecast errors (or uncertainty) is a fundamental requirement. In this case, the concept of simultaneous forecast intervals (SFI) can be found in the statistical literature (Chew, 1968). SFI differ from MFI since take into account the temporal interdependency of forecast errors and are constructed to have the observed temporal trajectory of the response variable fully contained inside the forecast intervals during all lead-times of the time horizon. The number of works that cover SFI is lower when compared to the MFI, but some examples are: methods based on Bonferroni- and product-type inequalities applied time series forecasting models likes ARIMA and Holt-Winters (Ravishanker, yen Wu, & Glaz, 1991); combination of bootstrap replications and an heuristic optimisation procedure to find an envelope of the temporal trajectories closest to the deterministic forecast (Staszewska-Bystrova, 2011, see also Section 2.7.5); sampling forecast errors at different horizons and estimate the SFI with the empirical Mahalanobis distance (Jordá, Knüppelc, & Marcellino, 2013).

Advances in Operations Research for decision-making problems under uncertainty imposed new requirements in forecast uncertainty estimation and representation. On one hand, stochastic optimisation requires a scenario representation for forecast uncertainty (Powell, 2019). This motivated research in methods that generate uncertainty forecasts represented by random vectors (term used in statistics) or path forecasts (term used in econometrics), such as parametric copula combined with MFI (Pinson, Madsen, Nielsen, Papaefthymiou, & Klöckl, 2009), parametric dynamic stochastic model (Li & Chan, 2011) or epi-spline basis functions (Rios, Wets, & Woodruff, 2015). On the other hand, robust optimisation does not make specific assumptions on probability distributions and the uncertain parameters are assumed to belong to a deterministic uncertainty set. Hence, some authors proposed new methods to shape forecast uncertainty as polyhedral or ellipsoidal regions to enable a direct integration of forecasts in this type of optimisation problem (Bertsimas & Pachamanova, 2008; Golestaneh, Pinson, & Gooi, 2019).

Finally, communication of forecast uncertainty (e.g., MFI, SFI, random vectors) to decision-makers requires further attention since it remains as a major bottleneck for a wide adoption by industry, particularly in uses cases with multivariate time series (Akram, Binning, & Maih, 2015) and adverse weather events (Ramos, Mathevet, Thielen, & Pappenberger, 2010). Please also see Section 3.7.5.

2.3.22. Forecasting under fat tails[31]

A non-negative continuous random variable X is fat-tailed, if its survival function $S(x) = P(X \geq x)$ is regularly varying, that is to say if $S(x) = L(x)x^{-\alpha}$, where $L(x)$ is a slowly varying function, for which $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ for $t > 0$ (Embrechts, Klüppelberg, & Mikosch, 2013). The parameter α is known as the tail parameter, and it governs the thickness of the tail – the smaller α the fatter the tail – and the existence of moments, so that $E[X^p] < \infty$ if and only if $\alpha > p$. Often α is re-parametrised as $\xi = 1/\alpha$.

[30] This subsection was written by Ricardo Bessa.

[31] This subsection was written by Pasquale Cirillo.

Fat tails are omnipresent in nature, from earthquakes to floods, and they are particularly common in human-related phenomena like financial markets, insurance, pandemics, and wars (see, for example, Mandelbrot, 1983; Taleb, 2020, and references therein).

Forecasting fat-tailed random variables is therefore pivotal in many fields of life and science. However, while a basic coverage of the topic is available in most time series and risk management manuals (e.g., McNeil, Frey, & Embrechts, 2015; Shumway & Stoffer, 2017), the profound implications of fat tails are rarely taken into consideration, and this can generate substantial errors in forecasts.

As observed in Taleb, Bar-Yam, and Cirillo (2020), any statistical forecasting activity about the mean – or another quantity – of a phenomenon needs the law of large numbers (LLN), which guarantees the convergence of the sample mean at a given known rate, when the number of observations n grows.

Fat-tailed phenomena with tail parameter $\alpha \leq 1$ are trivially not predictable. Since their theoretical mean is not defined, the LLN does not work, for there is nothing the sample mean can converge to. This also applies to apparently infinite-mean phenomena, like pandemics and wars, i.e., extremely tricky objects of study, as discussed in Cirillo and Taleb (2016a). In similar situations, one can rely on extreme value theory to understand tail risk properties, but should refrain from standard forecasting.

For random variables with $1 < \alpha \leq 2$, the LLN can be extremely slow, and an often unavailable number of observations is needed to produce reliable forecasts. Even for a well-behaved and non-erratic phenomenon, we all agree that a claim about the fitness or non-fitness of a forecasting approach, just on the basis of one single observation ($n = 1$), would be considered unscientific. The fact is that with fat-tailed variables that "$n = 1$" problem can be made with $n = 10^6$ observations (Embrechts et al., 2013; Taleb, 2020). In the case of events like operational losses, even a larger $n \to \infty$ can still be just anecdotal (Cirillo & Taleb, 2016a).

According to Taleb (2020), owing to preasymptotics, a conservative heuristic is to manage variables with $\alpha \leq 2.5$ as practically unpredictable (for example, see Sections 3.6.2 and 3.6.6). Their sample average is indeed too unstable and needs too many observations for forecasts to be reliable in a reasonable period of time. For $\alpha > 2.5$, conversely, forecasting can take place, and the higher α the better. In any case, more research is strongly needed and desirable (see also Section 4).

Observe that even discussing the optimality of any alarm system (see, for example, Svensson, Holst, Lindquist, & Lindgren, 1996; Turkman & Turkman, 1990) based on average forecasts would prove meaningless under extremely fat tails ($\alpha \leq 2$), when the LLN works very slowly or does not work. In fact, even when the expected value is well-defined (i.e., $1 < \alpha < 2$), the non-existence of the variance would affect all the relevant quantities for the verification of optimality (De Mare, 1980), like for instance the chance of undetected events. For all these quantities, the simple sample estimates commonly used would indeed be misleading.

2.4. Bayesian forecasting

2.4.1. Foundations of Bayesian forecasting[32]

The Bayesian approach to forecasting produces, by default, a probabilistic forecast (see also Section 2.6.2 and Section 2.12.4) describing the uncertainty about future values of the phenomenon of interest, conditional on all *known quantities*; and with uncertainty regarding all *unknown quantities* having been integrated out. In order to produce such forecasts, the Bayesian approach requires (*i*) a predictive model for the future value of the relevant phenomenon, conditional on the observed data and all model unknowns; (*ii*) a model for the observed data; and (*iii*) a distribution describing the (subjective or objective) prior beliefs regarding the model unknowns. Using these quantities, the standard calculus of probability distributions, and Bayes' theorem, then yield the Bayesian predictive (equivalently, forecast) density function (where density is used without loss of generality).

Stated more formally, given observed data up to time n, $\mathbf{y} = (y_1, \ldots, y_n)'$, and denoting the model unknowns by θ, Bayesian forecasting describes the behaviour of the future random variable Y_{n+1} via the predictive density:

$$p(y_{n+1}|\mathbf{y}) = \int p(y_{n+1}|\theta, \mathbf{y})p(\theta|\mathbf{y})d\theta, \qquad (2)$$

where y_{n+1} denotes a value in the support of Y_{n+1} and $p(y_{n+1}|\theta, \mathbf{y})$ is the predictive model for Y_{n+1} conditional on $\mathbf{y}$ and the model unknowns θ. Critically, and in contrast with frequentist approaches to forecasting, parameter uncertainty has been factored into $p(y_{n+1}|\mathbf{y})$ via the process of integration with respect to the posterior probability density function (pdf) for θ, $p(\theta|\mathbf{y})$. The posterior pdf is given, by Bayes' theorem, as $p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta) \times p(\theta)$, where $p(\mathbf{y}|\theta)$ defines the assumed model for $\mathbf{y}$ (equivalently, the likelihood function), and the prior pdf $p(\theta)$ captures prior beliefs about θ. Moreover, uncertainty about the assumed predictive model itself can be easily accommodated using Bayesian model averaging, which involves taking a weighted average of *model-specific* predictives, with posterior model probabilities (also obtained via by Bayes' theorem) serving as the weights. See Greenberg (2008), Koop (2003) and O'Hagan and Forster (2004) for textbook illustrations of all of these steps.

No matter what the data type, the form of predictive model, or the dimension of the unknowns, the basic manner in which all Bayesian forecast problems are framed is the same. What differs however, from problem to problem, is the way in which the forecasting problem is *solved*. To understand why, it is sufficient to recognise that in order to obtain the predictive density $p(y_{n+1}|\mathbf{y})$ we must be able to (somehow) perform the integration that defines this quantity. In almost any practical setting, this integration is infeasible analytically and we must rely on *computational methods* to access the predictive density. Therefore, the evolution of the practice of Bayesian forecasting has gone hand in hand with developments in Bayesian computation (Martin, Frazier, & Robert, 2020). Through the lens of computation, in Section 2.4.2 we briefly describe the methods of implementing Bayesian forecasting.

[32] This subsection was written by David T. Frazier & Gael M. Martin.

2.4.2. Implementation of Bayesian forecasting[33]

If the posterior is accessible via methods of *exact simulation* – e.g., Monte Carlo simulation, importance sampling, Markov chain Monte Carlo (MCMC) sampling, pseudo-marginal MCMC (Andrieu, Doucet, & Holenstein, 2011; Andrieu & Roberts, 2009) – an estimate of the predictive density $p(y_{n+1}|\mathbf{y})$ in (2) can be produced using draws of the unknown θ from the posterior pdf, $p(\theta|\mathbf{y})$. In most cases, this simulation-based estimate of $p(y_{n+1}|\mathbf{y})$ can be rendered arbitrarily accurate by choosing a very large number of posterior draws; hence the use of the term 'exact predictive' to reference this estimate. See Geweke and Whiteman (2006) for an early review of Bayesian forecasting implemented using exact simulation, Sections 2.3.15, 2.4.3, 2.5.3, 3.3.2 and 3.3.10 for further discussions and a range of relevant applications, and Chapters 3, 7, and 9 in Geweke, Koop, and van Dijk (2011) for applications in (general) state space, macroeconometric and finance settings respectively. In addition, a 2008 special issue of *International Journal of Forecasting* on Bayesian Forecasting in Economics provides coverage of forecasting applications (and methods of computation) that exploit exact simulation methods, as do selected chapters in Brooks, Gelman, Jones, and Meng (2011) and O'Hagan and West (2010).

In cases where the posterior is not readily accessible, due to either the intractability of the likelihood function or the high dimension of the model unknowns, or both, methods of *approximation* are required. Frazier et al. (2019), for instance, produce an 'approximate predictive' by replacing the exact posterior in (2), $p(\theta|\mathbf{y})$, with an 'approximate posterior' constructed using *approximate Bayesian computation* (ABC – Sisson, Fan, & Beaumont, 2019). In large samples, this approximate predictive is shown to be equivalent to the exact predictive. The approximate and exact predictives are also shown to be numerically indistinguishable in finite samples, in the cases investigated; see also Canale and Ruggiero (2016), and Kon Kam King, Canale, and Ruggiero (2019). Related work produces an approximate predictive by exploiting *variational Bayes* (Blei, Kucukelbir, & McAuliffe, 2017) approximations of the posterior (Chan & Yu, 2020; Koop & Korobilis, 2018; Loaiza-Maya, Smith, Nott & Danaher, 2020; Quiroz, Nott, & Kohn, 2018; Tran, Nott, & Kohn, 2017). The flavour of this work is broadly similar to that of Frazier et al. (2019); that is, computing the predictive $p(y_{n+1}|\mathbf{y})$ via an approximation to the posterior does not *significantly* reduce predictive accuracy.

The Bayesian paradigm thus provides a very natural and coherent approach to prediction that can be implemented successfully via one of any number of computational methods. Inherent to the approach, however, is the assumption that the model we are using to make predictions is an accurate description of the data generating process (DGP) that has generated the observed data; or if we are averaging across models using Bayesian model averaging, we must assume that the assumed set of models spans the DGP for the observed data. In response to this limitation, allied with the desire to drive prediction by user-specified measures of predictive loss, new approaches to Bayesian prediction have recently been proposed, and we briefly discuss two such classes of methods.

First are methods for combining predictives in which the weights are not equated to the posterior model probabilities (as in standard Bayesian model averaging) but, rather, are updated via problem-specific predictive criteria, or via predictive calibration (Dawid, 1982, 1985; Gneiting, Balabdaoui, & Raftery, 2007); see Bassetti, Casarin, and Ravazzolo (2018), Basturk, Borowska, Grassi, Hoogerheide, and van Dijk (2019), Billio, Casarin, Ravazzolo, and van Dijk (2013), Casarin, Leisen, Molina, and ter Horst (2015), McAlinn, Aastveit, Nakajima, and West (2020), McAlinn and West (2019), and Pettenuzzo and Ravazzolo (2016), for a selection of approaches, and Section 2.6.2 for related discussion. Importantly, these methods do not assume that the true model is spanned by the constituent model set. Second are methods in which the standard Bayesian posterior, which is itself based on a potentially misspecified model, is replaced by a generalised version that is designed for the specific predictive task at hand (e.g., accurate prediction of extreme values); with the overall goal then being to produce predictions that are accurate according to the particular measure of interest. See Frazier, Loaiza-Maya, Martin, and Koo (2021), Loaiza-Maya, Martin and Frazier (2020) and Syring and Martin (2020), for specific examples of this methodology, as well as proofs of its theoretical validity.

2.4.3. Bayesian forecasting with copulas[34]

Copulas provide intrinsic multivariate distribution structure that allows for modelling multivariate dependence with marginal distributions as the input, making it possible to forecast dependent time series and time series dependence. This review focuses on the Bayesian approach for copula forecasting. For rigorous topics on copula introduction (Joe, 1997; Nelsen, 2006; Trivedi & Zimmer, 2007), copula modelling techniques (Durante & Sempi, 2015), vine copulas (Joe, 2014), review on frequentist approaches for copula-based forecasting (Patton, 2013), see the aforementioned references and the references therein.

The advantages of the Bayesian copula approach compared to the frequentist treatments are (*i*) the Bayesian approach allows for jointly modelling the marginal models and the copula parameters, which improves the forecasting efficiency (Joe, 2005; Li & Kang, 2018), (*ii*) probabilistic forecasting is naturally implied with the Bayesian predictive density, and (*iii*) experts information can be seamlessly integrated into forecasting via the priors' setting.

Forecasting with copulas involves selecting between a different class of copulas. Common approaches include Bayesian hypothesis testing (see Section 2.4.1) where copula parameters are treated as nuisance variables (Huard, Évin, & Favre, 2006), or parsimonious modelling of covariance structures using Bayesian selection and model averaging (Min & Czado, 2011; Pitt, Chan, & Kohn, 2006; Smith, 2010).

33 This subsection was written by David T. Frazier & Gael M. Martin.

34 This subsection was written by Feng Li.

One particular interest in copulas is to forecast the dynamic dependencies between multivariate time series. Time-varying copula construction is possible via (*i*) an autoregressive or GARCH form (see Section 2.3.11) of dependence parameters (Lucas, Schwaab, & Zhang, 2014; Patton, 2006), (*ii*) factor copula construction (Oh & Patton, 2018; Tan, Panagiotelis, & Athanasopoulos, 2019) that simplifies the computation, (*iii*) the stochastic copula autoregressive model (Almeida & Czado, 2012) that dependence is modelled by a real-valued latent variable, and (*iv*) covariate-dependent copulas approach by parameterising the dependence as a function of possible time-dependent covariates (Li & Kang, 2018) that also improves the forecasting interpretability. ARMA-like and GARCH-like dependences in the tail can be considered as special cases of Li and Kang (2018).

In multivariate time series forecasting (see also Section 2.3.9), unequal length of data is a common issue. One possible approach is to partition the copula parameters into elements relating only to the marginal distributions and elements only relating to the copula (Patton, 2006a). For mixed frequency data, it is possible to decompose the copula dependence structure into linear and nonlinear components. Then the high and low frequency data is used to model the linear and nonlinear dependencies, respectively (Oh & Patton, 2016). Bayesian data augmentation is also used to forecast multivariate time series with mixed discrete and continuous margins (Smith & Khaled, 2012). For other treatments for discrete and continuous time series (see, for example, Panagiotelis, Czado, & Joe, 2012; Panagiotelis, Czado, Joe, & Stöber, 2017).

The Bayesian approach for lower dimensional copula forecasting ($d < 10$) is straightforward with traditional Gaussian copulas, Student's-t copulas, Archimedean copulas, or pair copula combinations. In higher dimensional settings, special considerations are required to save the computational burden, such as low rank approximation of covariates matrix (Salinas, Michael, Laurent, Roberto & Jan, 2019) or factor copula models with stochastic loadings (Creal & Tsay, 2015).

In the Bayesian setup, forecasting model performance is typically evaluated based on a K-fold out-of-sample log predictive score (LPS: Geweke & Amisano, 2010), and out-of-sample Value-at-Risk (VaR) or Expected Shortfall (ES) are particularly used in financial applications. The LPS is an overall forecasting evaluation tool based on predictive densities, serving out-of-sample probabilistic forecasting. LPS is ideal for decision makers (Geweke, 2001; Geweke & Amisano, 2010). The VaR gives the percentile of the conditional distribution, and the corresponding ES is the expected value of response variable conditional on it lying below its VaR.

2.5. Variable and model selection

2.5.1. Leading indicators and Granger causality[35]

Leading (economic) indicators are variables that try to capture changes in the development of economic activity before those changes materialise. Typically, market participants, policy makers, the general public, etc. are interested in which direction the economy of a country is developing ahead of the publication date of the respective quarterly GDP figures (see also Section 3.3.2), for which these leading indicators play a crucial role. This is of particular relevance for short-term forecasting and nowcasting of such a (low-frequency) variable characterised by a publication lag.

For some leading indicators, single variables are combined into a composite index such as The Conference Board Leading Economic Index published by The Conference Board for the United States. It consists of averages of ten variables, including average weekly hours in manufacturing, manufacturers' new orders, private housing building permits, and average consumer expectations (The Conference Board, 2020). Other leading indicators consist of one variable only and are purely survey-based, such as the monthly ifo Business Climate Index published by the ifo Institute for Germany, which covers the appraisal of current and future expected business of executives of approximately 9,000 German firms across industries (ifo Institute, 2020).

If a specific leading indicator is able to capture certain economic changes before they happen, such a leading indicator x_t is said to Granger-cause (or predictively cause) a forecast variable y_t (Granger, 1969), implying that a "cause" cannot happen after an effect. In (theoretical) econometric terms and following the notation of Lütkepohl (2005), y_t is Granger-caused by x_t if for at least one forecast horizon $h = 1, 2, \dots$ the following inequality holds:

$$\sum_y (h \mid \Omega_i) < \sum_y (h \mid \Omega_i \setminus \{x_t \mid t \leq i\}).$$

In other words, the mean square error of the optimal h-step-ahead predictor of the forecast variable, which includes the information contained in the leading indicator in the set of all relevant information to the forecaster available at the forecast origin $\sum_y(h \mid \Omega_i)$, must be smaller than the mean square error of that h-step-ahead predictor of the forecast variable without the information contained in said leading indicator $\sum_y(h \mid \Omega_i \setminus \{x_t \mid t \leq i\})$ (Lütkepohl, 2005).

Nonetheless, the concept of Granger causality also has its limitations. The notion that a "cause" cannot happen after an effect is only a necessary, yet not a sufficient condition for causality, hence the reason why "cause" has been written with quotation marks in this section. If one ignored this fact and equated causality with Granger causality, they would commit an informal logical fallacy called Post hoc ergo propter hoc (i.e., after this, therefore because of this; Walton, Reed, & Macagno, 2008). A bold example of this fallacy would be, "Because the birds migrate, winter is coming". This is a fallacy, as winter would come at about the same time every year, no matter if there were migrating birds or not. Moreover, hardly any economic activity is monocausal.

There are also different types of formal statistical Granger causality tests available for different data structures that are implemented in typical statistics/econometrics software packages. For the simple case of

[35] This subsection was written by Ulrich Gunter.

two variables (Granger, 1969), say the forecast variable and the leading indicator, the null hypothesis of a bivariate Granger causality test is that the leading indicator does not Granger-cause the forecast variable. Under this null hypothesis, the F-test statistic on the joint impact of the coefficients of the past realisations of the leading indicator employed as explanatory variables in a multiple linear regression with the forecast variable as dependent variable and its past realisations as additional explanatory variables will not be statistically significantly different from zero. The maximum lag for past realisations would be optimally determined, for instance, by some information criterion (e.g., AIC, BIC; see also Section 2.5.4). Practical applications of Granger causality and leading indicators in tourism demand forecasting can be found, for instance, in Section 3.8.1.

2.5.2. Model complexity[36]

A simple model must be easily understood by decision-makers. On the contrary, relationships in a complex model are opaque for its users (see also Section 3.7.4). In this context, complexity is not measured solely by the number of parameters, but also by the functional form of the relationships among variables.

Complex models are commonly believed to deliver better forecasts, as they well describe sophisticated economic structures and offer good fit the data. Consequently, they are favoured by researchers, decision-makers and academic journals. However, empirical studies provide little evidence about their forecasting superiority over simple models. This can be explained using the bias–variance trade-off framework, in which the mean squared error can be decomposed into

$$MSE = noise + variance + bias^2.$$

Noise is driven by the random term in the model. It cannot be avoided, even if we know the true DGP. Variance is caused by the need to estimate the model parameters, hence its value increases with model complexity and declines with the sample size. Bias is predominantly related to model mis-specification, which is most likely to occur for simple methods. The implications of the above framework are twofold: (i) the relationship between model complexity and MSE is U-shaped and (ii) the optimal model complexity increases with the sample size.

The illustration of the bias–variance trade-off for a simple autoregressive model (see Section 2.3.4) is provided by Ca' Zorzi, Muck, and Rubaszek (2016). They present analytical proof that for any persistent DGP of the form $y_t = c + \phi y_{t-1} + \epsilon_t$ characterised by half-life of over two years, the accuracy of forecasts from the random walk or the AR(1) model with parameter ϕ fixed at an arbitrary value consistent with half-life of five years tend to be higher than that from the estimated AR(1) model. This result explains why in numerous studies the random walk is a tough benchmark as well as why a simple, calibrated AR(1) model can be successful in forecasting inflation (Faust & Wright, 2013), exchange rates (Ca' Zorzi, Kolasa, & Rubaszek, 2017) or oil prices (Rubaszek, 2020) compared to a number of complex competitors.

Wide support to the view that model simplicity improves forecasting performance is presented by Green and Armstrong (2015), in an introductory article to the special issue of *Journal of Business Research* "Simple versus complex forecasting", as well as the results of M1 and M2 competitions (Makridakis et al., 1982, 1993). Stock and Watson (1998) also show that for most US monthly series complex non-linear autoregressive models deliver less accurate forecasts than their linear counterparts. On the contrary, the results of M4 competition tend to favour more complex models (Makridakis, Spiliotis & Assimakopoulos, 2020, and Section 2.12.7).

Why are then complex models preferred to simple ones, if the latter deliver more accurate forecasts? Brighton and Gigerenzer (2015) claim that there is a tendency to overemphasize the bias component and downplay the role of variance. This behaviour is leading to an implicit preference towards more complex forecasting methods, which is called by the authors as "bias bias". To avoid it, one can follow the golden rules of forecasting, which says: be conservative in the choice of over-ambitious models and be wary of the difficulty to forecast with complex methods (Armstrong, Green, & Graefe, 2015). The alternative is to use methods, which explicitly account for the bias–variance trade-off, e.g. machine learning (see Section 2.7.10).

2.5.3. Variable selection[37]

References to 'big data' have become somewhat ubiquitous both in the media and in academic literature in recent years (see Section 2.7.1 but also Section 2.2.5). Whilst in some disciplines (for example Finance) it has become possible to observe time series data at ever higher frequency, it is in the cross section that the amount of data available to analysts has seen exponential growth.

Ordinary Least Squares (OLS) is the standard tool for data analysis and prediction, but is well known to perform poorly when there many potential explanatory variables; Bishop (2006) sets out clearly why this is so. In situations where there is no obvious underlying model to suggest which of a potentially large set of candidate variables to focus on, the researcher needs to add new tools to the tool kit.

There are two principal sets of approaches to this problem. The first seeks to reduce the model dimensions by summarising the predictors in to a much smaller number of aggregated indices or factors. The second is to employ a regularisation strategy to reduce the effective dimension of the problem by shrinking the size of the regression coefficients. Some strategies reduce a subset of these coefficients to zero, removing some variables from the model entirely and these can hence be described as variable selection procedures.

Such procedures may be applicable either when a problem is believed to be truly sparse, with only a small number of variables having an effect, or alternatively when a sparse model can provide effective forecasts, even

36 This subsection was written by Michał Rubaszek.

37 This subsection was written by Ross Hollyman.

though the underling system is in fact more complex (see also Section 2.5.2).

In the frequentist framework, the Least Absolute Shrinkage and Selection Operator (LASSO) procedure of Tibshirani (1996) has proven effective in many applications. The LASSO requires choice of an additional regularisation parameter (usually selected by some statistical criteria or via a cross validation process). Various refinements to the original LASSO procedure have been developed, see in particular Rapach, Strauss, Tu, and Zhou (2019) for a recent forecasting application. An alternative frequentist approach to variable selection and shrinkage is the Complete Subset Regression (CSR) model of Elliott (2015), where separate OLS regressions are run on all possible combinations of potential regressors, with forecasts generated by averaging across the entire set of models. Kotchoni, Leroux, and Stevanovic (2019) combine CSR with LASSO in a comprehensive empirical economic forecasting exercise.

Where the underlying process is not properly sparse (see Giannone & Primiceri, 2017, for a discussion), it is perhaps more natural to work in a Bayesian framework where samples can be drawn from the variable selection part of model reflecting the estimated probability of inclusion of each variable. The appropriate degree of regularisation can also be selected in a similar way. Forecasts are then constructed as weighted averages over several sparse models. This approach has proven to be very effective in practice, for example in fitting Bayesian Vector Auto Regressions to forecast economic time series. Early examples include George and McCulloch (1993) and Mitchell and Beauchamp (1988), which use binary indicators to select variables in to the model. More recent approaches use continuous random variables to achieve a similar effect, making computation more tractable. Examples include the Horeshoe Prior (Carvalho, Polson, & Scott, 2010; Piironen & Vehtari, 2017) and the LN-CASS prior of Thomson, Jabbari, Taylor, Arlt, and Smith (2019). Cross (2020) is a recent example of an economic forecasting exercise using several such models.

2.5.4. Model selection[38]

Taxonomies of all possible sources of forecast errors from estimated models that do not coincide with their data generating process (DGP) have revealed that two mistakes determine forecast failures (i.e., systematic departures between forecasts and later outcomes), namely mis-measured forecast origins and unanticipated location shifts (Clements & Hendry, 1999; Hendry & Mizon, 2012). The former can be addressed by nowcasting designed to handle breaks (Bańbura, Giannone, & Reichlin, 2011; Castle & Hendry, 2010; Castle, Hendry, & Kitov, 2018). It is crucial to capture the latter as failure to do so will distort estimation and forecasts, so must be a focus when selecting models for forecasting facing an unknown number of in-sample contaminating outliers and multiple breaks at unknown times.

Consequently, selection must be jointly over both observations and variables, requiring computer learning methods (Castle, Doornik, & Hendry, 2020b; Hendry & Doornik, 2014, see also Section 3.6.2). *Autometrics*, a multiple-path block-search algorithm (see Doornik, 2018), uses impulse (IIS: Hendry, Johansen, & Santos, 2008a and Johansen & Nielsen, 2009) and step (SIS: Castle, Doornik, Hendry, & Pretis, 2015b) indicator saturation for discovering outliers and breaks, as well as selecting over other variables. Although knowing the in-sample DGP need not suffice for successful forecasting after out-of-sample shifts, like financial crises and pandemics, 'robust variants' of selected models can then avoid systematic mis-forecasts (see, for example, Doornik, Castle, & Hendry, 2020a; Martinez et al., 2021, and Section 3.3.4).

Saturation estimation has approximately $K = k + 2n$ candidates for k regressors, with 2^K possible models, requiring selection with more candidate variables, K, than observations, n (also see Section 2.7 and Section 2.3.9). Selection criteria like AIC (Akaike, 1973), BIC (Schwarz, 1978), and HQ (Hannan & Quinn, 1979) are insufficient in this setting. For saturation estimation, we select at a tight significance level, $\alpha = 1/K$, retaining subject-theory variables and other regressors. When forecasting is the aim, analyses and simulations suggest loose significance for then selecting other regressors, close to the 10% to 16% implied significance level of AIC, regardless of location shifts at or near the forecast origin (Castle, Doornik, & Hendry, 2018a). At loose levels, *Autometrics* can find multiple undominated terminal models across paths, and averaging over these, a univariate method and a robust forecasting device can be beneficial, matching commonly found empirical outcomes. The approach applies to systems, whence selection significance of both indicators and variables is judged at the system level. Capturing in-sample location shifts remains essential (Doornik, Castle, & Hendry, 2020b).

There are costs and benefits of selection for forecasting (also see Sections 2.5.3 and 2.11.3). Selection at a loose significance level implies excluding fewer relevant variables that contribute to forecast accuracy, but retaining more irrelevant variables that are adventitiously significant, although fewer than by simply averaging over all sub-models (Hoeting, Madigan, Raftery, & Volinsky, 1999). Retaining irrelevant variables that are subject to location shifts worsens forecast performance, but their coefficient estimates are driven towards zero when updating estimates as the forecast origin moves forward. Lacking omniscience about future values of regressors that then need to be forecast, not knowing the DGP need not be costly relative to selecting from a general model that nests it (Castle et al., 2018a). Overall, the costs of model selection for forecasting are small compared to the more fundamental benefit of finding location shifts that would otherwise induce systematic forecast failure.

2.5.5. Cross-validation for time-series data[39]

When building a predictive model, its purpose is usually not to predict well the already known samples, but to obtain a model that will generalise well to new, unseen

[38] This subsection was written by David F. Hendry.

[39] This subsection was written by Christoph Bergmeir.

data. To assess the out-of-sample performance of a predictive model we use a test set that consists of data not used to estimate the model (for a discussion of different error measures used in this context see Section 2.12.2). However, as we are now using only parts of the data for model building, and other parts of the data for model evaluation, we are not making the best possible use of our data set, which is a problem especially if the amount of data available is limited. Cross-validation (CV), first introduced by Stone (1974), is a widely used standard technique to overcome this problem (Hastie et al., 2009) by using all available data points for both model building and testing, therewith enabling a more precise estimation of the generalisation error and allowing for better model selection. The main idea of (k-fold) cross-validation is to partition the data set randomly into k subsets, and then use each of the k subsets to evaluate a model that has been estimated on the remaining subsets. An excellent overview of different cross-validation techniques is given by Arlot and Celisse (2010).

Despite its popularity in many fields, the use of CV in time series forecasting is not straightforward. Time series are often non-stationary and have serial dependencies (see also Section 2.3.4). Also, many forecasting techniques iterate through the time series and therewith have difficulties dealing with missing values (withheld for testing). Finally, using future values to predict data from the past is not in accordance with the normal use case and therewith seems intuitively problematic. Thus, practitioners often resort to out-of-sample evaluation, using a subset from the very end of a series exclusively for evaluation, and therewith falling back to a situation where the data are not used optimally.

To overcome these problems, the so-called time-series cross-validation (Hyndman & Athanasopoulos, 2018) extends the out-of-sample approach from a fixed origin to a rolling origin evaluation (Tashman, 2000). Data is subsequently moved from the out-of-sample block from the end of the series to the training set. Then, the model can be used (with or without parameter re-estimation) with the newly available data. The model re-estimation can be done on sliding windows with fixed window sizes or on expanding windows that always start at the beginning of the series (Bell & Smyl, 2018).

However, these approaches extending the out-of-sample procedure make again not optimal use of the data and may not be applicable when only small amounts of data are available. Adapting the original CV procedure, to overcome problems with serial correlations, blocked CV approaches have been proposed in the literature, where the folds are chosen in blocks (Bergmeir & Benítez, 2012; Racine, 2000) and/or data around the points used for testing are omitted (Burman, Chow, & Nolan, 1994; Racine, 2000). Finally, it has been shown that with purely autoregressive models, CV can be used without modifications, i.e., with randomly choosing the folds (Bergmeir, Hyndman, & Koo, 2018). Here, CV estimates the generalisation error accurately, as long as the model errors from the in-sample fit are uncorrelated. This especially holds when models are overfitting. Underfitting can be easily detected by checking the residuals for serial correlation, e.g., with a Ljung–Box test (Ljung & Box, 1978). This procedure is implemented in the *forecast* package (Hyndman et al., 2020) in R (R Core Team, 2020), in the function `CVar`.

2.6. Combining forecasts

2.6.1. Forecast combination: a brief review of statistical approaches[40]

Given N forecasts of the same event, forecast combination involves estimation of so called combination weights assigned to each forecast, such that the accuracy of the combined forecast generally outperforms the accuracy of the forecasts included. Early statistical approaches adopted a range of strategies to estimate combination weights including (*i*) minimising in-sample forecast error variance among forecast candidates (Bates & Granger, 1969; Min & Zellner, 1993; Newbold & Granger, 1974), (*ii*) formulation and estimation via ordinary least squares regression (Granger & Ramanathan, 1984; MacDonald & Marsh, 1994), (*iii*) use of approaches based on Bayesian probability theory (Bordley, 1982; Bunn, 1975; Clemen & Winkler, 1986; Diebold & Pauly, 1990; Raftery, 1993, and Section 2.4), (*iv*) and the use of regime switching and time varying weights recognising that weights can change over time (Diebold & Pauly, 1987; Elliott, Timmermann, & Komunjer, 2005; Lütkepohl, 2011; Tian & Anderson, 2014, and Section 2.3.12). Barrow and Kourentzes (2016) contains a very good documentation and empirical evaluation of a range of these early approaches, while De Menezes, Bunn, and Taylor (2000) and Armstrong (2001a) contain guidelines on their use.

Recent statistical approaches use a variety of techniques to generate forecasts and/or derive weights. Kolassa (2011) apply so called Akaike weights based on Akaike Information Criterion (Sakamoto, Ishiguro, & Kitagawa, 1986), while bootstrapping has been used to generate and combine forecast from exponential smoothing (Barrow, Kourentzes, Sandberg, & Niklewski, 2020; Bergmeir, Hyndman, & Benítez, 2016; Cordeiro & Neves, 2009, but also Section 2.7.5 and Section 2.7.6), artificial neural networks (Barrow & Crone, 2016a, 2016b, and Section 2.7.8), and other forecast methods (Athanasopoulos, Song, & Sun, 2018; Hillebrand & Medeiros, 2010; Inoue & Kilian, 2008). Barrow and Crone (2016b) developed cross-validation and aggregating (Crogging) using cross-validation to generate and average multiple forecasts, while more recently, combinations of forecasts generated from multiple temporal levels has become popular (Athanasopoulos, Hyndman, Kourentzes, & Petropoulos, 2017; Kourentzes & Athanasopoulos, 2019; Kourentzes, Petropoulos, & Trapero, 2014, and Section 2.10.2). These newer approaches recognise the importance of forecast generation in terms of uncertainty reduction (Petropoulos, Hyndman, & Bergmeir, 2018a), the creation of diverse forecasts (Brown, Wyatt, Harris, & Yao, 2005b; Lemke & Gabrys, 2010), and the pooling of forecasts (Kourentzes, Barrow, & Petropoulos, 2019; Lichtendahl Jr & Winkler, 2020).

[40] This subsection was written by Devon K. Barrow.

Now nearly 50 years on from the seminal work of Bates and Granger (1969), the evidence that statistical combinations of forecasts improves forecasting accuracy is near unanimous, including evidence from competitions (Makridakis et al., 1982; Makridakis & Hibon, 2000; Makridakis, Spiliotis et al., 2020, and Section 2.12.7), and empirical studies (Andrawis, Atiya, & El-Shishiny, 2011; Elliott et al., 2005; Jose & Winkler, 2008; Kourentzes et al., 2019). Still, researchers have tried to understand why and when combinations improve forecast accuracy (Atiya, 2020; Palm & Zellner, 1992; Petropoulos et al., 2018a; Timmermann, 2006), and the popularity of the simple average (Chan & Pauwels, 2018; Claeskens, Magnus, Vasnev, & Wang, 2016; Smith & Wallis, 2009a). Others have investigated properties of the distribution of the forecast error beyond accuracy considering issues such as normality, variance, and in out-of-sample performance of relevance to decision making (Barrow & Kourentzes, 2016; Chan, Kingsman, & Wong, 1999; Makridakis & Winkler, 1989).

Looking forward, evidence suggests that the future lies in the combination of statistical and machine learning generated forecasts (Makridakis, Spiliotis et al., 2020), and in the inclusion of human judgment (Gupta, 1994; Petropoulos, Kourentzes, Nikolopoulos, & Siemsen, 2018b; Wang & Petropoulos, 2016, but also Section 2.11.1). Additionally, there is need to investigate such issues as decomposing combination accuracy gains, constructing prediction intervals (Grushka-Cockayne & Jose, 2020; Koenker, 2005), and generating combined probability forecasts (Clements & Harvey, 2011; Hall & Mitchell, 2007; Raftery, Madigan, & Hoeting, 1997; Ranjan & Gneiting, 2010, and Section 2.6.2). Finally, there is need for the results of combined forecasts to be more interpretable and suitable for decision making (Barrow & Kourentzes, 2016; Bordignon, Bunn, Lisi, & Nan, 2013; Graefe, Armstrong, Jones Jr, & Cuzán, 2014; Todini, 2018).

2.6.2. Density forecast combinations[41]

Density forecasts provide an estimate of the future probability distribution of a random variable of interest. Unlike point forecasts (and point forecasts supplemented by prediction intervals) density forecasts provide a complete measure of forecast uncertainty. This is particularly important as it allows decision makers to have full information about the risks of relying on the forecasts of a specific model. Policy makers like the Bank of England, the European Central Bank and Federal Reserve Banks in the US routinely publish density forecasts of different macroeconomic variables such as inflation, unemployment rate, and GDP. In finance, density forecasts find application, in particular, in the areas of financial risk management and forecasting of stock returns (see, for example, Berkowitz, 2001; Guidolin & Timmermann, 2006; Shackleton, Taylor, & Yu, 2010; Tay & Wallis, 2000, *inter alia*). The reader is referred to Section 3.3 for a discussion of relevant applications.

Initial work on forecast combination focused on the combination of point forecasts (see Section 2.6.1). In recent years attention has shifted towards evaluation, comparison and combination of density forecasts with empirical applications that are mostly encountered in the areas of macroeconomics and finance. The improved performance of combined density forecasts stems from the fact that pooling of forecasts allows to mitigate potential misspecifications of the individual densities when the true population density is non-normal. Combining normal densities yields a flexible mixture of normals density which can accommodate heavier tails (and hence skewness and kurtosis), as well as approximate non-linear specifications (Hall & Mitchell, 2007; Jore, Mitchell, & Vahey, 2010).

The predictive density combination schemes vary across studies and range from simple averaging of individual density forecasts to complex approaches that allow for time-variation in the weights of prediction models, also called experts (see Aastveit, Mitchell, Ravazzolo, & van Dijk, 2019, for a comprehensive survey of density forecast combination methods). A popular approach is to combine density forecasts using a convex combination of experts' predictions, so called 'linear opinion pools' (see, for example, Geweke & Amisano, 2011; Hall & Mitchell, 2007; Kascha & Ravazzolo, 2010). In order to determine the optimal combination weights this method relies on minimising Kullback–Leibler divergence of the true density from the combined density forecast. These linear approaches have been extended by Gneiting and Ranjan (2013) to allow non-linear transformations of the aggregation scheme and by Kapetanios, Mitchell, Price, and Fawcett (2015), whose 'generalised pools' allow the combination weights to depend on the (forecast) value of the variable of interest.

Billio et al. (2013) developed a combination scheme that allows the weights associated with each predictive density to be time-varying, and propose a general state space representation of predictive densities and combination schemes. The constraint that the combination weights must be non-negative and sum to unity implies that linear and Gaussian state-space models cannot be used for inference and instead Sequential Monte Carlo methods (particle filters) are required. More recently, McAlinn and West (2019) developed a formal Bayesian framework for forecast combination (Bayesian predictive synthesis) which generalises existing methods. Specifying a dynamic linear model for the synthesis function, they develop a time-varying (non-convex/nonlinear) synthesis of predictive densities, which forms a dynamic latent (agent) factor model.

For a discussion on methods for evaluating probabilistic forecasts, see Sections 2.12.4 and 2.12.5.

2.6.3. Ensembles and predictive probability post processors[42]

Improved rational decisions are the final objective of modelling and forecasting. Relatively easy decisions among a number of alternatives with predefined and known outcomes become hard when they are conditioned by future unknown events. This is why one resorts to modelling and forecasting, but this is insufficient. To be successful, one must account for the future conditioning event uncertainty to be incorporated into the decision-making process using appropriate Bayesian approaches

[41] This subsection was written by Alisa Yusupova.

[42] This subsection was written by Ezio Todini.

(see also Section 2.4), as described by the decision theory literature (Berger, 1985; Bernardo, 1994; DeGroot, 2004). The reason is that taking a decision purely based on model forecasts is equivalent to assuming the future event (very unlikely, as we know) to equal the forecasted value. Therefore, the estimation of the predictive probability density is the essential prerequisite to estimating the expected value of benefits (or of losses) to be compared and traded-off in the decision-making process (Draper & Krnjajić, 2013). This highly increases the expected advantages together with the likelihood of success and the robustness of the decision (Todini, 2017, 2018).

In the past, the assessment of the prediction uncertainty was limited to the evaluation of the confidence limits meant to describe the quality of the forecast. This was done using continuous predictive densities, as in the case of the linear regression, or more frequently in the form of predictive ensembles. These probabilistic predictions, describing the uncertainty of the model forecasts given (knowing) the observations can be used within the historical period to assess the quality of the models (Todini, 2016). When predicting into the future, observations are no more available and what we look for (known as predictive probability) is the probability of occurrence of the unknown "future observations" given (knowing) the model forecasts. This can be obtained via Bayesian inversion, which is the basis of several uncertainty post-processors used in economy (Diebold, Gunther, & Tay, 1998, and Section 3.3), hydrology (Krzysztofowicz, 1999; Schwanenberg, Fan, Naumann, Kuwajima, Montero, & Assis dos Reis, 2015; Todini, 1999, 2008, and Section 3.5.4), meteorology (Economou, Stephenson, Rougier, Neal, & Mylne, 2016; Granger & Pesaran, 2000; Katz & Lazo, 2011; Reggiani & Boyko, 2019, see also Section 3.5.2), etc. Accordingly, one can derive a predictive probability from a single model forecast to estimate the expected value of a decision by integrating over the entire domain of existence all the possible future outcomes and theirs effects, weighted with their appropriate probability of occurrence.

When several forecasts are available to a decision maker, the problem of deciding on which of them one should rely upon becomes significant. It is generally hard to choose among several forecasts because one model could be best under certain circumstances but rather poor under others. Accordingly, to improve the available knowledge on a future unknown event, predictive densities are extended to multiple forecasts to provide decision makers with a single predictive probability, conditional upon several model's forecasts (Coccia & Todini, 2011; Raftery et al., 1997, see also Section 2.6.1 and Section 2.6.2).

A number of available uncertainty post processors can cope with multi-model approaches, such as Bayesian model averaging (Raftery, 1993; Raftery et al., 1997), model output statistics (Glahn & Lowry, 1972; Wilkd, 2005), ensemble model output statistics (Gneiting, Raftery, Westveld, & Goldman, 2005), and model conditional processor (Coccia & Todini, 2011; Todini, 2008).

Finally, important questions such as: (*i*) "what is the probability that an event will happen within the next x hours?" and (*ii*) "at which time interval it will most likely occur?" can be answered using a multi-temporal approach (Coccia, 2011; Krzysztofowicz, 2014, see also Section 2.10.2) and results of its applications were presented in Barbetta, Coccia, Moramarco, Brocca, and Todini (2017), Coccia (2011), and Todini (2017).

2.6.4. The wisdom of crowds[43]

Multiple experts' forecasts are collected in a wide variety of situations: medical diagnostics, weather prediction, forecasting the path of a hurricane, predicting the outcome of an election, macroeconomic forecasting, and more. One of the central findings from the forecasting literature is that there is tremendous power in combining such experts' forecasts into a single forecast. The simple average, or what Surowiecki refers to as 'the wisdom of crowds' (Surowiecki, 2005), has been shown to be a surprisingly robust combined forecast in the case of point forecasting (Armstrong, 2001b; Clemen, 1989; Clemen & Winkler, 1986, and Section 2.6.1). The average forecast is more accurate than choosing a forecast from the crowd at random and is sometimes even more accurate than nearly all individuals (Mannes, Larrick, & Soll, 2012). The average point forecast also often outperforms more complicated point aggregation schemes, such as weighted combinations (Smith & Wallis, 2009b; Soule, Grushka-Cockayne, & Merrick, 2020).

Mannes et al. (2012) highlight two crucial factors that influence the quality of the average point forecast: individual expertise and the crowd's diversity. Of the two: "The benefits of diversity are so strong that one can combine the judgments from individuals who differ a great deal in their individual accuracy and still gain from averaging" (Mannes et al., 2012, page 234).

Larrick and Soll (2006) define the idea of 'bracketing': In the case of averaging, two experts can either bracket the realisation (the truth) or not. When their estimates bracket, the forecast generated by taking their average performs better than choosing one of the two experts at random; when the estimates do not bracket, averaging performs equally as well as the average expert. Thus, averaging can do no worse than the average expert, and with some bracketing, it can do much better. Modern machine learning algorithms such as the random forest exploit this property by averaging forecasts from hundreds of diverse experts (here, each "expert" is a regression tree; Grushka-Cockayne, Jose & Lichtendahl, 2017).

Only when the crowd of forecasts being combined has a high degree of dispersion in expertise, some individuals in the crowd might stand out, and in such cases, there could be some benefits to chasing a single expert forecaster instead of relying on the entire crowd. Mannes, Soll, and Larrick (2014) suggest that combining a small crowd can be especially powerful in practice, offering some diversity among a subset of forecasters with an minimum level of expertise.

When working with probabilistic forecasting (see also Sections 2.6.2, 2.6.3 and 2.12.4), averaging probabilities is the most widely used probability combination method

[43] This subsection was written by Yael Grushka-Cockayne.

(Clemen, 2008; Cooke, 1991; Hora, 2004). Stone (1961) labelled such an average the linear opinion pool. O'Hagan, Buck, Daneshkhah, Richard Eiser, Garthwaite, Jenkinson, Oakley, and Rakow (2006) claimed that the linear opinion pool is: "hard to beat in practice".

Although diversity benefits the average point forecast, it can negatively impact the average probability forecast. As the crowd's diversity increases, the average probability forecast becomes more spread out, or more underconfident (Dawid, DeGroot, Mortera, Cooke, French, Genest, et al., 1995; Hora, 2004; Ranjan & Gneiting, 2010). Averaging quantiles, instead of probabilities, can offer sharper and better calibrated forecasts (Lichtendahl, Grushka-Cockayne, & Winkler, 2013). Trimmed opinion pools can be applied to probability forecasts, also resulting in better calibrated forecasts (Jose, Grushka-Cockayne, & Lichtendahl, 2014, see also Section 2.12.5).

The ubiquity of data and the increased sophistication of forecasting methods results in more use of probabilistic forecasts. While probabilities are more complex to elicit, evaluate, and aggregate compared to point estimates, they do contain richer information about the uncertainty of interest. The wisdom of combining probabilities, however, utilises diversity and expertise differently than combining point forecasts. When relying on a crowd, eliciting point forecasts versus eliciting probabilities can significantly influence the type of aggregation one might choose to use.

2.7. Data-driven methods

2.7.1. Forecasting with big data[44]

The last two decades have seen a proliferation of literature on forecasting using big data (Hassani & Silva, 2015; Swanson & Xiong, 2018; Varian, 2014) but the evidence is still uncertain as to whether the promised improvements in forecast accuracy can systematically be realised for macroeconomic phenomena. In this section we question whether big data will significantly increase the forecast accuracy of macroeconomic forecasts. Athey (2018) argues that machine learning methods are seen as an efficient approach to dealing with big data sets, and we present these methods before questioning their success at handling non-stationary macroeconomic data that are subject to shifts. Section 2.7.2 discusses big data in the context of distributed systems, and Section 2.7.11 evaluates a range of machine learning methods frequently applied to big data.

The tools used to analyse big data focus on regularization techniques to achieve dimension reduction, see Kim and Swanson (2014) for a summary of the literature. This can be achieved through selection (such as *Autometrics*, Doornik, 2018, but also see Section 2.5.3 and Section 2.5.4), shrinkage (including Ridge Regression, LASSO, and Elastic Nets, see Section 2.7.11 but also Section 3.3.13 for an applied example), variable combination (such as Principal Components Analysis and Partial Least Squares), and machine learning methods (including Artificial Neural Networks, see Section 2.7.8). Many of these methods are 'black boxes' where the algorithms are not easily interpretable, and so they are mostly used for forecasting rather than for policy analysis.

Big data has been effectively used in nowcasting, where improved estimates of the forecast origin lead to better forecasts, absent any later shifts. Nowcasting can benefit from large data sets as the events have happened and the information is available, see Castle et al. (2018) for a nowcasting application, and Section 2.5.1 on leading indicators. However, the benefits of big data are not as evident in a forecasting context where the future values of all added variables also need to be forecast and are as uncertain as the variable(s) of interest.

Macroeconomic time series data are highly non-stationary, with stochastic trends and structural breaks. The methods of cross-validation and hold-back, frequently used to handle bid data, often assume that the data generating process does not change over time. Forecasting models that assume the data are drawn from a stationary distribution (even after differencing) do not forecast well *ex ante*. So while there seems to be lots of mileage in improving forecasts using big data, as they allow for more flexible models that nest wider information sets, more general dynamics and many forms of non-linearity, the statistical problems facing 'small' data forecasting models do not disappear (Doornik & Hendry, 2015; Harford, 2014). Castle, Doornik, and Hendry (2020a) do not find improvements in forecasting from big data sets over small models. It is essential to keep in mind the classical statistical problems of mistaking correlation for causation, ignoring sampling biases, finding excess numbers of false positives and not handling structural breaks and non-constancies both in- and out-of-sample, in order to guard against these issues in a data abundant environment.

2.7.2. Forecasting on distributed systems[45]

Big data is normally accompanied by the nature that observations are indexed by timestamps, giving rise to big data time series characterised by high frequency and long-time span. Processing big data time series is obstructed by a wide variety of complications, such as significant storage requirements, algorithms' complexity and high computational cost (Galicia, Torres, Martínez-Álvarez, & Troncoso, 2018; L'heureux, Grolinger, Elyamany, & Capretz, 2017; Wang, Kang, Hyndman, & Li, 2020; Wang, Yang, Du, & Niu, 2018). These limitations accelerate the great demand for scalable algorithms. Nowadays, increasing attention has been paid to developing data mining techniques on distributed systems for handling big data time series, including but not limited to processing (Mirko & Kantelhardt, 2013), decomposition (Bendre & Manthalkar, 2019), clustering (Ding, Wang, Dang, Fu, Zhang, & Zhang, 2015), classification (Triguero, Peralta, Bacardit, García, & Herrera, 2015), and forecasting (Galicia et al., 2018). For forecasting problems based on big data sets and/or large sets of predictors, please refer to Sections 2.7.1 and 3.3.13.

Distributed systems, initially designed for independent jobs, do not support to deal with dependencies among observations, which is a critical obstacle in time series processing (Li, Noorian, Moss, & Leong, 2014; Wang, Kang,

[44] This subsection was written by Jennifer L. Castle.

[45] This subsection was written by Xiaoqian Wang.

Hyndman, & Li, 2020). Various databases (e.g., InfluxDB,[46] OpenTSDB,[47] RRDtool,[48] and Timely[49]) can function as storage platforms for time series data. However, none of these databases supports advanced analysis, such as modelling, machine learning algorithms and forecasting. Additional considerations are therefore required in further processing time series on distributed systems. Mirko and Kantelhardt (2013) developed the Hadoop.TS library for processing large-scale time series by creating a time series bucket. Li et al. (2014) designed an index pool serving as a data structure for assigning index keys to time series entries, allowing time series data to be sequentially stored on HDFS (Hadoop Distributed File System: Shvachko, Kuang, Radia, & Chansler, 2010) for MapReduce (Dean & Ghemawat, 2008) jobs. Chen, Li, Rong, Bilal, Li, and Philip (2019a) proposed a data compression and abstraction method for large-scale time series to facilitate the periodicity-based time series prediction in a parallel manner.

The evolution of the algorithms for efficiently forecasting big data time series on distributed systems is largely motivated by a wide range of applications including meteorology, energy, finance, transportation and farming (Chen et al., 2019a; Galicia et al., 2018; Hong & Pinson, 2019; Sommer, Pinson, Messner, & Obst, 2020). Researchers have made several attempts to make machine learning techniques available for big data time series forecasting on distributed systems (Galicia, Talavera-Llames, Troncoso, Koprinska, & Martínez-Álvarez, 2019; Li et al., 2014; Talavera-Llames, Pérez-Chacón, Martínez-Ballesteros, Troncoso, & Martínez-Álvarez, 2016; Xu, Liu, & Long, 2020). Talavera-Llames et al. (2016) presented a nearest neighbours-based algorithm implemented for Apache Spark (Zaharia, Xin, Wendell, Das, Armbrust, Dave, Meng, Rosen, Venkataraman, Franklin, Ghodsi, Gonzalez, Shenker, & Stoica, 2016) and achieved satisfactory forecasting performance. Galicia et al. (2018) proposed a scalable methodology which enables Spark's MLlib (Meng et al., 2016) library to conduct multi-step forecasting by splitting the multi-step forecasting problem into h sub-problems (h is the forecast horizon).

Another strand of the literature on forecasting big data time series is to improve time-consuming estimation methods using a MapReduce framework. Sheng, Zhao, Leung, and Wang (2013) learned the parameters of echo state networks for time series forecasting by designing a parallelised extended Kalman filter involving two MapReduce procedures. Recently, Sommer et al. (2020) accurately estimated coefficients of a high-dimensional ARX model by designing two online distributed learning algorithms. Wang, Kang, Hyndman, and Li (2020) resolved challenges associated with forecasting ultra-long time series from a new perspective that global estimators are approximated by combining the local estimators obtained from subseries by minimising a global loss function. Besides, inspired by the *no-free-lunch* theorem (Wolpert & Macready, 1997), model selection (see Section 2.5.4) and model combination (see Section 2.6) are involved in finalisation of algorithms for forecasting on distributed systems (e.g., Bendre & Manthalkar, 2019; Galicia et al., 2019; Li et al., 2014; Xu et al., 2020).

2.7.3. Agent-based models[50]

Time series forecasting involves use of historical data to predict values for a specific period time in future. This approach assumes that recent and historical patterns in the data will continue in the future. This assumption is overly ingenuous. However, this is not reliable in some situations. For example, (*i*) forecasting COVID-19 cases (see also Section 3.6.2) where, due to interventions and control measures taken by the governments and due to the change in personal behaviour, the disease transmission pattern changes rapidly, and (*ii*) forecasting sales of a new product (see also Section 3.2.6): external factors such as advertisement, promotions (see Section 3.2.5), social learning, and imitation of other individuals change the system behaviour.

In such circumstances to make reliable forecasts it is important to take into account all information that might influence the variable that is being forecast. This information includes a variety of environmental-level and individual-level factors. An agent-based modelling is a powerful tool to explore such complex systems. Agent-based modelling approach is useful when, (*i*) data availability is limited, (*ii*) uncertainty of various interventions in place and a rapidly changing social environment, and (*iii*) limited understanding of the dynamics of the variable of interest.

Agent-based modelling disaggregates systems into individual level and explores the aggregate impact of individual behavioural changes on the system as a whole. In other words, the key feature of agent-based modelling is the bottom-up approach to understand how a system's complexity arises, starting with individual level (see also Section 2.10.1). As opposed to this, the conventional time series forecasting approaches are considered top-down approaches.

Agent-based models have two main components: (*i*) Agents, and (*ii*) Rules, sometimes referred as procedures and interactions. Agents are individuals with autonomous behaviour. Agents are heterogeneous. Each agent individually assesses on the basis of a set of rules. An agent-based modelling approach simulates how heterogeneous agents interact and behave to assess the role of different activities on the target variable. According to Farmer and Foley (2009), "An agent-based model is a computerised simulation of a number of decision-makers (agents) and institutions, which interact through prescribed rules". Their paper highlights the importance of adopting agent-based models as a better way to help guide financial policies.

A general framework for agent-based modelling involves three main stages (See Fig. 4): (*i*) setup environments and agents, (*ii*) agent-based modelling, and (*iii*) calibration and validation. The first two steps are

---

[46] Available at https://www.influxdata.com/time-series-database/.

[47] Available at http://opentsdb.net/

[48] Available at https://oss.oetiker.ch/rrdtool/.

[49] Available at https://code.nsa.gov/timely/.

[50] This subsection was written by Thiyanga S. Talagala.

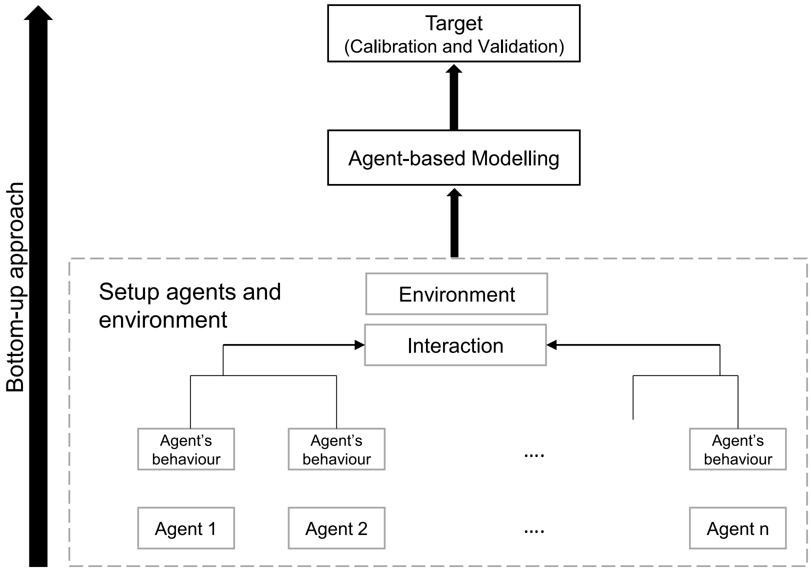

Fig. 4. Framework for Agent-based modelling.

self-explanatory. The final step involves calibration of the model with empirical data and then evaluates whether the agent-based model mirrors the real-world system/target. The validation step involves testing the significance of the difference between agent-based model results and real data collected about the target. One of the main challenges in designing an agent-based model is finding a balance between model simplicity and model realism (see also Section 2.5.2). The KISS principle (keep it simple, stupid), introduced by Axelrod (1997) is often cited as an effective strategy in agent-based modelling. A high level of expertise in the area of the subject is necessary when developing an agent-based model.

Despite these limitations and challenges, agent-based modelling has been used extensively to model infectious disease transmission and forecasting (Tracy, Cerdá, & Keyes, 2018; Venkatramanan, Lewis, Chen, Higdon, Vullikanti, & Marathe, 2018). Agent-based modelling approaches have been widely used in early phases of the COVID-19 outbreak, to assess the impact of different interventions on disease spread and forecasts (Wallentin, Kaziyeva, & Reibersdorfer-Adelsberger, 2020). In a review paper, Weron (2014) states some applications of agent-based models for electricity demand forecasting. Xiao and Han (2016) use agent-based models to forecast new product diffusion. Furthermore, thinking the other way around, Hassan, Arroyo, Galán Ordax, Antunes, and Pavón Mestras (2013) explain how forecasting principles can be applied in agent-based modelling.

2.7.4. *Feature-based time series forecasting*[51]

A time series *feature* can be any statistical representation of time series characteristics. A vast majority of time series mining tasks are based on similarity quantification using their feature representations, including but not limited to time series clustering (Kang, Belušić, & Smith-Miles, 2014, 2015; Wang, Smith-Miles, & Hyndman, 2006, and Section 2.7.12), classification (Fulcher, Little, & Jones, 2013; Nanopoulos, Alcock, & Manolopoulos, 2001), anomaly detection (Kang, 2012; Talagala, Hyndman, & Smith-Miles, 2020, and Section 2.2.3), and forecasting (Kang, Hyndman, & Smith-Miles, 2017; Montero-Manso, Athanasopoulos, Hyndman, & Talagala, 2020, see also Section 2.2.5). The choice of features depends on the nature of the data and the application context. The state-of-the-art time series feature representation methods quantify a wide range of time series characteristics, including simple summary statistics, stationarity (Montero-Manso et al., 2020; Wang, Kang, Petropoulos, & Li, 2021), model fits (Christ, Braun, Neuffer, & Kempa-Liehr, 2018; Fulcher & Jones, 2014), time series imaging (Li, Kang & Li, 2020), and others. In the forecasting community, two lines of forecasting approaches have been developed using time series features, namely feature-based model selection and feature-based model combination. The motivation behind them is no single model always performs the best for all time series. Instead of choosing one model for all the data, features can be used to obtain the most appropriate model or the optimal combination of candidate models, per series.

As early as in 1972, Reid (1972) argues that time series characteristics provide valuable information in forecast model selection, which is further echoed by Makridakis and Hibon (1979). One way to forecast an extensive collection of time series is to select the most appropriate method per series according to its features. Pioneer studies focus on rule-based methods (for example, Arinze, 1994; Wang, Smith-Miles, & Hyndman, 2009) to recommend the "best" forecasting model per series based on its features. Another line of approaches apply regression to study how useful features are in predicting which forecasting method performs best (for example, Meade, 2000; Petropoulos, Makridakis, Assimakopoulos,

[51] This subsection was written by Yanfei Kang.

& Nikolopoulos, 2014). With the advancement of machine learning (see also Section 2.7.10), more recent literature uses "meta-learning" to describe the process of automatically acquiring knowledge for forecast model selection. The first such study is by Prudêncio and Ludermir (2004), who apply decision trees for forecast model selection. Lemke and Gabrys (2010) compare different meta-learning approaches to investigate which model works best in which situation. Kang et al. (2017) propose using feature spaces to visualise the strengths and weaknesses of different forecasting methods. Other algorithms such as neural networks (see also Section 2.7.8) and random forecasts are also applied to forecast model selection (Kück, Crone, & Freitag, 2016; Talagala, Hyndman, & Athanasopoulos, 2018).

One of the pioneering studies in feature-based forecast combination is the rule-based approach by Collopy and Armstrong (1992), who develop 99 rules for forecast combination based on 18 features. Recently, Kang, Hyndman and Li (2020) use 26 features to predict the performances of nine forecasting methods with nonlinear regression models, and obtain the combination weights as a tailored softmax function of the predicted forecasting errors. The feature-based forecast model averaging (FFORMA) framework proposed by Montero-Manso et al. (2020) employ 42 features to estimate the optimal combination weights of nine forecasting methods based on extreme gradient boosting (XGBoost, Chen & Guestrin, 2016). Li, Kang et al. (2020) first transform time series into images, and use features extracted from images to estimate the optimal combination weights. For feature-based interval forecasting, Wang, Kang et al. (2021) investigate how time series features affect the relative performances of prediction intervals from different methods, and propose a general feature-based interval forecasting framework to provide reliable forecasts and their uncertainty estimation.

2.7.5. Forecasting with bootstrap[52]

The bootstrap methodology has been widely applied in many areas of research, including time series analysis. The bootstrap procedure (Efron, 1979) is a very popular methodology for independent data because of its simplicity and nice properties. It is a computer-intensive method that presents solutions in situations where the traditional methods fail or are very difficult to apply. However, Efron's bootstrap (iid bootstrap) has revealed itself inefficient in the context of dependent data, such as in the case of time series, where the dependence structure arrangement has to be kept during the resampling scheme.

Most of the resampling for dependent data consider segments of the data to define blocks, such that the dependence structure within each block can be kept. Different versions of blocking differ in the way as blocks are constructed: the nonoverlapping block bootstrap (Carlstein, 1990), the moving block bootstrap (Künsch, 1989), the circular block bootstrap (Politis & Romano, 1992), and the stationary block bootstrap (Politis & Romano, 1994). But, if the time series process is driven from iid innovations, another way of resampling can be used.

The iid Bootstrap can then be easily extended to a dependent setup. That was the spirit of sieve bootstrap proposed by Bühlmann (1997). This method is based on the idea of fitting parametric models first and resampling from the residuals. Such models include, for example, the linear regression (Freedman, 1981) and autoregressive time series (Efron & Tibshirani, 1986). This approach is different from the previous bootstrap methods for dependent data; the sample bootstrap is (conditionally) stationary and does not present a structure of dependence. Another different feature is that the sieve bootstrap sample is not a subsample from the original data, as in the previous methods. Observe that even if the sieve bootstrap is based on a parametric model, it is nonparametric in its spirit. The AR model (see Section 2.3.4) here is just used to filter the residuals series.

A few years ago, the sieve bootstrap was used for estimating forecast intervals (Andrés, Peña, & Romo, 2002; Zagdański, 2001). Motivated by these works, Cordeiro and Neves (2006, 2009, 2010) developed a procedure to estimate point forecasts. The idea of these authors was to fit an exponential smoothing model (see Section 2.3.1) to the time series, extract the residuals and then apply the sieve bootstrap to the residuals. Further developments of this procedure include the estimation of forecast intervals (Cordeiro & Neves, 2014) and also the detection, estimation and imputation of missing data (Cordeiro & Neves, 2013). In a recent work (Bergmeir et al., 2016), a similar approach was also consider, the residuals were extracted and resampled using moving block bootstrap (see Section 2.7.6 for further discussion).

Bickel and Freedman (1981) and later in Angus (1992) showed that in extreme value theory, the bootstrap version for the maximum (or minimum) does not converge to the extremal limit laws. Zelterman (1993) pointed out that "to resample the data for approximating the distribution of the k largest observations would not work because the 'pseudo-samples' would never have values greater than $X_{n:n}$"[53]. A method considering to resample a smaller size than the original sample was proposed in Hall (1990). Recently, Neves and Cordeiro (2020) used this idea and developed a preliminary work in modelling and forecasting extremes in time series.

2.7.6. Bagging for time series forecasting[54]

The term *bagging* was proposed by Breiman (1996) to describe the generation of several versions of a predictor, via Bootstrap procedures (introduced in Section 2.7.5), with a final aggregation stage. Thus, "**b**ootstrap **agg**regat**ing**" was established as bagging. The main idea is to improve predictors' accuracy once the data sets, draw randomly with replacement, will approximating the original distribution. The author argues that bagging works for unstable procedures, but it was not tested for time series. Years after, Kilian and Inoue (2004) suggested

52 This subsection was written by Clara Cordeiro.

53 $\max(X_1, \ldots, X_n)$.

54 This subsection was written by Fernando Luiz Cyrino Oliveira.

the first attempts for temporal dependent data. For data-driven methods, to forecasting and simulation time series and deal with predictors ensembles, bagging has shown as a powerful tool.

A general framework for ensemble forecasting methods involves four main stages: (*i*) data treatment, (*ii*) resampling, (*iii*) forecasting, and (*iv*) aggregation. However, for time series, bootstrap should be done carefully, as the serial dependence and non-stationarity must be considered.

As mentioned in Section 2.7.5, this led Bergmeir et al. (2016) to propose a bagging version for exponential smoothing, the Bagged ETS. As pre-treatment, after a Box–Cox transformation, the series is decomposed into trend, seasonal, and remainder components via STL decomposition (Cleveland et al., 1990). The resampling stage uses moving block bootstrap (MBB: Lahiri & Lahiri, 2003), applied to the remainder. There are several discussions in the literature about this procedure, mainly regarding the size of the blocks. MBB resampling the collection of overlapping (consecutive) blocks of observations. The idea is to keep the structure still present in the remainder. The forecasts are obtained via ETS methods (see Section 2.3.1) and, for the final aggregation, the authors adopted the median. Their method is evaluated on the M3 data set and outperformed the original benchmarks. The work of Bergmeir et al. (2016) inspired many others: Dantas, Cyrino Oliveira, and Varela Repolho (2017) applied the idea for air transport demand data and de Oliveira and Cyrino Oliveira (2018) for energy consumption, proposing the so-called remainder sieve bootstrap (RSB).

Dantas and Cyrino Oliveira (2018) proposed an extension to the Bagged ETS where bagging and exponential smoothing are combined with clustering methods (clustering-based forecasting methods are discussed in Section 2.7.12). The approach aims to consider and reduce the covariance effects among the ensemble time series, creating clusters of similar forecasts – since it could impact the variance of the group. A variety of forecasts are selected from each cluster, producing groups with reduced variance.

In light of the aforementioned, there are several possibilities for each stage of the mentioned framework. In this context, to investigate the reasons why bagging works well for time series forecasting, Petropoulos et al. (2018a) explored three sources of uncertainty: model form, data, and parameter. While arguably bagging can handle all of them, Petropoulos et al. (2018a) showed that simply tackling model uncertainty is enough for achieving a superior performance, leading to the proposal of a Bootstrap Model Combination (BMC) approach, where different model forms are identified in the ensemble and fitted to the original data.

Finally, Meira, Cyrino Oliveira, and Jeon (2020) proposed "treating and pruning" strategies to improve the performance of prediction intervals for both model selection and forecast combinations. Testing over a large set of real time series from the M forecasting competitions (see also Section 2.12.7), their results highlighted the importance of analysing the prediction intervals of the ensemble series before the final aggregation.

2.7.7. Multi-step ahead forecasting[55]

Given a univariate time series comprising n observations, $y_1, y_2, \ldots, y_n$, multi-step ahead point forecasting involves producing point estimates of the H future values $y_{n+1}, y_{n+2}, \ldots, y_{n+H}$, where $H > 1$, is the forecast horizon (Ben Taieb, 2014).

The (naive) recursive strategy estimates a one-step-ahead autoregressive model to predict y_{t+1} from $y_t, y_{t-1}, \ldots$, by minimising the one-step-ahead forecast errors. Each forecast is then obtained dynamically by iterating the model H times, and by plugging in the missing lagged values with their respective forecasts. The direct strategy builds separate h-step-ahead models to predict y_{t+h} from $y_t, y_{t-1}, \ldots$ for $h = 1, 2, \ldots, H$, by minimising h-step-ahead forecast errors, and forecasts are computed directly by the estimated models.

In theory, with linear models, model misspecification plays an important role in the relative performance between the recursive and direct strategy (Chevillon, 2007). If the model is correctly specified, the recursive strategy benefits from more efficient parameter estimates, while the direct strategy is more robust to model misspecification. With nonlinear models, recursive forecasts are known to be asymptotically biased (Fan & Yao, 2005; Lin & Granger, 1994; Teräsvirta, Tjostheim, & Granger, 2010), and the direct strategy is often preferred over the recursive strategy since it avoids the accumulation of forecast errors. In practice, the results are mixed (Atiya, El-shoura, Shaheen, & El-sherif, 1999; Kline, 2004; Marcellino, Stock, & Watson, 2006; Pesaran, Pick, & Timmermann, 2011; Sorjamaa, Hao, Reyhani, Ji, & Lendasse, 2007), and depend on many interacting factors including the model complexity (see also Section 2.5.2), the (unknown) underlying data generating process, the number of observations, and the forecast horizon (see also Section 2.7.4).

Hybrids and variants of both recursive and direct strategies have been proposed in the literature. For example, one of the hybrid strategies (Sorjamaa & Lendasse, 2006; Zhang & Hutchinson, 1994; Zhang, Zhou, Chang, Yang, & Li, 2013) first produce recursive forecasts, then adjust these forecasts by modelling the multi-step forecast errors using a direct strategy (Ben Taieb & Hyndman, 2014). Variants of the recursive strategy match the model estimation and forecasting loss functions by minimising the implied h-step-ahead recursive forecast errors (Bhansali & Kokoszka, 2002; Bontempi, Birattari, & Bersini, 1999; McNames, 1998). Variants of the direct strategy exploit the fact that the errors of different models are serially correlated (Chen, Yang, & Hafner, 2004; Franses & Legerstee, 2009c; Lee & Billings, 2003; Pesaran et al., 2011). The idea is to reduce the forecast variance of independently selected models by exploiting the relatedness between the forecasting tasks, as in multi-task learning (Caruana, 1997). For example, a multi-horizon strategy will measure forecast accuracy (see Section 2.12.2) by averaging the forecast errors over multiple forecast horizons (Bontempi & Ben Taieb, 2011; Kline, 2004). Different multi-horizon strategies can be specified, with different formulation of the objective function (Ben Taieb, Sorjamaa, & Bontempi,

[55] This subsection was written by Souhaib Ben Taieb.

2010). One particular case is the multi-output strategy which estimates a single model for all horizons by minimising the average forecast error over the entire forecast horizon (Bontempi & Ben Taieb, 2011).

Forecasting strategies are often model-dependent, especially with machine learning models (see Section 2.7.10). Furthermore, model architecture and parameters are often trained by taking into account the chosen forecasting strategy. For example, we can naturally implement and train recursive forecasting models using recurrent neural networks (see also Section 2.7.8). Also, different specifications of the decoder in sequence-to-sequence models will induce different forecasting strategies, including variants of direct and multi-horizon strategies. For more details, we refer the reader to Hewamalage, Bergmeir, and Bandara (2021) and Section 4.2 in Benidis et al. (2020).

Which forecasting strategy is best is an empirical question since it involves a tradeoff between forecast bias and variance (Ben Taieb & Atiya, 2015; Ben Taieb, Bontempi, Atiya, & Sorjamaa, 2012). Therefore, the forecasting strategy should be part of the design choices and the model selection procedure of any multi-step-ahead forecasting model.

2.7.8. Neural networks[56]

Neural Networks (NNs) or Artificial Neural Networks (ANNs) are mathematical formulations inspired by the work and functioning of biological neurons. They are characterized by their ability to model non-stationary, nonlinear and high complex datasets. This property along with the increased computational power have put NNs in the frontline of research in most fields of science (De Gooijer & Hyndman, 2006; Zhang, Eddy Patuwo, & Y. Hu, 1998).

A typical NN topology is consisted by three types of layers (input, hidden and output) and each layer is consisted by nodes. The first layer in every NN, is the input layer and the number of its nodes corresponds to the number of explanatory variables (inputs). The last layer is the output layer and the number of nodes corresponds to the number of response variables (forecasts). Between the input and the output layer, there is one or more hidden layers where the nodes define the amount of complexity the model is capable of fitting. Most NN topologies in the input and the first hidden layer contain an extra node, called the bias node. The bias node has a fixed value of one and serves a function similar to the intercept in traditional regression models. Each node in one layer has connections (weights) with all or a subset (for example, for the convolutional neural network topology) of the nodes of the next layer.

NNs process the information as follows: the input nodes contain the explanatory variables. These variables are weighted by the connections between the input and the first hidden nodes, and the information reaches to the hidden nodes as a weighted sum of the inputs. In the hidden nodes, there is usually a non-linear function (such as the sigmoid or the RelU) which transform the information received. This process is repeated until the information reaches the output layer as forecasts. NNs are trained by adjusting the weights that connect the nodes in a way that the network maps the input value of the training data to the corresponding output value. This mapping is based on a loss function, the choice of which depends on the nature of the forecasting problem. The most common NN procedure, is the back-propagation of errors (for additional details on training, see Section 2.7.11).

The simpler and most common NN topology, is the Multilayer Forward Perceptron (MLP). In MLP, the hidden nodes contain the sigmoid function and the information moves in forward direction (from the inputs to the output nodes). An another NN topology where the information moves also only in a forward direction is the Radial Basis Function NN (RBF). Now the hidden neurons compute the Euclidean distance of the test case from the neuron's centre point and then applies the Gaussian function to this distance using the spread values. Recurrent Neural Networks (RNNs) are NN topologies that allow previous outputs to be used as inputs while having hidden states. The information moves both forwards and backwards. RNNs have short-term memory and inputs are taken potentially from all previous values. MLPs, RBFs and RNNs are universal function approximators (Hornik, 1991; Park & Sandberg, 1991; Schäfer & Zimmermann, 2006). However, the amount of NN complexity in terms of hidden layers and nodes to reach this property, might make the NN topology computationally unfeasible to train (see also the discussion in Section 2.7.11). For the interaction of NNs with the probability theory, we refer the reader to last part of Section 2.7.9.

2.7.9. Deep probabilistic forecasting models[57]

Neural networks (Section 2.7.8) can be equipped to provide not only a single-valued forecast, but rather the entire range of values possible in a number of ways (see also Sections 2.6.2 and 2.6.3). We will discuss three selected approaches in the following, but remark that this is a subjective selection and is by far not comprehensive.[58]

1. Analogously to linear regression and Generalised Linear Models, obtaining probabilistic forecasts can be achieved by the neural network outputting not the forecasted value itself but rather parameters of probability distribution or density (Bishop, 2006). In forecasting, a prominent example is the DeepAR model (Salinas, Valentin, Jan & Januschowski, 2019), which uses a recurrent neural network architecture and assumes the probability distribution to be from a standard probability density function (e.g., negative binomial or Student's t). Variations are possible, with either non-standard output distributions in forecasting such as the multinomial distribution (Rabanser, Januschowski, Flunkert, Salinas, & Gasthaus, 2020) or via representing the probability

[56] This subsection was written by Georgios Sermpinis.

[57] This subsection was written by Tim Januschowski.

[58] For example, contemporary topics in machine learning such as generative adversarial networks can be naturally lifted to forecasting and similarly, more traditional probabilistic machine learning approaches such as Gaussian Processes (Maddix, Wang, & Smola, 2018). We ignore the important area of Bayesian deep learning (see Wang & Yeung, 2016, for a survey) entirely here for lack of space.

density as cumulative distribution function (Salinas, Michael et al., 2019) or the quantile function (Gasthaus et al., 2019).
2. An alternative approach is to apply concepts for quantile regression (Koenker, 2005) to neural networks, e.g., by making the neural network produce values for selected quantiles directly (Wen, Torkkola, Narayanaswamy, & Madeka, 2017).
3. It is possible to combine neural networks with existing probabilistic models. For example, neural networks can parametrise state space models (Durbin & Koopman, 2012) as an example for another class of approaches (Rangapuram, Seeger, Gasthaus, Stella, Wang, & Januschowski, 2018), dynamic factor models (Geweke, 1977) with neural networks (Wang, Smola, Maddix, Gasthaus, Foster, & Januschowski, 2019) or deep temporal point processes (Turkmen, Wang, & Januschowski, 2019).

The appeals of using neural networks for point forecasts carry over to probabilistic forecasts, so we will only comment on the elegance of modern neural network programming frameworks. To the forecasting model builder, the availability of auto gradient computation, the integration with highly-tuned optimisation algorithms and scalability considerations built into the frameworks, means that the time from model idea to experimental evaluation has never been shorter. In the examples above, we brushed over the need to have loss functions with which we estimate the parameters of the neural networks. Standard negative log-likelihood based approaches are easily expressible in code as are approaches based on non-standard losses such as the continuous ranked probability score (Gneiting et al., 2007, and Section 2.12.4). With open source proliferating in the deep learning community, most of the above examples for obtaining probabilistic forecasts can readily be test-driven (see, for example, Alexandrov et al., 2019).

For the future, we see a number of open challenges. Most of the approaches mentioned above are univariate, in the following sense. If we are interested in forecasting values for all time series in a panel, we may be interested in modelling the relationship among these time series. The aforementioned approaches mostly assume independence of the time series. In recent years, a number of multivariate probabilistic forecasting models have been proposed (Rangapuram, de Bezenac, Benidis, Stella, & Januschowski, 2020; Salinas, Michael et al., 2019), but much work remains to obtain a better understanding. Another counter-intuitive challenge for neural networks is to scale them down. Neural networks are highly parametrised, so in order to estimate parameters correctly, panels with lots of time series are needed. However, a large part of the forecasting problem landscape (Januschowski & Kolassa, 2019) consists of forecasting problems with only a few time series. Obtaining good uncertainty estimates with neural networks in these settings is an open problem.

2.7.10. Machine learning[59]

Categorising forecasting methods into statistical and machine learning (ML) is not trivial, as various criteria can be considered for performing this task (Januschowski et al., 2020). Nevertheless, more often than not, forecasting methods are categorised as ML when they do not prescribe the data-generating process, e.g., through a set of equations, thus allowing for data relationships to be automatically learned (Barker, 2020). In this respect, methods that build on unstructured, non-linear regression algorithms (see also Section 2.3.2), such as Neural Networks (NN), Decision Trees, Support Vector Machines (SVM), and Gaussian Processes, are considered as ML (Makridakis et al., 2018).

Since ML methods are data-driven, they are more generic and easier to be adapted to forecast series of different characteristics (Spiliotis, Kouloumos, Assimakopoulos & Makridakis, 2020). However, ML methods also display some limitations. First, in order for ML methods to take full advantage of their capacity, sufficient data are required. Thus, when series are relatively short and display complex patterns, such as seasonality and trend, ML methods are expected to provide sub-optimal forecasts if the data are not properly pre-processed (Makridakis et al., 2018; Zhang et al., 1998). On the other hand, when dealing with long, high-frequency series, typically found in energy (Chae, Horesh, Hwang, & Lee, 2016, but also Section 3.4), stock market (Moghaddam, Moghaddam, & Esfandyari, 2016, and Section 3.3), and demand (Carmo & Rodrigues, 2004, but also Section 3.2) related applications, ML methods can be applied with success. Second, computational intensity may become relevant (Makridakis, Spiliotis et al., 2020), especially when forecasting numerous series at the weekly and daily frequency (Seaman, 2018) or long-term accuracy improvements over traditional methods are insignificant (Nikolopoulos & Petropoulos, 2018). Third, given that the effective implementation of ML methods strongly depends on optimally determining the values of several hyper-parameters, related both with the forecasting method itself and the training process, considerable complexity is introduced, significant resources are required to set up the methods, and high experience and a strong background in other fields than forecasting, such as programming and optimisation, are needed.

In order to deal with these limitations, ML methods can be applied in a cross-learning (CL) fashion instead of a series-by-series one (Makridakis, Spiliotis et al., 2020), i.e., allow the methods to learn from multiple series how to accurately forecast the individual ones (see also Section 2.12.7). The key principle behind CL is that, although series may differ, common patterns may occur among them, especially when data are structured in a hierarchical way and additional information, such as categorical attributes and exogenous/explanatory variables (see Section 2.2.5), is provided as input (Fry & Brundage, 2020). The CL approach has several advantages. First, computational time can be significantly reduced

[59] This subsection was written by Evangelos Spiliotis.

as a single model can be used to forecast multiple series simultaneously (Semenoglou, Spiliotis, Makridakis, & Assimakopoulos, 2021). Second, methods trained in a particular dataset can be effectively used to provide forecasts for series of different datasets that display similar characteristics (transfer-learning), thus allowing the development of generalised forecasting methods (Oreshkin, Carpov, Chapados, & Bengio, 2020a). Third, data limitations are mitigated and valuable information can be exploited at global level, thus allowing for patterns shared among the series, such as seasonal cycles (Dekker, van Donselaar, & Ouwehand, 2004) and special events (Huber & Stuckenschmidt, 2020), to be effectively captured.

Based on the above, CL is currently considered the most effective way of applying ML for batch time series forecasting. Some state-of-the-art implementations of CL include long short-term memory NNs (Smyl, 2020), deep NNs based on backward and forward residual links (Oreshkin, Carpov, Chapados, & Bengio, 2020b), feature-based XGBoost (Montero-Manso et al., 2020), and gradient boosted decision trees (Bojer & Meldgaard, 2020).

2.7.11. Machine learning with (very) noisy data[60]

With the advent of big data, machine learning now plays a leading role in forecasting.[61] There are two primary reasons for this. First, conventional ordinary least squares (OLS) estimation is highly susceptible to *overfitting* in the presence of a large number of regressors (or features); see also Sections 2.5.2 and 2.5.3. OLS maximises the fit of the model over the estimation (or training) sample, which can lead to poor out-of-sample performance; in essence, OLS over-responds to *noise* in the data, and the problem becomes magnified as the number of features grows. A class of machine-learning techniques, which includes the popular least absolute shrinkage and selection operator (LASSO: Tibshirani, 1996) and elastic net (ENet: Zou & Hastie, 2005), employs *penalised regression* to improve out-of-sample performance with large numbers of features. The LASSO and ENet guard against overfitting by *shrinking* the parameter estimates toward zero.

Very noisy data – data with a very low signal-to-noise ratio – exacerbate the overfitting problem. In such an environment, it is vital to induce adequate shrinkage to guard against overfitting and more reliably uncover the predictive signal amidst all the noise. For LASSO and ENet estimation, a promising strategy is to employ a stringent information criterion, such as the Bayesian information criterion (BIC, Schwarz, 1978), to select (or tune) the regularisation parameter governing the degree of shrinkage (often denoted by λ). Fan and Tang (2013) and Wang, Li, and Leng (2009) modify the BIC penalty to account for a diverging number of features, while Hui, Warton, and Foster (2015) refine the BIC penalty to include the value of λ. These BIC variants induce a greater degree of shrinkage by strengthening the BIC's penalty term, making them useful for implementing the LASSO and ENet in noisy data environments; see Filippou, Rapach, Taylor, and Zhou (2020) for a recent empirical application.

A second reason for the popularity of machine learning in the era of big data is the existence of powerful tools for accommodating complex predictive relationships. In many contexts, a linear specification appears overly restrictive, as it may neglect important nonlinearities in the data that can potentially be exploited to improve forecasting performance. Neural networks (NNs; see Section 2.7.8) are perhaps the most popular machine-learning device for modelling nonlinear predictive relationships with a large number of features. Under a reasonable set of assumptions, a sufficiently complex NN can approximate any smooth function (for example, Barron, 1994; Cybenko, 1989; Funahashi, 1989; Hornik, Stinchcombe, & White, 1989).

By design, NNs are extremely flexible, and this flexibility means that a large number of parameters (or weights) need to be estimated, which again raises concerns about overfitting, especially with very noisy data. The weights of a NN are typically estimated via a stochastic gradient descent (SGD) algorithm, such as *Adam* (Kingma & Ba, 2015). The SGD algorithm itself has some regularising properties, which can be strengthened by adjusting the algorithm's hyperparameters. We can further guard against overfitting by shrinking the weights via LASSO or ENet penalty terms, as well as imposing a dropout rate (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012; Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).

Perhaps the quintessential example of a noisy data environment is forecasting asset returns, especially at short horizons (e.g., monthly). Because many asset markets are reasonably efficient, most of the fluctuations in returns are inherently unpredictable – they reflect the arrival of new information, which, by definition, is unpredictable. This does not mean that we should not bother trying to forecast returns, as even a seemingly small degree of return predictability can be economically significant (e.g., Campbell & Thompson, 2008). Instead, it means that we need to be particularly mindful of overfitting when forecasting returns in the era of big data. Section 3.3.13 discusses applications of machine-learning techniques for stock return forecasting.

2.7.12. Clustering-based forecasting[62]

The robustness of the forecasting process depends mainly on the characteristics of the target variable. In cases of high nonlinear and volatile time series, a forecasting model may not be able to fully capture and simulate the special characteristics, a fact that may lead to poor forecasting accuracy (Pradeepkumar & Ravi, 2017). Contemporary research has proposed some approaches to increase the forecasting performance (Sardinha-Lourenço, Andrade-Campos, Antunes, & Oliveira, 2018). Clustering-based forecasting refers to the application of unsupervised machine learning in forecasting tasks. The scope is to increase the performance by employing the information of data structure and of the existing similarities among

[60] This subsection was written by David E. Rapach.

[61] We do not make a sharp distinction between statistical learning and machine learning. For brevity, we use the latter throughout this subsection.

[62] This subsection was written by Ioannis Panapakidis.

the data entries (Goia, May, & Fusai, 2010); see also Sections 2.7.4 and 2.7.10. Clustering is a proven method in pattern recognition and data science for deriving the level of similarity of data points within a set. The outputs of a clustering algorithm are the centroids of the clusters and the cluster labels, i.e., integer numbers that denote the number of cluster that a specific data entry belongs to (Xu & Wunsch, 2005).

There are two approaches in clustering-based forecasting: (*i*) Combination of clustering and supervised machine learning, and (*ii*) solely application of clustering. In the first case, a clustering algorithm is used to split the training set into smaller sub-training sets. These sets contain patterns with high similarity. Then for each cluster a dedicated forecaster is applied (Chaouch, 2014; Fan, Chen, & Lee, 2008). Thus, the number of forecasting algorithms is equal to the number of clusters. This approach enables to train forecasters with more similar patterns and eventually achieve better training process. The forecasting systems that involve clustering are reported to result in lower errors (Fan, Mao, & Chen, 2006; Mori & Yuihara, 2001). The combination of clustering and forecasting has been presented in the literature earlier than the sole application of clustering. One of the first articles in the literature on combining clustering and forecasting sets up the respective theoretical framework (Kehagias & Petridis, 1997).

In the second case, a clustering algorithm is used to both cluster the load data set and perform the forecasting (López, Valero, Senabre, Aparicio, & Gabaldon, 2012). In the sole clustering applications, either the centroids of the clusters can be utilised or the labels. Pattern sequence-based forecasting is an approach that employs the cluster labels. In this approach, a clustering of all days prior to the test day is held and this results in sequences of labels of a certain length. Next, the similarity of the predicted day sequence with the historical data sequences is examined. The load curve of the predicted day is the average of the curves of the days following the same sequences (Kang, Spiliotis, Petropoulos, Athiniotis, Li & Assimakopoulos, 2020; Martinez Alvarez, Troncoso, Riquelme, & Aguilar Ruiz, 2011).

There is variety of clustering algorithms that have been proposed in forecasting such as the *k*-means, fuzzy C-means (FCM), self-organising map (SOM), deterministic annealing (DA), ant colony clustering ACC, and others. Apart from the clustering effectiveness, a selection criterion for an algorithm is the complexity. *k*-means and FCM are less complex compared to the SOM that needs a large number of variables to be calibrated prior to its application. Meta-heuristics algorithms, such as ACC, strongly depend on the initialisation conditions and the swarm size. Therefore, a comparison of clustering algorithms should take place to define the most suitable one for the problem under study (Elangasinghe, Singhal, Dirks, Salmond, & Samarasinghe, 2014; Li, Han, & Li, 2008; Mori & Yuihara, 2001; Wang, Pedrycz, & Liu, 2015).

The assessment of clustering-based forecasting is held via common evaluation metrics for forecasting tasks (see also Section 2.12). The optimal number of clusters, which is a crucial parameter of a clustering application, is selected via trial-and-error, i.e., the optimal number corresponds to the lowest forecasting error (Nagi, Yap, Nagi, Tiong, & Ahmed, 2011).

2.7.13. Hybrid methods[63]

Hybrid approaches combine two or more of the above-mentioned advanced methods. In general, when methods based on AI-based techniques, physical, and statistical approaches are combined together, the result is often improved forecasting accuracy as a benefit from the inherent integration of the single methods. The idea is to mix diverse methods with unique features to address the limitations of individual techniques, thus enhancing the forecast performance (Mandal, Madhira, Haque, Meng, & Pineda, 2012; Nespoli, Ogliari, Leva, Massi Pavan, Mellit, Lughi, & Dolara, 2019, see also Section 2.6). The performance of the hybrid methods depends on the performance of the single methods, and these single methods should be specifically selected for the problem that has to be addressed.

Hybrid methods can be categorised based on the constituent methods, but also considering that these base methods may not necessarily act only on the forecasting stage but also on data treatment and parameters identification stages. In data pre-processing combining approaches (see also Section 2.2), different methods can be used for decomposing the time series into subseries (Son, Yang, & Na, 2019) or the signal into different frequencies (Zang, Cheng, Ding, Cheung, Liang, Wei, & Sun, 2018), and for classifying the historical data (Huang, Chen, Yang & Kuo, 2015). An advantage of such hybrid methods is robustness against sudden changes in the values of the main parameters. However, they require additional knowledge and understanding of the base methods, and have the disadvantage of slow response time to new data.

The purpose of the parameter selection stage is to optimise the parameters of the model, in terms of extracting nonlinear features and invariant structures (Behera, Majumder, & Nayak, 2018; Ogliari, Niccolai, Leva, & Zich, 2018) but also in terms of estimation of the parameter adopted for the prediction; for example, meteorological factors such as temperature, humidity, precipitation, snowfall, cloud, sunshine, wind speed, and wind direction (Qu, Kang, Zhang, Jiang, & Ma, 2016). Hybrid methods feature straightforward determination of the parameters with relatively basic structures. However, the implementation is sometimes challenging, and depends on the knowledge and expertise of the designer.

Finally, the data post-processing hybrid approaches forecast the residual errors resulted from the forecasting model. Since these hybrid methods consider residual errors from the model, they aim in further improving the predictions of the base methods by applying corrections in the forecasts. However, a disadvantage of these hybrid methods is the increased calculation time, as the residual errors must also be estimated. Also, such hybrid methods are not general and will depend on the field of application. In many cases, hybrids approaches outperform other (single) approaches such as *k*NN, NN, and ARIMA-based models (Mellit, Massi Pavan, Ogliari, Leva, & Lughi, 2020). A great example is the hybrid method by Smyl (2020), which achieved the best performance in the M4 forecasting competition (see also Section 2.12.7). In particular, in energy applications (see Section 3.4), a

[63] This subsection was written by Sonia Leva.

combination of physical and AI-based techniques can lead to improved forecasting performance. Furthermore, machine learning methods (see Section 2.7.10) based on historical data of meteorological variables combined with an optimal learning algorithm and weather classification can further improve the forecasting accuracy of single methods. However, in general, the weak point of such hybrid approaches is that they underperform when meteorological conditions are unstable (Chicco, Cocina, Di Leo, Spertino, & Massi Pavan, 2015).

2.8. Methods for intermittent demand

2.8.1. Parametric methods for intermittent demand forecasting[64]

Demand forecasting is the basis for most planning and control activities in any organisation. Demand will typically be accumulated in some pre-defined 'time buckets' (periods), such as a day, a week or a month. On many occasions, demand may be observed in every time period, resulting in what is sometimes referred to as 'non-intermittent demand'. Alternatively, demand may appear sporadically, with no demand at all in some periods, leading to an intermittent appearance of demand occurrences. Intermittent demand items monopolise the stock bases in the after sales industry and are prevalent in many other industries, including the automotive, IT, and electronics sectors. Their inventory implications are dramatic and forecasting their requirements is a very challenging task.

Methods to forecast intermittent demand may broadly be classified as parametric and non-parametric. The former suppose that future demand can be well represented by a statistical distribution (say Poisson or Negative Binomial) which has parameters that are unknown but may be forecasted using past data. These methods are discussed in this sub-section. In the latter, the data are not assumed to follow any standard probability distribution. Instead, direct methods are used to assess the distributions required for inventory management (see also Section 3.2.3). Such methods are discussed in Section 2.8.2.

Simple Exponential Smoothing (SES; see Section 2.3.1) is often used in practice to forecast intermittent demand series. However, SES fails to recognise that intermittent demand is built from two constituent elements: (*i*) the inter-demand intervals, which relate to the probability of demand occurring, and (*ii*) the demand sizes, when demand occurs. The former indicates the degree of intermittence, whereas the latter relates to the behaviour of the positive demands. Croston (1972) showed that this inherent limitation leads to SES being (positively) biased after a demand occurring period; this is sometimes referred to as an 'issue point' bias. Subsequently, he proposed a method that forecasts separately the sizes of demand, when demand occurs, and the inter-demand intervals. Both forecasts are produced using SES, and the ratio of the former over the latter gives a forecast of the mean demand per period. Croston's method was shown by Syntetos and Boylan (2001) to suffer from another type of bias (inversion bias) and the same researchers (Syntetos & Boylan, 2005) proposed a modification to his method that leads to approximately unbiased estimates. This method is known in the literature as the Syntetos-Boylan Approximation (SBA). It has been found repeatedly to account for considerable empirical inventory forecasting improvements (Eaves & Kingsman, 2004; Gutierrez, Solis, & Mukhopadhyay, 2008; Nikolopoulos, Babai, & Bozos, 2016; van Wingerden, Basten, Dekker, & Rustenburg, 2014) and, at the time of writing, it constitutes the benchmark against which other (new) proposed methodologies in the area of intermittent demand forecasting are assessed.

Croston's method is based upon the assumption of a Bernoulli demand arrival process. Alternatively, demand may be assumed to arrive according to a Poisson process. It is also possible to adapt Croston's method so that sizes and intervals are updated based on a simple moving average (SMA) procedure instead of SES. Boylan and Syntetos (2003), Shale, Boylan, and Johnston (2006), and Syntetos, Babai, and Luo (2015) presented correction factors to overcome the bias associated with Croston's approach under a Poisson demand arrival process and/or estimation of demand sizes and intervals using an SMA.

For a detailed review of developments in intermittent demand forecasting interested readers are referred to Boylan and Syntetos (2021).

2.8.2. Non-parametric intermittent demand methods[65]

Two main non-parametric forecasting approaches have dominated the intermittent demand literature: the bootstrapping approach and the Overlapping/Non-Overlapping aggregation Blocks approach (Boylan & Syntetos, 2021).

The bootstrapping approach relies upon a resampling (with or without replacement) of the historical demand data to build the empirical distribution of the demand over a specified interval. As discussed in Section 2.7.5, this approach was initially introduced by Efron (1979). Since then, it has been developed by Willemain, Smart, and Schwarz (2004) and Zhou and Viswanathan (2011) to deal with intermittent demand items (Babai, Tsadiras, & Papadopoulos, 2020). Willemain et al. (2004) have proposed a method that resamples demand data by using a Markov chain to switch between no demand and demand periods. The empirical outperformance of this method has been shown when compared to Simple Exponential Smoothing (SES) and Croston's method (see also Section 2.8.1). However, the findings of Willemain et al. (2004)'s work have been challenged by Gardner and Koehler (2005) and some limitations have been addressed by Syntetos, Zied Babai and Gardner (2015). Zhou and Viswanathan (2011) have developed an alternative bootstrapping method. Their method samples separately demand intervals and demand sizes and it has been shown to be associated with a good performance for long lead-times. Teunter and Duncan (2009) and Hasni, Aguir, Babai, and Jemai (2019a) have developed adjustments of the bootstrapping methods, where the lead-time demand forecast is adjusted by assuming that the first period in the lead-time bucket

[64] This subsection was written by Aris A. Syntetos.

[65] This subsection was written by Mohamed Zied Babai.

corresponds to a non-zero demand. They have demonstrated the outperformance of the adjusted bootstrapping methods in a periodic order-up-to-level inventory control system. A review of the bootstrapping methods in the context of intermittent demand is provided by Hasni, Aguir, Babai, and Jemai (2019b).

Porras and Dekker (2008) were the first to consider aggregation with overlapping and non-overlapping blocks (OB and NOB) approach in forecasting the demand of spare parts. In the NOB approach, a demand series is divided into consecutive non-overlapping blocks of time, whereas in OB, at each period the oldest observation is dropped and the newest is included (Rostami-Tabar, Babai, Syntetos, & Ducq, 2013). Boylan and Babai (2016) have compared the statistical and inventory performance of the OB and NOB methods. They found that, unless the demand history is short, there is a clear advantage of using OB instead of NOB. More recently, based on extreme value theory (EVT), Zhu, Dekker, van Jaarsveld, Renjie, and Koning (2017) have proposed an improvement of the OB method that models better the tail of lead-time demand. They have shown that the empirical-EVT method leads to higher achieved target cycle service levels when compared to the original method proposed by Porras and Dekker (2008). Temporal aggregation is further discussed in Section 2.10.2.

2.8.3. *Classification methods*[66]

In many application areas, forecasts are required across a wide collection of products, services or locations. In this situation, it is convenient to introduce classification rules that allow subsets of time series to be forecasted using the same approaches and methods. Categorisation rules, such as the ABC inventory classification, serve the forecasting function only coincidentally. They do not necessarily align to the selection of the best forecasting method.

Within certain modelling frameworks, classification of time series is well established. For example, within an ARIMA framework (Box, Jenkins, & Reinsel, 2008, and Section 2.3.4), or within a state-space framework for exponential smoothing (Hyndman et al., 2002, and Section 2.3.1), series may be classified, for example based on the AIC (Akaike, 1973). It is more challenging to classify series according to their recommended forecasting method if some of the candidate methods, such as Croston's method (see Section 2.8.1), lack a fully satisfactory model base. In the field of intermittent demand forecasting, Syntetos, Boylan, and Croston (2005) proposed the SBC classification scheme, enabling time series to be classified according to their length of average demand interval and coefficient of variation of demand sizes (when demand occurs). These rules were based on assumptions of independent and identically distributed (iid) demand, and a comparison of expected mean square error between methods. The scheme has been extended by Kostenko and Hyndman (2006) and by Petropoulos and Kourentzes (2015). In an empirical case-study, Boylan, Syntetos, and Karakostas (2008) examined series not necessarily conforming to iid assumptions and found the rules to be robust to inexact specification of cut-off values. Moon, Simpson, and Hicks (2013) used logistic regression to classify time series of demand for spare parts in the South Korean Navy. The classification was designed to identify superior performance (accuracy and inventory costs) of direct and hierarchical forecasting methods, based on the serial correlation of demands, the coefficient of variation of demand volume of spare parts (see also Section 3.2.7), and the functionality of the naval equipment.

Bartezzaghi, Verganti, and Zotteri (1999) identified five factors that contribute towards intermittence and 'lumpiness' (intermittence with highly variable demand sizes): number of potential customers, frequency of customers' requests, heterogeneity of customers, variety of individual customer's requests, and correlations between customers' requests. These may contribute towards useful classifications, for example by the number of customers for an item. When this number is low and some of the customers are large, then direct communication with the customers can inform judgmental forecasts. Similarly, if a customer's requests are highly variable, then 'advance demand information' from customers can help to improve judgmental estimates. These strategies can be very useful in a business-to-business environment, where such strategies are feasible.

An alternative approach to classification is combination of forecasts (see Section 2.6 for a review, and Section 2.6.1 in particular). Petropoulos and Kourentzes (2015) investigated combining standard forecasting methods for intermittent demand (e.g., SES, Croston, Syntetos-Boylan Approximation; see also Section 2.8.1). They did not find this to improve accuracy directly, but obtained good results from the use of combinations of forecasts at different temporal frequencies, using methods selected from the extended SBC classification scheme.

2.8.4. *Peak over the theshold*[67]

In the forecasting literature, Nikolopoulos (2020) argues that great attention has been given to modelling fast-moving time series with or without cues of information available (Nikolopoulos, Goodwin, Patelis, & Assimakopoulos, 2007). Less attention has been given to intermittent/count series (see Sections 2.3.8, 2.8.1 and 2.8.2), which are more difficult to forecast given the presence of two sources of uncertainty: demand volume, and timing.

Historically there have been few forecasting methods developed specifically for such data (Syntetos, Zied Babai & Gardner, 2015). We believe that through a time series decomposition approach á la Leadbetter (1991) we can isolate 'peaks over threshold' (POT) data points, and create new intermittent series from any time series of interest. The derived series present almost identical characteristics with the series that Croston (1972) analysed. In essence one could use such decomposition forecasting techniques to tackle much more difficult phenomena and problems coming from finance, politics, healthcare, humanitarian logistics, business, economics, and social sciences.

[66] This subsection was written by John E. Boylan.

[67] This subsection was written by Konstantinos Nikolopoulos.

Any time series can be decomposed into two sub-series: one containing the baseline (*white swans*) and one containing the extreme values over an arbitrary-set or rule-based-set threshold (*grey* and *black swans*) as proposed by Taleb (2008); see also Section 2.3.22. Unfortunately, major decision-related risks and most of the underlying uncertainty lay with these extremes. So, it is very important to be able to effectively model and forecast them.

It is unlikely that any forecasting approach will accurately give the exact timing of the forthcoming extreme event, but it will instead provide a satisfactory cumulative forecast over a long period of time. The question still stands what can one do with this forecast? For earthquake data, although even if we know that a major earthquake is going to hit a region, it is almost impossible to decide to evacuate cities, but still we can influence and legislate the way structures are built and increase the awareness, training, preparedness and readiness of the public; and also ensure enough capital on hold to cope with the aftermath of the major event. For epidemics/pandemics (see Section 3.6.2) there are clear implications, as we have evidenced with COVID-19, on how proactively we can source and secure human resources, medical supplies, etc.

What is the current doctrine when forecasting in such a context: advanced probabilistic models. These methods typically require a lot of data and reconstruct the distributions of the underlying phenomena. These come with common successes and a plethora of constraints: *big data sets* needed for training the models, high mathematical *complexity*, and invisibility to practitioners how these methods do actually work and thus *less acceptance in practice*. Yet again, forecasting accuracy is the name of the game and thus these forecasting methods are serious contenders for the task in hand.

Extreme Value Theory (EVT) analyses extreme deviations from statistical measures of central location to estimate the probability of events that are more extreme than anything observed in the time series. This is usually done in the following two ways (Nikolopoulos, 2020): (*i*) deriving maxima and/or minima series as a first step and then having the Generalised Extreme Value Distribution fitted (often the number of extreme events is limited), and (*ii*) isolating the values that exceed a threshold (point over threshold) that can also lead to only a few instances extracted – so a very intermittent series in nature. The analysis involves fitting a Poisson distribution for the number of events in a basic time period and a second distribution – usually a Generalised Pareto Distribution – for the size of the resulting POT values.

2.9. Reasoning and mining

2.9.1. Fuzzy logic[68]

The "classical" Boolean logic is not able to handle for uncertainties and/or vagueness that are necessary when dealing with many real world problems. This is in part due to the fact that the Boolean logic is based on only two values (i.e., a statement can only be true or false). Fuzzy logic tries to overcome this issue by admitting that a statement/variable could be partially true or partially false. Mathematically, the fuzzy logic framework is based on the work of Zadeh (1965) who introduced the theory of *fuzzy sets*. The main point of this theory is the definition of two kinds of sets:

1. *Crisp sets* are the "classical" sets in the Boolean logic. An element can belong (or not) to a certain set.
2. *Fuzzy sets*, where an element can belong to the sets with a certain *membership grade*, with a value that varies in the interval [0, 1].

The definition of the fuzzy sets allows the framework to take into account the uncertainty and vagueness of information. An extension of this approach is related to the fact that a certain variable can assume a crisp value (classical theory) or can belong to different fuzzy sets with different membership grade. For example, in a system implemented to forecast the daily pollutant concentration in atmosphere, one of the inputs could relate to the weather conditions, such as the wind speed. In the classical approach, the system must have as an input the value of the wind speed at a certain day. In the fuzzy approach, the input of the system could be the membership grade of the input variable to three different fuzzy sets: (*i*) "Not windy", (*ii*) "average windy", and (*iii*) "strong windy". On the other hand, the user of the forecasting system may be only interested in a classification of the output variable instead of the crisp value. In this case, the fuzzy approach is applied to the pollutant concentration which could belong with a certain degree to the fuzzy sets (*i*) "not polluted day", (*ii*) "medium polluted day", (*iii*) "high polluted day", and (*iv*) "critical polluted day".

In fuzzy theory, each fuzzy set is characterised by a (generally nonlinear) function, called the *membership function*, linking crisp values to the membership of the different sets. The association of a crisp value to its membership for a set is called *fuzzyfication*, while the inverse operation (from a membership value to a crisp value) is called *defuzzification*. As with the logic theory, the *inference system* assumes a key role in the fuzzy theory. A Fuzzy Inference System (FIS) allows the interpretation of the membership grades of the input variable(s) and, given some sets of fuzzy rules, assigns the corresponding values to the output variable(s). In the literature, two main fuzzy inference systems are presented:

1. Mamdani system (Mamdani & Assilian, 1975), where both the input and output of the inference system are membership functions.
2. Sugeno system (Sugeno, 1985), where the output of the inference system is a crisp value, usually obtained by applying a linear function to the defuzzified value of the input.

2.9.2. Association rule mining[69]

Association rule mining is an exploratory data-driven approach which is able to automatically and exhaustively

[68] This subsection was written by Claudio Carnevale.

[69] This subsection was written by Daniele Apiletti.

extract all existing correlations in a data set of categorical features. It is a powerful but computationally intensive technique, successfully applied in different forecasting contexts (Acquaviva et al., 2015; Apiletti & Pastor, 2020; Di Corso, Cerquitelli, & Apiletti, 2018). Its results are in a human-readable form.

The data set must be in the form of transactions, i.e., a collection of events, each described by categorical features. If the phenomena under analysis are modelled by continuous-valued variables, discretisation can be applied to obtain a suitable data set.

Association rule mining core task is the frequent itemset extraction, which consists in finding frequently-occurring relationships among items in a data set (Han, Pei, & Kamber, 2011). Given a data set of records characterised by several attributes, an item refers to a pair of (attribute = value), while a set of items is called itemset. The support count of an itemset is the number of records r containing that itemset. The support of an itemset is the percentage of records containing it with respect to the total number of records in the data set. An itemset is frequent when its support is greater than or equal to a minimum support threshold.

An association rule is an implication in the form $A \rightarrow B$, where A and B are disjoint itemsets (i.e., $A \cap B = \emptyset$) (Tan, Steinbach, & Kumar, 2005). A is called rule body or antecedent and B rule head or consequent.

To evaluate the quality of an association rule, the support, confidence, and lift metrics are commonly exploited (Han et al., 2011). *Rule support* is the fraction of records containing both A and B, indicating the probability that a record contains every item in these itemsets. The support of the rule is computed as the support of the union of A and B.

Rule confidence represents the strength of the implication, and is the conditional probability that a transaction containing A also contains B, $P(B|A)$, i.e., the proportion of records that contain A with respect to those that also contain B.

Finally, the *lift* of a rule measures the correlation between antecedent and consequent. It is defined as the ratio between the rule $A \rightarrow B$ confidence and the support of B. A lift ratio equal to 1.0 implies that itemsets A and B are not correlated. A lift higher than 1.0 indicates a positive correlation, meaning that the occurrence of A likely leads to the occurrence of B with the given confidence. The greater the lift, the stronger the association. Finally, a lift lower than 1.0 indicates a negative correlation between A and B.

The problem of association rule mining consists in the extraction of all the association rules having rule support and confidence greater than the respective support and confidence thresholds, *MinConf* and *MinSup*, defined as parameters of the mining process (Tan et al., 2005). These thresholds allow to control the statistical relevance of the extracted rules.

The process of rule mining consists of two steps. The first step is the computation of frequent itemsets, i.e., itemsets with support greater or equal to *MinSup*. The second step is the extraction of association rules from frequent itemsets. Let be F a frequent itemset, hence having a support higher than *MinSup*, pairs A and $B = F - A$ are derived so that the confidence of $A \rightarrow B$ is higher than *MinConf*. The first step of the process is the most computationally expensive. Thus, several algorithms have been proposed to solve the problem of frequent itemset extraction (Zaki, 2000), some specifically addressing high-dimensionality issues (Apiletti, Baralis, Cerquitelli, Garza, Michiardi, & Pulvirenti, 2015; Apiletti, Baralis, Cerquitelli, Garza, Pulvirenti, & Michiardi, 2017). Despite being computationally demanding, association rule mining is an exhaustive approach, i.e., all and only statistically relevant correlations are extracted. Section 3.8.11 offers an example of applying association rule mining to forecast the quality of beverages.

2.9.3. *Forecasting with text information*[70]

Text data, such as social media posts, scholar articles and company reports, often contains valuable information that can be used as predictors in forecasting models (Aggarwal & Zhai, 2012). Before extracting useful features from the data, a text document needs to be cleaned and normalised for further processing. The first step of preparing the data is to filter out *stop words* – the words that do not add much meaning to a sentence, e.g., "a", "is", and "me". For grammatical reasons, it is necessary for documents to use different forms of words. Stemming and lemmatisation can be applied to reduce inflectional forms or relate different forms of a word to a common base form. Stemming often chops off the end of a word, while lemmatisation uses vocabularies and morphological analysis to return the base form (or the *lemma*) of a word (Lovins, 1968; Manning, Schütze, & Raghavan, 2008). For example, the word "industries" will be turned into "industri" or "industry" if stemming or lemmatisation is applied.

To model and analyse text data, we need to transform it into numerical representations so the forecasting models can process them as predictors. One way of transforming the data is through sentiment analysis. Sentiment analysis is often applied to detect polarity within customer materials such as reviews and social media posts (Archak, Ghose, & Ipeirotis, 2011; Das & Chen, 2007). An easy way to obtain the sentiment score of a word is to look it up in a happiness dictionary (for example, the hedonometer dictionary, Hedonometer, 2020). Another common way of representing the sentiment is to use a vector of numeric values that denote the word's positivity, negativity and neutrality based on existing lexical databases such as the *WordNet* (Baccianella, Esuli, & Sebastiani, 2010; Godbole, Srinivasaiah, & Skiena, 2007). Once the sentiment of each word is calculated, we can apply an aggregation algorithm (e.g., simple average) to measure the sentiment of an entire sentence or paragraph.

In scholar articles and company reports, context features might be more important than sentiments. The bag-of-words model and word embeddings are often applied to generate numeric representations of such text. A bag-of-words model simply returns a matrix that describes the

[70] This subsection was written by Xiaojia Guo.

occurrence of words within a document (Goldberg, 2017). When we use this matrix as input to a forecasting model, each word count can be considered as a feature. The Word2Vec method is a widely used embedding method that is built based on the context of a word. Specifically, it trains a two-layer neural network that takes each word as an input to predict its surrounding words (see Section 2.7.8 for a discussion of neural networks for forecasting). The weights from the input layer to the hidden layer are then utilised as the numerical representation for the input word (Le & Mikolov, 2014). Once the text is turned into numeric representations, they can be used as predictors in any forecasting models. The most challenging part in this process is to find the right technique to extract features from the text.

In terms of software implementation, the Natural Language Toolkit (NLTK) and SpaCy library in Python can be applied to remove stop words and stem or lemmatise text (Honnibal, 2015; Loper & Bird, 2002). The bag-of-words technique is also available in NLTK. A particular implementation of the Word2Vec model is available on Google code (2013). Moreover, a public data set of movie reviews that is commonly studied in literature is available from the Stanford NLP Group (2013).

2.10. Forecasting by aggregation

2.10.1. Cross-sectional hierarchical forecasting[71]

In many applications time series can be aggregated at several levels of aggregation based on geographic or logical reasons to form hierarchical structures. These are called hierarchical time series. In the retail industry (see also Section 3.2.4), for example, individual sales of products at the bottom-level of the hierarchy can be grouped in categories and families of related products at increasing aggregation levels, with the total sales of the shop or distribution centre at the top level (Oliveira & Ramos, 2019; Pennings & van Dalen, 2017; Villegas & Pedregal, 2018). Similarly, cross-sectional hierarchies can be used for spatial aggregation to help model housing prices or traffic in transportation networks, or otherwise formed geographical demarcations (for example, Athanasopoulos, Ahmed, & Hyndman, 2009; Kourentzes & Athanasopoulos, 2019). The forecasts of hierarchical time series produced independently of the hierarchical structure generally will not add up according to the aggregation constrains of the hierarchy, i.e., they are not coherent. Therefore, hierarchical forecasting methods that generate coherent forecasts should be considered to allow appropriate decision-making at the different levels. Actually, by taking advantage of the relationships between the series across all levels these methods have shown to improve forecast accuracy (Athanasopoulos et al., 2009; Shang & Hyndman, 2017; Yagli, Yang, & Srinivasan, 2019). One of the main reasons behind this improved performance is that forecast reconciliation is effectively a special case of forecast combinations (Hollyman, Petropoulos, & Tipping, 2021); see also Section 2.6.

The most common approaches to hierarchical forecasting are bottom-up and top-down. In the bottom-up approach forecasts for each time series at the bottom-level are first produced and then these are added up to obtain forecasts for all other series at the hierarchy (Dunn, Williams, & Dechaine, 1976). Since forecasts are obtained at the bottom-level no information is lost due to aggregation. In the top-down approach forecasts for the top-level series are first generated and then these are disaggregated generally using historical proportions to obtain forecasts for the bottom-level series, which are then aggregated (Gross & Sohl, 1990). Hyndman, Ahmed, Athanasopoulos, and Shang (2011) claim that this approach introduces bias to the forecasts, however Hollyman et al. (2021) showed that it is possible to calculate unbiased top-down forecasts.

Recent research on hierarchical forecasting tackles the problem using a two-stage approach. Forecasts for the series at all levels of the hierarchy are first obtained independently without considering any aggregation constrains (we refer to these as base forecasts). Then, base forecasts are adjusted so that they become coherent (we refer to these as reconciled forecasts). This adjustment is achieved by a matrix that maps the base forecasts into new bottom-level forecasts which are then added up (Hyndman et al., 2011).

Wickramasuriya, Athanasopoulos, and Hyndman (2019) found the optimal solution for this matrix, which minimises the trace of the covariance matrix of the reconciled forecast errors (hence MinT reconciliation). This optimal solution is based on the covariance matrix of the base forecast errors which incorporates the correlation structure of the hierarchy. Wickramasuriya et al. (2019) presented several alternative estimators for this covariance matrix: (*i*) proportional to the identity which is optimal only when base forecast errors are uncorrelated and equivariant (referred to as OLS), (*ii*) proportional to the sample covariance estimator of the in-sample one-step-ahead base forecast errors with off-diagonal elements null accounts for the differences in scale between the levels of the hierarchy (referred to as WLS), (*iii*) proportional to the previous estimator unrestricted also accounts for the relationships between the series (referred to as MinT-Sample), and (*iv*) proportional to a shrinkage estimator based on the two previous estimators, parameterising the shrinkage in terms of variances and correlations, accounts for the correlation across levels (referred as MinT-Shrink). Other researchers focus on simple (equal-weighted) combinations of the forecasts produced at different hierarchical levels (Abouarghoub, Nomikos, & Petropoulos, 2018; Hollyman et al., 2021). Pritularga, Svetunkov, and Kourentzes (2021) showed that more complex reconciliation schemes result in more variability in the forecasts, due to the estimation of the elements in the covariance matrix, or the implicit combination weights. They provide approximations for the covariance matrix that balance this estimation uncertainty with the benefits of more finely tuned weights.

More recently these techniques were extended to probabilistic forecasting (Ben Taieb, Taylor, & Hyndman, 2020).

71 This subsection was written by Patrícia Ramos.

When base forecasts are probabilistic forecasts characterised by elliptical distributions, Panagiotelis, Gamakumara, Athanasopoulos, and Hyndman (2021) showed that reconciled probabilistic forecasts also elliptical can be obtained analytically. When it is not reasonable to assume elliptical distributions, a non-parametric approach based on bootstrapping in-sample errors can be used.

2.10.2. Temporal aggregation[72]

Temporal aggregation is the transformation of a time series from one frequency to another of lower frequency. As an example, a time series of length n that is originally sampled at a monthly frequency can be transformed to a quarterly series of length $n/3$ by using equally-sized time buckets of three periods each. It is usually applied in an non-overlapping manner, but overlapping aggregation can also be considered. The latter is preferred in the case when the original series is short, but has the disadvantage of applying lower weights on the few first and last observations of the series and introducing autocorrelations (Boylan & Babai, 2016, and Section 2.8.2).

Temporal aggregation is appealing as it allows to investigate the original series through different lenses. By changing the original frequency of the data, the apparent series characteristics also change. In the case of slow-moving series, temporal aggregation leads to decrease of intermittence (Nikolopoulos, Syntetos, Boylan, Petropoulos, & Assimakopoulos, 2011, see also Section 2.8). In the case of fast-moving series, higher levels of aggregation (i.e., lower frequencies) allow for better modelling of trend patterns, while lower aggregation levels (i.e., higher frequencies) are more suitable for capturing seasonal patterns (Kourentzes et al., 2014; Spithourakis, Petropoulos, Nikolopoulos, & Assimakopoulos, 2014).

Research has found evidence of improved forecasting performance with temporal aggregation for both slow (Nikolopoulos et al., 2011) and fast (Spithourakis, Petropoulos, Babai, Nikolopoulos, & Assimakopoulos, 2011) moving time series. This led to characterise temporal aggregation as a "self-improving mechanism". The good performance of temporal aggregation was reconfirmed by Babai, Ali, and Nikolopoulos (2012), who focused on its utility performance rather than the forecast error. However, one challenge with single levels of aggregation is the choice of a suitable aggregation level for each series (Kourentzes, Rostami-Tabar, & Barrow, 2017).

Instead of focusing on a single aggregation level, Andrawis et al. (2011), Kourentzes et al. (2014), Petropoulos and Kourentzes (2014), and Petropoulos and Kourentzes (2015) suggested the use of multiple levels of aggregation, usually abbreviated as MTA (multiple temporal aggregation). This not only tackles the need to select a single aggregation level, but also partly addresses the issue of model uncertainty, instead of relying on model selection and parametrisation at a single aggregation level. Using this property, Kourentzes et al. (2017) showed that MTA will typically lead to more accurate forecasts, even if in theory suboptimal. Different frequencies allow for better identification of different series patterns, so it is intuitive to consider multiple temporal levels and benefit from the subsequent forecast combination across frequencies. Kourentzes and Petropoulos (2016) showed how multiple temporal aggregation can be extended to incorporate exogenous variables (see also Section 2.2.5). However, forecasting at a single level of aggregation can still result in better performance when the seasonal pattern is strong (Spiliotis, Petropoulos & Assimakopoulos, 2019; Spiliotis, Petropoulos, Kourentzes & Assimakopoulos, 2020).

Athanasopoulos et al. (2017) expressed multiple temporal aggregation within the hierarchical forecasting framework (see Section 2.10.1) using the term "temporal hierarchies". Temporal hierarchies allow for the application of established hierarchical reconciliation approaches directly to the temporal dimension. Jeon, Panagiotelis, and Petropoulos (2019) show how temporal hierarchies can be used to obtain reconciled probabilistic forecasts, while (Spiliotis, Petropoulos et al., 2019) explored empirical bias-adjustment strategies and a strategy to avoid excessive seasonal shrinkage. Nystrup, Lindström, Pinson, and Madsen (2020) proposed estimators for temporal hierarchies suitable to account for autocorrelation in the data. Finally, Kourentzes and Athanasopoulos (2020) applied temporal hierarchies on intermittent data, and showed that higher aggregation levels may offer structural information which can improve the quality of the forecasts.

2.10.3. Cross-temporal hierarchies[73]

In the last two subsections (Sections 2.10.1 and 2.10.2), we saw two complimentary hierarchical structures, cross-sectional and temporal. Although the machinery behind both approaches is similar, often relying on the hierarchical framework by Athanasopoulos et al. (2009) and Hyndman et al. (2011), and work that followed from these (particularly Wickramasuriya et al., 2019), they address different forecasting problems. Cross-sectional hierarchies change the unit of analysis but are fixed in the period of analysis. For example, a manufacturer may operate using a hierarchy across products. The different nodes in the hierarchy will correspond to different products, product groups, super-groups, and so on, but will all refer to the same period, for example, a specific week. Temporal hierarchies do the opposite, where the unit of analysis is fixed, but the period is not. For example, we may look at the sales of a specific Stock Keeping Unit (SKU) at a daily, weekly, monthly, quarterly and annual levels. However, one can argue that having annual forecasts at the SKU level may not be useful. Similarly, having aggregate sales across an organisation at a weekly frequency is also of little value.

In connecting these to organisational decisions, we can observe that there is only a minority of problems that either cross-sectional or temporal hierarchies are natural, as typically decisions can differ across both the unit and the period (planning horizon) of analysis. In the latter case, both hierarchical approaches are more akin to statistical devices that can improve forecast accuracy through the

[72] This subsection was written by Fotios Petropoulos.

[73] This subsection was written by Nikolaos Kourentzes.

use of forecast combinations, rather than satisfy the motivating argument behind hierarchical forecasting that is to provide coherent predictions for decisions at different levels of the hierarchy.

Cross-temporal hierarchies attempt to overcome this limitation, providing coherent forecasts across all units and periods of analysis, and therefore a common outlook for the future across decision-makers at different functions and levels within an organisation. The literature remains sparse on how to construct cross-temporal forecasts, as the size of the hierarchy can easily become problematic. Kourentzes and Athanasopoulos (2019) propose a heuristic approach to overcome the ensuing estimation issues. The approach works by compartmentalising the estimation. First, they obtain estimates of the cross-sectional reconciliation weights for each temporal level of the hierarchy. Then, these are combined across temporal levels, to a unique set that satisfies all coherency constraints. Using these combined weights, they obtain the reconciled bottom level forecasts, which can be aggregated as needed. Although they recognise that their approach can result in suboptimal results in terms of reconciliation errors, it guarantees coherent forecasts. Cross-temporal forecasts are more accurate than either temporal or cross-sectional hierarchical forecasts and provide a holistic view of the future across all planning levels and demarcations. Spiliotis, Petropoulos et al. (2020) also identify the problem, however, they do not focus on the coherency of forecasts and propose a sequential reconciliation across the two dimensions. This is shown to again be beneficial, but it does not achieve coherency. Arguably one can adapt the iterative correction algorithm by Kourentzes and Athanasopoulos (2020) to enforce coherency in this approach as well.

2.10.4. Ecological inference forecasting[74]

Ecological inference forecasting (EIF) aims to predict the inner-cells values of a set of contingency tables when only the margins are known. It defines a fundamental problem in disciplines such as political science, sociology and epidemiology (Salway & Wakefield, 2004). Cleave, Brown, and Payne (1995), Greiner (2007) and Pavia, Cabrer, and Sala (2009) describe other areas of application. The fundamental difficulty of EIF lies in the fact that this is a problem with more unknowns than observations, giving rise to concerns over identifiability and indeterminacy: many sets of substantively different internal cell counts are consistent with a given marginal table. To overcome this issue, a similarity hypothesis (and, sometimes, the use of covariates) is routinely assumed. The basic hypothesis considers that either conditional row (underlying) probabilities or fractions are similar (related) among contingency tables (Greiner & Quinn, 2010). The covariations among row and column margins of the different tables are then used to learn about the internal cells.

The above hypothesis is not a cure-all to the main drawback of this approach. EIF is exposed to the so-called ecological fallacy (Robinson, 1950): the presence of inconsistencies in correlations and association measures across different levels of aggregation. This is closely related to the well-known Simpson's Paradox (Simpson, 1951). In this setting, the ecological fallacy is manifested through aggregation bias (Wakefield, 2004) due to contextual effects and/or spatial autocorrelation (Achen & Phillips Shively, 1995). This has led many authors to disqualify ecological inference forecasts (see, for example, Anselin & Tam Cho, 2002; Freedman, Klein, Ostland, & Roberts, 1998; Herron & Shotts, 2004; Tam Cho, 1998) and many others to study under which circumstances ecological inference predictions would be reliable (Firebaugh, 1978; Forcina & Pellegrino, 2019; Gelman, Park, Ansolabehere, Price, & Minnite, 2001; Guseo, 2010). Despite the criticisms, many algorithms for solving the EIF problem can be found in the literature, mainly from the ecological regression and mathematical programming frameworks (some of them available in functions of the R statistical software).

The ecological regression literature has been prolific since the seminal papers of Duncan and Davis (1953) and Goodman (1953, 1959) and is undergoing a renaissance after King (1997): new methods generalised from 2×2 tables to $R \times C$ tables have been proposed (King, Rosen, & Tanner, 1999; Rosen, Jiang, King, & Tanner, 2001), the geographical dimension of the data is being explicitly considered (Calvo & Escolar, 2003; Puig & Ginebra, 2015), and new procedures combining aggregated and individual level data, including exit polls (see also Section 3.8.5), are introduced (Glynn & Wakefield, 2010; Greiner & Quinn, 2010; Klima, Schlesinger, Thurner, & Küchenhoff, 2019). See King, Tanner, and Rosen (2004) for a wide survey and Klima, Thurner, Molnar, Schlesinger, and Küchenhoff (2016) and Plescia and De Sio (2018) for an extensive evaluation of procedures. In mathematical programming exact and inequality constraints for the inner-cell values are incorporated in a natural way. Hence, this approach has shown itself to be a proper framework for generating ecological inference forecasts. The proposals from this approach can be traced back to Hawkes (1969) and Irwin and Meeter (1969). After them, some key references include Corominas, Lusa, and Dolors Calvet (2015), McCarthy and Ryan (1977), Pavía and Romero (2021), Romero, Pavía, Martín, and Romero (2020), and Tziafetas (1986). Solutions based on other strategies, for instance, entropy maximization, have been also suggested (see, for example, Bernardini Papalia & Fernandez Vazquez, 2020; Johnston & Pattie, 2000).

2.11. Forecasting with judgment

2.11.1. Judgmental forecasting[75]

People may use judgment alone to make forecasts or they may use it in combination with statistical methods. Here the focus is on pure judgmental forecasting (for judgmental adjustments, see Section 2.11.2). Different types of judgment heuristic (mental 'rules of thumb') can be used to make forecasts. The heuristic used depends on the nature of the information available to the forecaster (Harvey, 2007).

74 This subsection was written by Jose M. Pavía.

75 This subsection was written by Nigel Harvey.

Consider cases where the only relevant information is held in the forecaster's memory. For example, someone might be asked whether Manchester United or Burnley will win next week's match. Here one memory-based heuristic that might be applicable is the recognition heuristic: if one recognises one object but not the other, then one should infer that the recognised object has higher value (Goldstein & Gigerenzer, 2002). In the above example, most people who recognise just one of the teams would be likely to make a correct forecast that Manchester United will win (Ayton, Önkal, & McReynolds, 2011). The availability heuristic is another memory-based heuristic that may be applicable: objects that are most easily brought to mind are those which are more likely. Thus, if we are asked which team is likely to come top of the premier league, we would say Manchester United if that is the one that most easily comes to mind. The availability heuristic is often effective because more likely events are encountered more often and more recently and are hence better remembered. However, it can be disrupted by, for example, greater media coverage of more unlikely (and hence more interesting) events.

Consider next cases in which forecasters possess information about values of one or more variables correlated with the variable to be forecast. For example, teachers may wish to forecast the grades of their students in a final examination on the basis of past records of various other measures. Kahneman and Tversky (1973) suggested that people use the representativeness heuristic to deal with this type of situation. Forecasters first select a variable that they think is able to represent the one that must be predicted. For example, a teacher may consider that frequency in attending voluntary revision classes represents a student's ability in the final examination. Thus, if a student attended 15 of the 20 revision classes, they are likely to obtain 75% in the final examination.

Finally, consider situations in which people forecast future values of a variable on the basis of a record of previous values of that variable. There is some evidence that, when forecasting from time series, people use anchor-and-adjustment heuristics (Hogarth & Makridakis, 1981; Lawrence & O'Connor, 1992). For example, (*i*) when forecasting from an upward trended series, they anchor on the last data point and then make an upward adjustment to take the trend into account and (*ii*) when forecasting from an untrended series containing autocorrelation, they anchor on the last data point and make an adjustment towards the mean to take the autocorrelation into account.

Kahneman (2011) and others have divided cognitive processes into those which are intuitive (System 1) and those which are deliberative (System 2). We have discussed only intuitive processes underlying judgmental forecasting (Gigerenzer, 2007). However, they can be supplemented by deliberative (System 2) processes (Theocharis & Harvey, 2019) in some circumstances.

2.11.2. Judgmental adjustments of computer-based forecasts[76]

Judgmental adjustments to algorithmic computer-based forecasts can enhance accuracy by incorporating important extra information into forecasts (Fahimnia, Sanders, & Siemsen, 2020; McNees, 1990; Perera, Hurley, Fahimnia, & Reisi, 2019). However, cognitive factors (see, for example, Section 2.11.1), and motivational biases (see Section 3.2.2), can lead to the inefficient use of information (Fildes, Goodwin, & Önkal, 2019a), unwarranted adjustments and reductions in accuracy (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Franses & Legerstee, 2009a).

People may 'egocentrically discount' a computer's forecasts when its rationale is less clear than their own reasoning (Bonaccio & Dalal, 2006). They can also be less tolerant of errors made by algorithms than those made by humans (Dietvorst, Simmons, & Massey, 2015; Önkal, Goodwin, Thomson, Gönül, & Pollock, 2009, and Section 3.7.4). The random errors associated with algorithmic forecasts, and the salience of rare large errors, can therefore lead to an unjustified loss of trust in computer forecasts. Adjustments may also give forecasters a sense of ownership of forecasts or be used to justify their role (Önkal & Gönül, 2005).

Computer-based forecasts are designed to filter randomness from time-series. In contrast, humans tend to perceive non-existent systematic patterns in random movements (O'Connor, Remus, & Griggs, 1993; Reimers & Harvey, 2011, and Section 3.7.3) and apply adjustments to reflect them. This can be exacerbated by the narrative fallacy (Taleb, 2008), where people invent stories to explain these random movements, and hindsight bias (Fischhoff, 2007), where they believe, in retrospect, that these movements were predictable. Recent random movements, and events, are particularly likely to attract undue attention, so long-run patterns identified by the computer are given insufficient weight (Bolger & Harvey, 1993). Damaging interventions are also probable when they result from political interference (Oliva & Watson, 2009) or optimism bias (Fildes et al., 2009), or when they reflect information already factored into the computer's forecast, leading to double counting (Van den Broeke, De Baets, Vereecke, Baecke, & Vanderheyden, 2019).

How can interventions be limited to occasions when they are likely to improve accuracy? Requiring people to document reasons justifying adjustments can reduce gratuitous interventions (Goodwin, 2000b). Explaining the rationale underlying statistical forecasts also improved adjustment behaviour when series had a simple underlying pattern in a study by Goodwin and Fildes (1999). However, providing guidance on when to adjust was ineffective in an experiment conducted by Goodwin, Fildes, Lawrence, and Stephens (2011), as was a restriction preventing people from making small adjustments.

When determining the size of adjustments required, decomposing the judgment into a set of easier tasks improved accuracy in a study by Webby, O'Connor, and Edmundson (2005). Providing a database of past outcomes that occurred in circumstances analogous to those expected in the forecast period also improved adjustments in a study by Lee, Goodwin, Fildes, Nikolopoulos, and Lawrence (2007). Outcome feedback, where the forecaster is informed of the most recent outcome is unlikely to be useful since it contains noise and exacerbates the tendency to over-focus on recent events (Goodwin & Fildes,

76 This subsection was written by Paul Goodwin.

1999; Petropoulos, Fildes, & Goodwin, 2016). However, feedback on biases in adjustments over several recent periods may improve judgments (Petropoulos, Goodwin, & Fildes, 2017). Feedback will be less useful where interventions are rare so there is insufficient data to assess performance.

The evidence for this section is largely based on laboratory-based studies of adjustment behaviour. Section 3.7.3 gives details of research into forecast adjustment in practice and discusses the role of forecasting support systems in improving the effectiveness of judgmental adjustments.

2.11.3. Judgmental model selection[77]

Forecasters – practitioners and researchers alike – use Forecast Support Systems (FSS) in order to perform their forecasting tasks. Usually, such an FSS allows the forecaster to load their historical data and they can then apply many different types of forecasting techniques to the selected data. The idea is that the forecaster selects the method which leads to highest forecast accuracy. Yet, there is no universally 'best' method, as it depends on the data that is being forecasted (see also Section 2.5.4). Thus, selection is important in achieving high accuracy. But how does this selection occur?

Research on judgmental selection of statistical forecasts is limited in quantity. Lawrence, Goodwin, and Fildes (2002) found that participants were not very adept at selecting good forecasting algorithms from a range offered to them by an FSS and had higher error than those who were presented with the optimal algorithm by an FSS. Petropoulos et al. (2018b) compared judgmental selection of forecasting algorithms with automatic selection based on predetermined information criteria. They found that judgmental selection was better than automatic selection at avoiding the 'worst' models, but that automatic selection was superior at choosing the 'best' ones. In the end, overall accuracy of judgmental selection was better than that of algorithmic selection. If their experiment had included more variation of the data (trends, fractals, different autoregressive factors) and variation of proposed models, this could possibly have led to better algorithmic than judgmental performance (Harvey, 2019). Time series that are more complex will place a higher cognitive load on judgmental selection. This was confirmed in a study by Han, Wang, Petropoulos, and Wang (2019), who used an electroencephalogram (EEG) for the comparison of judgmental forecast selection versus (judgmental) pattern identification. They found that pattern identification outperformed forecast selection, as the latter required a higher cognitive load, which in turn led to a lower forecasting accuracy.

It is likely that, in practice, judgmental selection is much more common than automatic selection. This preference for human judgment over advice from an algorithm has been shown in an experiment by Önkal et al. (2009). But how apt are forecasters in distinguishing 'good' models from 'bad' models? This was investigated by De Baets and Harvey (2020) in an experiment. People were asked to select the best performing model out of a choice of two different qualities (accuracies) of models (different combinations of good versus medium versus bad). People's choice outperformed forecasts made by averaging the model outputs, lending credence to the views of Fifić and Gigerenzer (2014). The performance of the participants improved with a larger difference in quality between models and a lower level of noise in the data series. In a second experiment, De Baets and Harvey (2020) found that participants adjusted more towards the advice of what they perceived to be a good quality model than a medium or bad quality one.

Importantly, in selecting an algorithm and seeing it err, people are quick to abandon it. This phenomenon is known as 'algorithm aversion' (Dietvorst et al., 2015, see also Section 2.11.6) and is due to a 'perfection schema' we have in our heads where algorithms are concerned (Madhavan & Wiegmann, 2007). We do not expect them to 'fail' and thus react strongly when they do. While a model may not perform as it should for a particular dataset and may thus elicit algorithm aversion for that particular method, one should not abandon it for all datasets and future forecasts.

2.11.4. Panels of experts[78]

Panels of experts are often used in practice to produce judgmental forecasts (see, for example, Sections 3.2.6 and 3.8.5). This is especially true in cases with limited available quantitative data and with the level of uncertainty being very high. In this section, three methods for eliciting judgmental forecasts from panels of experts are presented: the Delphi method, interaction groups (IG), and structured analogies (SA).

The Delphi method is centred around organising and structuring group communication (Rao, Anderson, Sukumar, Beauchesne, Stein, & Frankel, 2010), which aims to achieve a convergence of opinion on a specific real-world issue. It is a multiple-round survey in which experts participate anonymously to provide their forecasts and feedback (Rowe & Wright, 2001). At the end of each round, the facilitator collects and prepares statistical summaries of the panel of experts' forecasts. These summaries are presented as feedback to the group, and may be used towards revising their forecasts. This loop continues until a consensus is reached, or the experts in the panel are not willing to revise their forecasts further. In some implementations of the Delphi method, justification of extreme positions (forecasts) is also part of the (anonymous) feedback process. The Delphi method results in a more accurate outcome in the decision-making process (Dalkey, 1969; Steurer, 2011). Rowe and Wright (2001) mentioned that, by adopting the Delphi method, groups of individuals can produce more accurate forecasts than simply using unstructured methods. A drawback of the Delphi method is the additional cost associated with the need to run multiple rounds, extending the forecasting process as well as increasing the potential drop-out rates. On the other hand, the anonymity in the Delphi method eliminates

[77] This subsection was written by Shari De Baets.

[78] This subsection was written by Konstantia Litsiou.

issues such as groupthink and the 'dominant personalities' effects (Van de Ven & Delbeco, 1971).

The IG method suggests that the members of the panel of experts actively interact and debate their points to the extent they have to reach an agreement on a common forecast (Litsiou, Polychronakis, Karami, & Nikolopoulos, 2019). Sniezek and Henry (1989) found that members of interacting groups provide more accurate judgments compared to individuals. However, there is mixed evidence about the forecasting potential of IG (Boje & Murnighan, 1982; Graefe & Armstrong, 2011; Scott Armstrong, 2006). Besides, the need for arranging and facilitating meetings for the IG makes it a less attractive option.

Another popular approach to judgmental forecasting using panels of experts is SA, which refers to the recollection of past experiences and the use analogies (Green & Armstrong, 2007). In the SA method, the facilitator assembles a panel of experts. The experts are asked to recall and provide descriptions, forecasts, and similarities/differences for cases analogous to the target situation, as well as a similarity ranking for each of these analogous cases. The facilitator gathers the lists of the analogies provided by the experts, and prepares summaries, usually using weighted averages of the recalled cases based on their similarity to the target situation (see also Section 2.6.4). Semi-structured analogies (sSA) have also been proposed in the literature, where the experts are asked to provide a final forecasts based on the analogous cases they recalled, which essentially reduces the load for the facilitator (Nikolopoulos, Litsa, Petropoulos, Bougioukos, & Khammash, 2015). Nikolopoulos et al. (2015) supported that the use of SA and IG could result to forecasts that are 50% more accurate compared to unstructured methods (such as unaided judgment). One common caveat of using panels of experts is the difficulty to identify who a real expert is. Engaging experts with high level of experience, and encouraging the interaction of experts are also supported by Armstrong and Green (2018).

2.11.5. Scenarios and judgmental forecasting[79]

Scenarios provide exhilarating narratives about conceivable futures that are likely to occur. Through such depictions they broaden the perspectives of decision makers and act as mental stimulants to think about alternatives. Scenarios enhance information sharing and provide capable tools for communication within organisations. By virtue of these merits, they have been widely used in corporate planning and strategy setting since 1960's (Godet, 1982; Goodwin & Wright, 2010; Schoemaker, 1991; Wright & Goodwin, 1999, 2009). Even though utilisation of scenarios as decision advice to judgmental forecasting has been proposed earlier (Bunn & Salo, 1993; Schnaars & Topol, 1987), the research within this domain remained limited until recently when the interest in the subject has rekindled (Goodwin, Gönül, Önkal, Kocabıyıkoğlu, & Göğüş, 2019b; Önkal, Sayım, & Gönül, 2013; Wicke, Dhami, Önkal, & Belton, 2019).

The recent research has used behavioural experimentation to examine various features of scenarios and their interactions with judgmental adjustments (see Section 2.11.2) of model-based forecasts. Önkal et al. (2013) explored the 'content' effects of scenarios where through the narration either a bleak/negative future (a pessimistic scenario) or a bright/positive future (an optimistic scenario) was portrayed. On a demand forecasting context for mobile phones, the participants first received time-series data, model-based forecasts and then asked to generate point and interval forecasts as well as provide a confidence measure. With respect to the existence of scenarios, there were four conditions where the participants may receive: (*i*) no scenarios, (*ii*) optimistic scenarios, (*iii*) pessimistic scenarios, and (*iv*) both scenarios. Findings indicate that decision makers respond differently to optimistic and pessimistic scenarios. Receiving optimistic scenarios resulted in making larger adjustments to the model-based forecasts. At the same time, led to an increased confidence of the participants in their predictions. On the other hand, participants who received negative scenarios tend to lower their predictions the most among the four groups. An intriguing finding was the balancing effect of scenarios on the interval forecast symmetry. The lower interval bounds were adjusted upwards the most towards the centre-point of the interval (i.e., model-based predictions) when optimistic scenarios were received. Similarly, the upper bounds were adjusted downwards the most towards the centre-point of the interval in the presence of pessimistic scenarios.

The prospects of receiving a single scenario versus multiple scenarios were further explored in Goodwin et al. (2019b). The researchers investigated whether assimilation or contrast effects will occur when decision makers see optimistic (pessimistic) forecasts followed by pessimistic (optimistic) ones compared against receiving a single scenario in solitude. In case of assimilation, a scenario presenting an opposing world view with the initial one would cause adjustments in the opposite direction creating an offset effect. On the other hand, in case of contrast, the forecasts generated after the initial scenarios would be adjusted to more extremes when an opposing scenario is seen. In two experiments conducted in different contexts the researchers found resilient evidence for contrast effects taking place. Interestingly, seeing an opposing scenario also increased the confidence of the forecasters in their initial predictions.

In terms of the effects of scenario presence on the forecasting performance, however, the experimental evidence indicates the benefits are only circumstantial. Goodwin, Gönül, and Önkal (2019a) found that providing scenarios worsened forecast accuracy and shifted the resultant production order decisions further away from optimality. Despite this performance controversy, the decision makers express their fondness in receiving scenarios and belief in their merits (Goodwin et al., 2019b; Önkal et al., 2013). Therefore, we need more tailored research on scenarios and judgmental forecasting to reveal the conditions when scenarios can provide significant improvements to the forecasting accuracy.

[79] This subsection was written by M. Sinan Gönül.

2.11.6. Trusting model and expert forecasts[80]

Defined as "firm belief in the reliability, truth, and ability of someone/something" (Oxford English Dictionary), trust entails accepting vulnerability and risk (Rousseau, Sitkin, Burt, & Camerer, 1998). Given that forecasts are altered or even discarded when distrusted by users, examining trust is a central theme for both forecasting theory and practice.

Studies examining individual's trust in model versus expert forecasts show that individuals often distrust algorithms (Burton, Stein, & Jensen, 2020; Meehl, 2013) and place higher trust on human advice (Diab, Pui, Yankelevich, & Highhouse, 2011; Eastwood, Snook, & Luther, 2012, but also Sections 2.11.2, 2.11.3 and 3.7.4). We live in an era where we are bombarded with news about how algorithms get it wrong, ranging from COVID-19 forecasts affecting lockdown decisions to algorithmic grade predictions affecting university admissions. Individuals appear to prefer forecasts from humans over those from statistical algorithms even when those forecasts are identical (Önkal et al., 2009). Furthermore, they lose trust in algorithms quicker when they see forecast errors (Dietvorst et al., 2015; Prahl & Van Swol, 2017). Such 'algorithm aversion' and error intolerance is reduced when users have opportunity to adjust the forecasting outcome, irrespective of the extent of modification allowed (Dietvorst, Simmons, & Massey, 2018). Feedback appears to improve trust, with individuals placing higher trust in algorithms if they can understand them (Seong & Bisantz, 2008). Overuse of technical language may reduce understanding of the forecast/advice, in turn affecting perceptions of expertise and trustworthiness (Joiner, Leveson, & Langfield-Smith, 2002). Explanations can be helpful (Goodwin, Gonul & Onkal, 2013), with their packaging affecting judgments of trustworthiness (Elsbach & Elofson, 2000). Algorithmic appreciation appears to easily fade with forecasting expertise (Logg, Minson, & Moore, 2019), emphasising the importance of debiasing against overconfidence and anchoring on one's own predictions.

Trusting experts also presents challenges (Hendriks, Kienhues, & Bromme, 2015; Hertzum, 2014; Maister, Galford, & Green, 2012). Expert forecasts are typically seen as predisposed to group-based preconceptions (Brennan, 2020; Vermue, Seger, & Sanfey, 2018), along with contextual and motivational biases (Burgman, 2016). Misinformed expectations, distorted exposures to 'forecast failures', and over-reliance on one's own judgments may all contribute to distrusting experts as well as algorithms.

Credibility of forecast source is an important determinant in gaining trust (Önkal, Gönül, & De Baets, 2019). Studies show that the perceived credibility of system forecasts affects expert forecasters' behaviours and trust (Alvarado-Valencia & Barrero, 2014), while providing information on limitations of such algorithmic forecasts may reduce biases (Alvarado-Valencia, Barrero, Önkal, & Dennerlein, 2017). Previous experience with the source appears to be key to assessing credibility (Hertzum, 2002) and trust (Cross & Sproull, 2004). Such 'experienced' credibility appears to be more influential on users' acceptance of given forecasts as opposed to 'presumed' credibility (Önkal, Gönül, Goodwin, Thomson, & Öz, 2017). Source credibility can be revised when forecast (in)accuracy is encountered repetitively (Jiang, Muhanna, & Pick, 1996), with forecaster and user confidence playing key roles (Sah, Moore, & MacCoun, 2013).

Trust is critical for forecasting efforts to be translated into sound decisions (Choi, Özer, & Zheng, 2020; Özer, Zheng, & Chen, 2011). Further work on fostering trust in individual/collaborative forecasting will benefit from how trusted experts and models are selected and combined to enhance decision-making.

2.12. Evaluation, validation, and calibration

2.12.1. Benchmarking[81]

When a new forecasting model or methodology is proposed, it is common for its performance to be benchmarked according to some measure of forecast accuracy against other forecasting methods using a sub-sample of some particular time series. In this process, there is the risk that either the measures of accuracy, competing forecasting methods or test data, are chosen in a way that exaggerates the benefits of a new method. This possibility is only exacerbated by the phenomenon of publication bias (Dickersin, 1990).

A rigorous approach to benchmarking new forecasting methods should follow the following principles:

1. New methods should always be compared to a larger number of suitable benchmark methods. These should at a minimum include naïve methods such as a random walk and also popular general purpose forecasting algorithms such as ARIMA models, Exponential Smoothing, Holt Winters and the Theta method (see Section 2.3 and references therein).
2. Forecasts should be evaluated using a diverse set of error metrics for point, interval and probabilistic forecasts (see Section 2.12.2). Where the forecasting problem at hand should be tailored to a specific problem, then appropriate measures of forecast accuracy must be used. As an example, the literature on Value at Risk forecasting has developed a number of backtesting measures for evaluating the quality of quantile forecasts (see Zhang & Nadarajah, 2018, and references therein).
3. Testing should be carried out to discern whether differences between forecasting methods are statistically significant. For discussion see Section 2.12.6. However, there should also be a greater awareness of the debate around the use of hypothesis testing both in forecasting (Armstrong, 2007) and more generally in statistics (Wasserstein & Lazar, 2016).
4. Sample sizes for rolling windows should be chosen with reference to the latest literature on rolling window choice (see Inoue et al., 2017, and references therein).

[80] This subsection was written by Dilek Önkal.

[81] This subsection was written by Anastasios Panagiotelis.

5. All code used to implement and benchmark new forecasting methods should, where possible, be written in open source programming languages (such as C, Python and R). This is to ensure replicability of results (for more on the replicablity crisis in research see Peng, 2015, and references therein)
6. Methods should be applied to appropriate benchmark datasets.

Regarding the last of these points there are some examples in specific fields, of datasets that already play a de facto role as benchmarks. In macroeconomic forecasting, the U.S. dataset of Stock and Watson (2012, see §2.7.1) is often used to evaluate forecasting methods that exploit a large number of predictors, with Forni, Hallin, Lippi, and Reichlin (2003) and Panagiotelis, Athanasopoulos, Hyndman, Jiang, and Vahid (2019) having constructed similar datasets for the EU and Australia respectively. In the field of energy, the GEFCom data (Hong, Pinson, Fan, Zareipour, Troccoli, & Hyndman, 2016) discussed in Section 3.4.3 and the IEEE 118 Bus Test Case data (Peña, Martinez-Anido, & Hodge, 2018) are often used as benchmarks. Finally, the success of the M Forecasting competitions (Makridakis, Spiliotis et al., 2020) provide a benchmark dataset for general forecasting methodologies (see Section 2.12.7 and references therein).

A recent trend that has great future potential is the publication of websites that demonstrate the efficacy of different forecasting methods on real data. The Covid-19 Forecast Hub[82] and the Business Forecast Lab[83] provide notable examples in the fields of epidemiology and macroeconomics and business respectively.

2.12.2. Point, interval, and pHDR forecast error measures[84]

Point forecasts are single number forecasts for an unknown future quantity also given by a single number. *Interval forecasts* take the form of two point forecasts, an upper and a lower limit. Finally, a less common type of forecast would be a *predictive Highest Density Region* (pHDR), i.e., an HDR (Hyndman, 1996) for the conditional density of the future observable. pHDRs would be interesting for multimodal (possibly implicit) predictive densities, e.g., in scenario planning. Once we have observed the corresponding realisation, we can evaluate our point, interval and pHDR forecasts.

There are many common point forecast error measures (PFEMs), e.g., the mean squared error (MSE), mean absolute error (MAE), mean absolute scaled error (MASE), mean absolute percentage error (MAPE) or many others (see Section 3.4 in Hyndman & Athanasopoulos, 2018). Which one is most appropriate for our situation, or should we even use multiple different PFEMs?

There are many common point forecast error measures (PFEMs), e.g., the MSE, MAE, MASE, (s)MAPE, the quantile score or pinball loss or many others (e.g., sections 5.8 and 5.9 in Hyndman & Athanasopoulos, 2021). Assuming n historical periods with observations $y_1, \ldots, y_n$ and a forecasting horizon H with observations $y_{n+1}, \ldots, y_{n+H}$ and point forecasts $f_{n+1}, \ldots, f_{n+H}$, we have:

$$\mathrm{MSE} = \sum_{t=n+1}^{n+H} (y_t - f_t)^2,$$

$$\mathrm{MAE} = \sum_{t=n+1}^{n+H} |y_t - f_t|, \quad \mathrm{MASE} = \frac{\sum_{t=n+1}^{n+H} |y_t - f_t|}{\sum_{t=2}^{n} |y_t - y_{t-1}|}$$

$$\mathrm{MAPE} = \sum_{t=n+1}^{n+H} \frac{|y_t - f_t|}{y_t}, \quad \mathrm{sMAPE} = \sum_{t=n+1}^{n+H} \frac{|y_t - f_t|}{\frac{1}{2}(y_t + f_t)}$$

$$Q_\alpha = \sum_{t=n+1}^{n+H} (1-\alpha)(f_t - y_t)1_{y_t<f_t} + \alpha(y_t - f_t)1_{y_t \geq f_t}.$$

Let us take a step back. Assume we have a full density forecast and wish to "condense" it to a point forecast that will minimise some PFEM in expectation. The key observation is that *different PFEMs will be minimised by different point forecasts derived from the same density forecast* (Kolassa, 2020b).

- The MSE is minimised by the expectation.
- The MAE and MASE are minimised by the median (Hanley, Joseph, Platt, Chung, & Belisle, 2001).
- The MAPE is minimised by the (-1)-median (Gneiting, 2011a, p. 752 with $\beta = -1$).
- The sMAPE is minimised by an unnamed functional that would need to be minimised numerically (Gonçalves, 2015).
- The hinge/tick/pinball Q_p loss is minimised by the appropriate p-quantile (Gneiting, 2011b).
- In general, there is no loss function that is minimised by the mode (Heinrich, 2014).

We note that intermittent demand (see Section 2.8) poses specific challenges. On the one hand, the MAPE is undefined if there are zeros in the actuals. On the other hand, the point forecasts minimising different PFEMs will be very different. For instance, the conditional median (minimising the MAE) may well be a flat zero, while the conditional mean (minimising the MSE) will usually be nonzero. Our forecasting algorithm may not output an explicit density forecast. It is nevertheless imperative to think about which functional of the implicit density we want to elicit (Gneiting, 2011a), and tailor our error measure – and forecasting algorithm! – to it. It usually makes no sense to evaluate a point forecast with *multiple* PFEMs (Kolassa, 2020b).

Interval forecasts can be specified in multiple ways. We can start with a probability coverage and require two appropriate quantiles – e.g., we could require a 2.5% and a 97.5% quantile forecast, yielding a symmetric or equal-tailed 95% interval forecast. Interval forecasts (ℓ_t, u_t) of this form can be evaluated by the interval score (Brehmer & Gneiting, 2020; Winkler, 1972), a proper scoring rule (section 6.2 in Gneiting & Raftery, 2007):

$$\mathrm{IS}_\alpha = \sum_{t=n+1}^{n+H} (u_t - \ell_t) + \frac{2}{\alpha}(\ell_t - y_t)1_{y_t<\ell_t} + \frac{2}{\alpha}(y_t - u_t)1_{y_t>u_t}. \tag{3}$$

82 https://viz.covid19forecasthub.org/.

83 https://business-forecast-lab.com/.

84 This subsection was written by Stephan Kolassa.

We can also use the hinge loss to evaluate the quantile forecasts separately.

Alternatively, we can require a shortest interval subject to a specified coverage. This interval is not elicitable relative to practically relevant classes of distributions (Brehmer & Gneiting, 2020; Fissler, Frongillo, Hlavinová, & Rudloff, 2020).

Yet another possibility is to maximise the interval forecast's probability coverage, subject to a maximum length ℓ. This modal interval forecast $(f_t, f_t+\ell)$ is elicitable by an appropriate ℓ-zero–one-loss (Brehmer & Gneiting, 2020):

$$L_\ell = \sum_{t=n+1}^{n+H} 1_{f_t<y_t<f_t+\ell} \\ = \#\left\{t \in n+1,\ldots,n+H \,\middle|\, f_t < y_t < f_t+\ell\right\}. \quad (4)$$

The pHDR is not elicitable even for unimodal densities (Brehmer & Gneiting, 2020). In the multimodal case, the analysis is likely difficult. Nevertheless, a variation of the Winkler score has been proposed to evaluate pHDRs on an ad hoc basis (Hyndman, 2020). One could also compare the achieved to the nominal coverage, e.g., using a binomial test – which disregards the volume of the pHDR (Kolassa, 2020a).

In conclusion, there is a bewildering array of PFEMs, which require more thought in choosing among than is obvious at first glance. The difficulties involved in evaluating interval and pHDR forecasts motivate a stronger emphasis on full density forecasts (cf. Askanazi, Diebold, Schorfheide, & Shin, 2018, and Section 2.12.4).

2.12.3. Scoring expert forecasts[85]

Evaluating forecasting capabilities can be a difficult task. One prominent way to evaluate an expert's forecast is to score the forecast once the realisation of the uncertainty is known. Scoring forecasts using the outcome's realisations over multiple forecasts offers insights into an individual's expertise. Experts can also use scoring information to identify ways to improve future forecasts. In addition, scoring rules and evaluation measures can be designed to match decision-making problems, incentivising forecasts that are most useful in a specific situation (Winkler, Grushka-Cockayne, Lichtendahl, & Jose, 2019).

Scoring rules were first suggested for evaluate meteorological forecasts in work by Brier (1950). Scoring rules have since been used in a wide variety of settings, such as business and other applications. When forecasting a discrete uncertainty with only two possible outcomes (e.g., a loan with be defaulted on or not, a customer will click on an ad or not), the Brier score assigns a score of $-(1-p)^2$, where p is the probability forecast reported that the event will occurs. The greater the probability reported for an event that occurs, the higher the score the forecast receives. Over multiple forecasts, better forecasters will tend to have higher average Brier scores. For discrete events with more than two outcomes, a logarithmic scoring rule can be used.

The scoring rules are attractive to managers in practice since they are considered proper. Proper scoring rules (see also Section 2.12.4) incentivise honest forecasts from the experts, even prior to knowing the realisation of an uncertainty, since ex ante the expected score is maximised only when reported probabilities equals true beliefs (Bickel, 2007; Gneiting & Raftery, 2007; Merkle & Steyvers, 2013; O'Hagan et al., 2006; Winkler, Muñoz, Cervera, Bernardo, Blattenberger, Kadane, Lindley, Murphy, Oliver, & Ríos-Insua, 1996). Examples of a scoring rule that is not proper yet still commonly used are the linear score, which simply equals the reported probability or density for the actual outcome, or the skill score, which is the percentage improvement of the Brier score for the forecast relative to the Brier score of some base line naive forecast (Winkler et al., 2019).

For forecasting continuous quantities, forecasts could be elicited by asking for an expert's quantile (or fractile) forecast rather than a probability forecast. For instance, the 0.05, 0.25, 0.50, 0.75 and 0.95 quantiles are often elicited in practice, and in some cases every 0.01 quantile, between 0–1 are elicited (e.g., the 2014 Global Energy Forecasting Competition, Hong et al., 2016). Proper scoring rules for quantiles are developed in Jose and Winkler (2009).

When forecasts are used for decision-making, it is beneficial if the scoring rule used relates in some manner to the decision problem itself. In certain settings, the connection of the scoring rule to the decision context is straight forward. For example, Jose, Nau, and Winkler (2008) develop scoring rules that can be mapped to decision problems based on the decision maker's utility function. Johnstone, Jose, and Winkler (2011) develop tailored scoring rules aligning the interest of the forecaster and the decision maker. Grushka-Cockayne, Lichtendahl, Jose and Winkler (2017) link quantile scoring rules to business profit-sharing situations.

2.12.4. Evaluating probabilistic forecasts[86]

Probabilistic forecasting is a term that is not strictly defined, but usually refers to everything beyond point forecasting (Gneiting, 2011a). However, in this section we consider only the evaluation of full predictive distributions or equivalent characterisations. For the evaluation of prediction of quantiles, intervals and related objects, see Section 2.12.2.

One crucial point for evaluating probabilistic forecasts is the reporting, which is highly influenced from meteorologic communities. From the theoretical point of view, we should always report the predicted cumulative distribution function $\widehat{F}$ of our prediction target F. Alternatively for continuous data, reporting the probability density function is a popular choice. For univariate prediction problems a common alternative is to report quantile forecast on a dense grid of probabilities, as it approximates the full distribution (Hong et al., 2016). For multivariate forecasts, it seems to become standard to report a large ensemble (a set of simulated trajectories/paths) of the full predictive

85 This subsection was written by Yael Grushka-Cockayne.

86 This subsection was written by Florian Ziel.

distribution. The reason is that the reporting of a multivariate distribution (or an equivalent characterisation) of sophisticated prediction models is often not feasible or practicable, especially for non-parametric or copula-based forecasting methods.

In general, suitable tools for forecasting evaluation are proper scoring rules as they address calibration and sharpness simultaneously (Gneiting & Katzfuss, 2014; Gneiting & Raftery, 2007). Preferably, we consider strictly proper scoring rules which can identify the true predicted distribution among a set of forecast candidates that contains the true model.

In the univariate case the theory is pretty much settled and there is quite some consensus about the evaluation of probabilistic forecasts (Gneiting & Katzfuss, 2014). The continuous ranked probability score (CRPS) and logarithmic scores (log-score) are popular strictly proper scoring rules, while the quadratic and pseudospherical score remain strictly proper alternatives. The CRPS can be well approximated by averaging across quantile forecasts on an equidistant grid of probabilities (Nowotarski & Weron, 2018).

For multivariate forecast evaluation the situation is more complicated and many questions remain open (Gneiting & Raftery, 2007; Meng, Taylor, Ben Taieb, & Li, 2020). The multivariate version of the log-score is a strictly proper scoring rule, but it requires the availability of a multivariate density forecast. This makes it impracticable for many applications. Gneiting and Raftery (2007) discuss the energy score, a multivariate generalisation of the CRPS, that is strictly proper. Still, it took the energy score more than a decade to increase its popularity in forecasting. A potential reason is the limited simulation study of Pinson and Tastu (2013) that concludes that the energy score can not discriminate well differences in the dependency structure. In consequence other scoring rules were proposed in literature, e.g., the variogram score (Scheuerer & Hamill, 2015) which is not strictly proper. Ziel and Berk (2019) consider a strictly proper scoring method for continuous variables using copula techniques. In contrast to Pinson and Tastu (2013), recent studies (Lerch et al., 2020; Ziel & Berk, 2019) show that the energy score discriminates well when used together with significance tests like the Diebold–Mariano (DM) test. In general, we recommended scoring be applied with reliability evaluation (see Section 2.12.5) and significance tests (see Section 2.12.6). Additionally, if we want to learn about the performance of our forecasts it is highly recommended to consider multiple scoring rules and evaluate on lower-dimensional subspaces. For multivariate problems, this holds particularly for the evaluation of univariate and bivariate marginal distributions.

2.12.5. Assessing the reliability of probabilistic forecasts[87]

Probabilistic forecasts in the form of predictive distributions are central in risk-based decision making where reliability, or calibration, is a necessary condition for the optimal use and value of the forecast. A probabilistic forecast is calibrated if the observation cannot be distinguished from a random draw from the predictive distribution or, in the case of ensemble forecasts, if the observation and the ensemble members look like random draws from the same distribution. Additionally, to ensure their utility in decision making, forecasts should be sharp, or specific, see Sections 2.12.2 and 2.12.4 as well as (Gneiting et al., 2007).

In the univariate setting, several alternative notions of calibration exist for both a single forecast (Gneiting et al., 2007; Tsyplakov, 2013) and a group of forecasts (Strähl & Ziegel, 2017). The notion most commonly used in applications is probabilistic calibration (Dawid, 1984); the forecast system is probabilistically calibrated if the probability integral transform (PIT) of a random observation, that is, the value of the predictive cumulative distribution function in the observation, is uniformly distributed. If the predictive distribution has a discrete component, a randomised version of the PIT should be used (Gneiting & Ranjan, 2013).

Probabilistic calibration is assessed visually by plotting the histogram of the PIT values over a test set. A calibrated forecast system will return a uniform histogram, a ∩-shape indicates overdispersion and a ∪-shape indicates underdispersion, while a systematic bias results in a biased histogram (e.g. Thorarinsdottir & Schuhen, 2018). The discrete equivalent of the PIT histogram, which applies to ensemble forecasts, is the verification rank histogram (Anderson, 1996; Hamill & Colucci, 1997). It shows the distribution of the ranks of the observations within the corresponding ensembles and has the same interpretation as the PIT histogram.

For small test sets, the bin number of a PIT/rank histogram must be chosen with care. With very few bins, the plot may obscure miscalibration while with many bins, even perfectly calibrated forecasts can yield non-uniformly appearing histograms (Heinrich, 2020; Thorarinsdottir & Schuhen, 2018). The bin number should be chosen based on the size of the test set, with the bin number increasing linearly with the size of the test set (Heinrich, 2020). More specifically, the uniformity of PIT/rank values can be assessed with statistical tests (Delle Monache, Hacker, Zhou, Deng, & Stull, 2006; Taillardat, Mestre, Zamo, & Naveau, 2016; Wilks, 2019), where the test statistics can be interpreted as a distance between the observed and a flat histogram (Heinrich, 2020; Wilks, 2019). Testing predictive performance is further discussed in Section 2.12.6.

Calibration assessment of multivariate forecasts is complicated by the lack of a unique ordering in higher dimensions and the many ways in which the forecasts can be miscalibrated (Wilks, 2019). Gneiting, Stanberry, Grimit, Held, and Johnson (2008) propose a general two-step approach where an ensemble forecast and the corresponding observation are first mapped to a single value by a pre-rank function. Subsequently, the pre-rank function values are ranked in a standard manner. The challenge here is to find a pre-rank function that yields informative and discriminative ranking (Gneiting et al., 2008; Thorarinsdottir, Scheuerer, & Heinz, 2016; Wilks, 2004); see Thorarinsdottir et al. (2016) and Wilks (2019) for comparative studies. Alternatively, Ziegel and Gneiting (2014) propose a direct multivariate extension of the univariate setting based on copulas.

[87] This subsection was written by Thordis Thorarinsdottir.

2.12.6. Statistical tests of forecast performance[88]

A natural consequence of growth in forecasting methodologies was the development of statistical tests for predictive ability in the last thirty years. These tests provided forecasters some formal reassurance that the predictive superiority of a leading forecast is statistically significant and is not merely due to random chance.

One of the early papers that undoubtedly sparked growth in this field was Diebold and Mariano (1995, DM hereafter). In their seminal paper, DM provided a simple yet general approach for testing equal predictive ability, i.e., if two forecasting sources ($f_{1,t}$ and $f_{2,t}$, $t = 1, \ldots, h$) are equally accurate on average. Mathematically, if we denote the error $e_{i,t} = y_t - f_{i,t}$ for $i = 1, 2$ and $t = 1, \ldots, h$, the hypotheses for this DM test is H_0: $E[L(-e_{1,t}) - L(-e_{2,t})] = 0$ for all t versus H_1: $E[L(-e_{1,t}) - L(-e_{2,t})] \neq 0$ under a loss function L. Their population-level predictive ability test has very few assumptions (e.g., covariance stationary loss differential) and is applicable to a wide range of loss functions, multi-period settings, and wide class of forecast errors (e.g., non-Gaussian, serially and/or contemporaneously correlated). This test though not originally intended for models has been widely used by others to test forecasting models' accuracy (Diebold, 2015).

Modifications were later introduced by Harvey, Leybourne, and Newbold (1998) to improve small sample properties of the test. Generalisations and extensions have emerged to address issues that DM tests encountered in practice such as nested models (Clark & McCracken, 2001, 2009), parameter estimation error (West, 1996), cointegrated variables (Corradi, Swanson, & Olivetti, 2001), high persistence (Rossi, 2005), and panel data (Timmermann & Zhu, 2019). Finite-sample predictive ability tests also emerged from the observation that models may have equal predictive ability in finite samples, which generated a class called conditional predictive accuracy tests (Clark & McCracken, 2013; Giacomini & White, 2006).

An alternative approach to comparing forecast accuracy is through the notion of forecast encompassing, which examines if a forecast encompasses all useful information from another with respect to predictions (Chong & Hendry, 1986; Clark & McCracken, 2001; Harvey et al., 1998). Though it has a few more assumptions, forecast encompassing tests in certain contexts might be preferable to the mean square prediction error tests à la Diebold–Mariano (Busetti & Marcucci, 2013).

Another stream of available statistical tests looks at multiple forecasts simultaneously instead of pairs. Addressing a need for a reality check on "data snooping", White (2000) later modified by Hansen (2005) developed a multiple model test that uses a null hypothesis of "superior predictive ability" instead of the equal predictive ability used in DM tests. These have also been generalised to deal with issues such as cointegrated variables (Corradi et al., 2001) and multi-horizon forecasts (Quaedvlieg, 2019). Recently, Li, Liao and Quaedvlieg (2020) proposed a conditional superior predictive ability test similar to Giacomini and White (2006)'s innovation to the DM test. A different approach for studying performance of multiple forecasting models is through the use of multiple comparison tests such as multiple comparison with a control and multiple comparison with the best (Edwards & Hsu, 1983; Horrace & Schmidt, 2000; Hsu, 1981). These tests often are based on jointly estimated confidence intervals that measure the difference between two parameters of interest such as the forecast accuracies of a model and a benchmark. Koning, Franses, Hibon, and Stekler (2005) illustrates how they can be ex post used to analyse forecasting performance in the M3 forecasting competition (Makridakis & Hibon, 2000) using model ranking instead of forecast accuracy scores as its primitives. The multiple comparison of the best was used in the analysis of the subsequent M4 and M5 Competitions (Makridakis, Spiliotis et al., 2020; Makridakis, Spiliotis & Assimakopoulos, 2021, and Section 2.12.7).

2.12.7. Forecasting competitions[89]

Forecasting competitions provide a "playground" for academics, data scientists, students, practitioners, and software developers to compare the forecasting performance of their methods and approaches against others. Organisers of forecasting competitions test the performance of the participants' submissions against some hidden data, usually the last window of observations for each series. The benefits from forecasting competitions are multifold. Forecasting competitions (*i*) motivate the development of innovative forecasting solutions, (*ii*) provide a deeper understanding of the conditions that some methods work and others fail, (*iii*) promote knowledge dissemination, (*iv*) provide a much-needed, explicit link between theory and practice, and (*v*) leave as a legacy usable and well-defined data sets. Participation in forecasting competitions is sometimes incentivised by monetary prizes. However, the stakes are usually much higher, including reputational benefits.

The most famous forecasting competitions are the ones organised by Spyros Makridakis. Initially, the research question focused on the relative performance of simple versus complex forecast. M and M3 competitions (Makridakis et al., 1982; Makridakis & Hibon, 2000) empirically showed that simple methods (such as exponential smoothing; see Section 2.3.1) are equally good compared to other more complex methods and models (such as ARIMA and neural networks; see Section 2.3.4 and Section 2.7.8 respectively) in point-forecast accuracy – if not better. Moreover, the early Makridakis competitions showed the importance of forecast combinations in increasing predictive accuracy. For example, the winner of the M3 competition was the Theta method (see Section 2.3.3), a simple statistical method that involved the combination of linear regression and simple exponential smoothing forecasts (Assimakopoulos & Nikolopoulos, 2000).

The M4 competition (Makridakis, Spiliotis et al., 2020) challenged researchers and practitioners alike with a task of producing point forecasts and prediction intervals for 100 thousand time series of varied frequencies. This time,

[88] This subsection was written by Victor Richmond R. Jose.

[89] This subsection was written by Fotios Petropoulos.

the main hypothesis focused on the ability of machine learning and neural network approaches in the task of time series forecasting. Machine learning approaches (see Section 2.7.10) that focused on each series independently performed poorly against statistical benchmarks, such as Theta, Damped exponential smoothing or simple averages of exponential smoothing models. However, the best two performing submissions in the M4 competition (Montero-Manso et al., 2020; Smyl, 2020) used neural network and machine learning algorithms towards utilising cross-learning. So, the main learning outcome from the M4 competition is that, if utilised properly, machine learning can increase the forecasting performance. Similarly to previous competitions, M4 demonstrated again the usefulness of combining across forecasts, with five out of the top six submissions offering a different implementation of forecast combinations.

Several other forecasting competitions focused on specific contexts and applications. For example, M2 competition (Makridakis et al., 1993) suggested that the benefits from additional information (domain expertise) are limited; see also Section 2.11.4. The tourism forecasting competition (Athanasopoulos, Hyndman, Song, & Wu, 2011) also showed that exogenous variables do not add value, while naive forecasts perform very well on a yearly frequency (for a discussion on tourism forecasting applications, see Section 3.8.1). The NN3 competition (Crone, Hibon, & Nikolopoulos, 2011) confirmed the superior performance of statistical methods, but noted that neural network approaches are closing the distance. Tao Hong's series of energy competitions (Hong, Pinson, & Fan, 2014; Hong et al., 2016; Hong, Xie, & Black, 2019) demonstrated best practices for load, price, solar, and wind forecasting, with extensions to probabilistic and hierarchical forecasts (for energy forecasting applications, see Section 3.4). Finally, many companies have hosted forecasting challenges through the Kaggle platform. Bojer and Meldgaard (2020) reviewed the Kaggle competitions over the last five years, and concluded that access to hierarchical information, cross-learning, feature engineering, and combinations (ensembles) can lead to increased forecasting performance, outperforming traditional statistical methods. These insights were a forerunner to the results of the M5 competition, which focused on hierarchically organised retail data (Makridakis, Spiliotis, Assimakopoulos, 2021; Makridakis, Spiliotis, Assimakopoulos, Chen & Winkler, 2021).

Makridakis, Fry, Petropoulos and Spiliotis (2021) provide a list of design attributes for forecasting competitions and propose principles for future competitions.

2.13. The future of forecasting theory[90]

The theory of forecasting appears mature today, based on dedicated developments at the interface among a number of disciplines, e.g., mathematics and statistics, computer sciences, psychology, etc. A wealth of these theoretical developments have originated from specific needs and challenges in different application areas, e.g., in economics, meteorology and climate sciences, as well as management science among others. In this section, many aspects of the theory of forecasting were covered, with aspects related to data, modelling and reasoning, forecast verification. Now, the fact that forecasting is mature does not mean that all has been done – we aim here at giving a few pointers at current and future challenges.

First of all, it is of utmost importance to remember that forecasting is a process that involves both quantitative aspects (based on data and models) and humans, at various levels, i.e., from the generation of forecasts to their use in decision-making. A first consequence is that we always need to find, depending on the problem at hand, an optimal trade-off between data-driven approaches and the use of expert judgment. In parallel, forecasting is to be thought of in a probabilistic framework in a systematic manner (Gneiting & Katzfuss, 2014). This allows us to naturally convey uncertainty about the future, while providing the right basis to make optimal decisions in view of the characteristics of the decision problem, as well as the loss (or utility) function and risk aversion of the decision maker. Another consequence is that using forecasts as input to decision-making often affects the outcome to be predicted itself – a problem known as self-negating forecasts (possibly also self-fulfilling) or the prophet dilemma. With advances in the science of dynamic systems and game theory, we should invest in modelling those systems as a whole (i.e., forecasting and decision-making) in order to predict the full range of possible outcomes, based on the decisions that could be made.

In parallel, it is clear that today, the amount of data being collected and possibly available for forecasting is growing at an astounding pace. This requires re-thinking our approaches to forecasting towards high-dimensional models, online learning, etc. Importantly, the data being collected is distributed in terms of ownership. And, due to privacy concerns and competitive interests, some may not be ready to share their data. Novel frameworks to learning and forecasting ought to be developed with that context in mind, for instance focusing on distributed and privacy-preserving learning – an example among many others is that of Google pushing forward federated learning (Abadi et al., 2016), an approach to deep learning where the learning process is distributed and with a privacy layer. Eventually the access and use of data, as well as the contribution to distributed learning (and collaborative analytics, more generally), may be monetised, bringing a mechanism design component to the future theory of forecasting. A simple and pragmatic example is that of forecast reconciliation: if asking various agents to modify their forecasts to make them coherent within a hierarchy, such modifications could be monetised to compensate for accuracy loss.

A large part of today's modelling and forecasting approaches uses a wealth of data to identify and fit models, to be eventually used to forecast based on new data and under new conditions. Different approaches have been proposed to maximise the generalisation ability of those models, to somewhat maximise chances to do well out-of-sample. At the root of this problem is the effort to go beyond correlation only, and to identify causality (see,

[90] This subsection was written by Pierre Pinson.

e.g., Pearl (2009) for a recent extensive coverage). While causality has been a key topic of interest to forecasters for a long time already, new approaches and concepts are being pushed forward for identification of and inference in causal models (Peters, Janzing, & Schölkopf, 2017), which may have a significant impact on the theory of forecasting.

Eventually, the key question of *what a good forecast is* will continue to steer new developments in the theory of forecasting in the foreseeable future. The nature of goodness of forecasts (seen from the meteorological application angle) was theorised a few decades ago already (Murphy, 1993), based on consistency, quality and value. We still see the need to work further on that question – possibly considering these 3 pillars, but possibly also finding other ways to define desirable properties of forecasts. This will, in all cases, translates to further developing frameworks for forecast verification, focusing on the interplay between forecast quality and value, but also better linking to psychology and behavioural economics. In terms of forecast verification, some of the most pressing areas most likely relate to (multivariate) probabilistic forecasting and to the forecasting of extreme events. When it comes to forecast quality and value, we need to go beyond the simple plugging of forecasts into decision problems to assess whether this yields better decisions, or not. Instead, we ought to propose suitable theoretical frameworks that allow assessing whether certain forecasts are fundamentally better (than others) for given classes of decision problems. Finally, the link to psychology and behavioural economics should ensure a better appraisal of how forecasts are to be communicated, how they are perceived and acted upon.

Most of the advances in the science of forecasting have come from the complementarity between theoretical developments and applications. We can then only be optimistic for the future since more and more application areas are relying heavily on forecasting. Their specific needs and challenges will continue fuelling upcoming developments in the theory of forecasting.

3. Practice

3.1. Introduction to forecasting practice[91]

The purpose of forecasting is to improve decision making in the face of uncertainty. To achieve this, forecasts should provide an unbiased guess at what is most likely to happen (the point forecast), along with a measure of uncertainty, such as a prediction interval (PI). Such information will facilitate appropriate decisions and actions.

Forecasting should be an objective, dispassionate exercise, one that is built upon facts, sound reasoning, and sound methods. But since forecasts are created in social settings, they are influenced by organisational politics and personal agendas. As a consequence, forecasts will often reflect aspirations rather than unbiased projections.

In organisations, forecasts are created through processes that can involve multiple steps and participants. The process can be as simple as executive fiat (also known as evangelical forecasting), unencumbered by what the data show. More commonly, the process begins with a statistical forecast (generated by forecasting software), which is then subject to review and adjustment, as illustrated in Fig. 5.

In concept, such an elaborate multi-stage process allows "management intelligence" to improve forecast quality, incorporating information not accounted for in the statistical model. In reality, however, benefits are not assured. Lawrence, Goodwin, O'Connor, and Önkal (2006) reviewed more than 200 studies, concluding that human judgment can be of significant benefit but is also subject to significant biases. Among the many papers on this subject, there is general agreement on the need to track and review overrides, and the need to better understand the psychological issues around judgmental adjustments.

The underlying problem is that each human touch point subjects the forecast to the interests of the reviewers – and these interests may not align with creating an accurate, unbiased forecast. To identify where such problems are occurring, Forecast Value Added (FVA) analysis is an increasingly popular approach among practitioners.

FVA is defined as the change in a forecasting performance metric that can be attributed to a particular step or participant in the forecasting process (Gilliland, 2002). Any activity that fails to deliver positive FVA (i.e., fails to improve forecast quality) is considered process waste.

Starting with a naive forecast, FVA analysis seeks to determine whether each subsequent step in the process improves upon the prior steps. The "stairstep report" of Table 1 is a familiar way of summarising results, as in this example from Newell Rubbermaid (Schubert & Rickard, 2011).

Here, averaged across all products, naive (random walk) achieved forecast accuracy of 60%. The company's statistical forecast delivered five percentage points of improvement, but management review and adjustment delivered negative value. Such findings – not uncommon – urge further investigation into causes and possible process corrections (such as training reviewers or limiting adjustments). Alternatively, the management review step could be eliminated, providing the dual benefits of freeing up management time spent on forecasting and, on average, more accurate forecasts.

Morlidge (2014c) expanded upon FVA analysis to present a strategy for prioritising judgmental adjustments, finding the greatest opportunity for error reduction in products with high volume and high relative absolute error. Chase (2021) described a machine learning (ML) method to guide forecast review, identifying which forecasts are most likely to benefit from adjustment along with a suggested adjustment range. Baker (2021) used ML classification models to identify characteristics of non-value adding overrides, proposing the behavioural economics notion of a "nudge" to prompt desired forecaster behaviour. Further, Goodwin, Petropoulos, and Hyndman (2017) derived upper bounds for FVA relative to naive forecasts. And de Kok (2017) created a Stochastic Value Added (SVA) metric to assess the difference between actual and forecasted distributions, knowledge of which is valuable for inventory management.

[91] This subsection was written by Michael Gilliland.

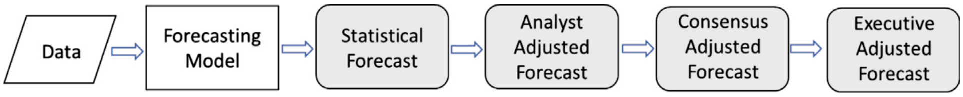

Fig. 5. Multi-stage forecasting process.

Table 1
Stairstep report showing FVA results.

Process step	Forecast accuracy (100%–MAPE)	FVA vs. Naive	FVA vs. statistical
Naive forecast	60%		
Statistical forecast	65%	5%	
Adjusted forecast	62%	2%	−3%

Including an indication of uncertainty around the point forecast remains an uncommon practice. Prediction intervals in software generally underestimate uncertainty, often dramatically, leading to unrealistic confidence in the forecast. And even when provided, PIs largely go unused by practitioners. Goodwin (2014) summarised the psychological issues, noting that the generally poor calibration of the PIs may not explain the reluctance to utilise them. Rather, "an interval forecast may accurately reflect the uncertainty, but it is likely to be spurned by decision makers if it is too wide and judged to be uninformative" (Goodwin, 2014, page 5).

It has long been recognised (Chatfield, 1986; Lawrence, 2000) that the practice of forecasting falls well short of the potential exhibited in academic research, and revealed by the M forecasting competitions. In the M4, a simple benchmark combination method (the average of Single, Holt, and Damped exponential smoothing) reduced the overall weighted average (OWA) error by 17.9% compared to naive. The top six performing methods in M4 further reduced OWA by over 5% compared to the combination benchmark (Makridakis, Spiliotis et al., 2020). But in forecasting practice, just bettering the accuracy of naive has proven to be a surprising challenge. Morlidge's (2014b) study of eight consumer and industrial businesses found 52% of their forecasts failed to do so. And, as shown, Newel Rubbermaid beat naive by just two percentage points after management adjustments.

Ultimately, forecast accuracy is limited by the nature of the behaviour being forecast. But even a highly accurate forecast is of little consequence if overridden by management and not used to enhance decision making and improve organisational performance.

Practitioners need to recognise limits to forecastability and be willing to consider alternative (non-forecasting) approaches when the desired level of accuracy is not achievable (Gilliland, 2010). Alternatives include supply chain re-engineering – to better react to unforeseen variations in demand, and demand smoothing – leveraging pricing and promotional practices to shape more favourable demand patterns.

Despite measurable advances in our statistical forecasting capabilities (Makridakis, Hyndman & Petropoulos, 2020), it is questionable whether forecasting practice has similarly progressed. The solution, perhaps, is what Morlidge (2014a, page 39) suggests that "users should focus less on trying to optimise their forecasting process than on detecting where their process is severely suboptimal and taking measures to redress the problem". This is where FVA can help.

For now, the challenge for researchers remains: To prompt practitioners to adopt sound methods based on the objective assessment of available information, and avoid the "worst practices" that squander resources and fail to improve the forecast.

3.2. Operations and supply chain management

3.2.1. Demand management[92]

Demand management is one of the dominant components of supply chain management (Fildes, Goodwin, & Lawrence, 2006). Accurate demand estimate of the present and future is a first vital step for almost all aspects of supply chain optimisation, such as inventory management, vehicle scheduling, workforce planning, and distribution and marketing strategies (Kolassa & Siemsen, 2016). Simply speaking, better demand forecasts can yield significantly better supply chain management, including improved inventory management and increased service levels. Classic demand forecasts mainly rely on qualitative techniques, based on expert judgment and past experience (e.g., Weaver, 1971), and quantitative techniques, based on statistical and machine learning modelling (e.g., Bacha & Meyer, 1992; Taylor, 2003b). A combination of qualitative and quantitative methods is also popular and proven to be beneficial in practice by, e.g., judgmental adjustments (Önkal & Gönül, 2005; Syntetos, Kholidasari & Naim, 2016; Turner, 1990, and Section 2.11.2), judgmental forecast model selection (Han et al., 2019; Petropoulos et al., 2018b, and Section 2.11.3), and other advanced forecasting support systems (Arvan, Fahimnia, Reisi, & Siemsen, 2019; Baecke, De Baets, & Vanderheyden, 2017, see also Section 3.7.1).

The key challenges that demand forecasting faces vary from domain to domain. They include:

1. The existence of intermittent demands, e.g., irregular demand patterns of fashion products. According to Nikolopoulos (2020), limited literature has focused on intermittent demand. The seminal work by Croston (1972) was followed by other representative methods such as the SBA method by Syntetos

[92] This subsection was written by Yanfei Kang.

and Boylan (2001), the aggregate–disaggregate intermittent demand approach (ADIDA) by Nikolopoulos et al. (2011), the multiple temporal aggregation by Petropoulos and Kourentzes (2015), and the k nearest neighbour (kNN) based approach by Nikolopoulos et al. (2016). See Section 2.8 for more details on intermittent demand forecasting and Section 2.10.2 for a discussion on temporal aggregation.
2. The emergence of new products. Recent studies on new product demand forecasting are based on finding analogies (Hu, Acimovic, Erize, Thomas, & Van Mieghem, 2019; Wright & Stern, 2015), leveraging comparable products (Baardman, Levin, Perakis, & Singhvi, 2018), and using external information like web search trends (Kulkarni, Kannan, & Moe, 2012). See Section 3.2.6 for more details on new product demand forecasting.
3. The existence of short-life-cycle products, e.g., smartphone demand (e.g., Chung, Niu, & Sriskandarajah, 2012; Shi, Yin, Cai, Cichocki, Yokota, Chen, Yuan, & Zeng, 2020; Szozda, 2010).
4. The hierarchical structure of the data such as the electricity demand mapped to a geographical hierarchy (e.g., Athanasopoulos et al., 2009; Hong et al., 2019; Hyndman et al., 2011, but also Section 2.10.1).

With the advent of the big data era, a couple of coexisting new challenges have drawn the attention of researchers and practitioners in the forecasting community: the need to forecast a large volume of related time series (e.g., thousands or millions of products from one large retailer: Salinas, Michael et al., 2019), and the increasing number of external variables that have significant influence on future demand (e.g., massive amounts of keyword search indices that could impact future tourism demand (Law, Li, Fong, & Han, 2019)). Recently, to deal with these new challenges, numerous empirical studies have identified the potentials of deep learning based global models, in both point and probabilistic demand forecasting (e.g., Bandara, Bergmeir, & Smyl, 2020b; Rangapuram et al., 2018; Salinas, Michael et al., 2019; Wen et al., 2017). With the merits of cross-learning, global models have been shown to be able to learn long memory patterns and related effects (Montero-Manso & Hyndman, 2020), latent correlation across multiple series (Smyl, 2020), handle complex real-world forecasting situations such as data sparsity and cold-starts (Chen, Kang, Chen, & Wang, 2020), include exogenous covariates such as promotional information and keyword search indices (Law et al., 2019), and allow for different choices of distributional assumptions (Salinas, Michael et al., 2019).

3.2.2. Forecasting in the supply chain[93]

A supply chain is 'a network of stakeholders (e.g., retailers, manufacturers, suppliers) who collaborate to satisfy customer demand' (Perera et al., 2019). Forecasts inform many supply chain decisions, including those relating to inventory control, production planning, cash flow management, logistics and human resources (also see Section 3.2.1). Typically, forecasts are based on an amalgam of statistical methods and management judgment (Fildes & Goodwin, 2007). Hofmann and Rutschmann (2018) have investigated the potential for using big data analytics in supply chain forecasting but indicate more research is needed to establish its usefulness.

In many organisations forecasts are a crucial element of Sales and Operations Planning (S&OP), a tool that brings together different business plans, such as those relating to sales, marketing, manufacturing and finance, into one integrated set of plans (Thomé, Scavarda, Fernandez, & Scavarda, 2012). The purposes of S&OP are to balance supply and demand and to link an organisation's operational and strategic plans. This requires collaboration between individuals and functional areas at different levels because it involves data sharing and achieving a consensus on forecasts and common objectives (Mello, 2010). Successful implementations of S&OP are therefore associated with forecasts that are both aligned with an organisation's needs and able to draw on information from across the organisation. This can be contrasted with the 'silo culture' identified in a survey of companies by Moon, Mentzer, and Smith (2003) where separate forecasts were prepared by different departments in 'islands of analysis'. Methods for reconciling forecasts at different levels in both cross-sectional hierarchies (e.g., national, regional and local forecasts) and temporal hierarchies (e.g., annual, monthly and daily forecasts) are also emerging as an approach to break through information silos in organisations (see Section 2.10.1, Section 2.10.2, and Section 2.10.3). Cross-temporal reconciliation provides a data-driven approach that allows information to be drawn from different sources and levels of the hierarchy and enables this to be blended into coherent forecasts (Kourentzes & Athanasopoulos, 2019).

In some supply chains, companies have agreed to share data and jointly manage planning processes in an initiative known as Collaborative Planning, Forecasting, and Replenishment (CPFR) (Seifert, 2003, also see Section 3.2.3). CPFR involves pooling information on inventory levels and on forthcoming events, like sales promotions. Demand forecasts can be shared, in real time via the Internet, and discrepancies between them reconciled. In theory, information sharing should reduce forecast errors. This should mitigate the 'bullwhip effect' where forecast errors at the retail-end of supply chains cause upstream suppliers to experience increasingly volatile demand, forcing them to hold high safety stock levels (Lee et al., 2007). Much research demonstrating the benefits of collaboration has involved simulated supply chains (Fildes, 2017). Studies of real companies have also found improved performance through collaboration (e.g., Boone & Ganeshan, 2008; Eksoz, Mansouri, Bourlakis, & Önkal, 2019; Hill, Zhang, & Miller, 2018), but case study evidence is still scarce (Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016). The implementation of collaborative schemes has been slow with many not progressing beyond the pilot stage (Galbreth, Kurtuluş, & Shor, 2015; Panahifar, Byrne, & Heavey, 2015). Barriers to successful implementation include a lack of trust between organisations, reward systems that foster a silo mentality, fragmented forecasting

[93] This subsection was written by Paul Goodwin.

systems within companies, incompatible systems, a lack of relevant training and the absence of top management support (Fliedner, 2003; Thomé, Hollmann, & Scavarda do Carmo, 2014).

Initiatives to improve supply chain forecasting can be undermined by political manipulation of forecasts and gaming. Examples include ‘enforcing’: requiring inflated forecasts to align them with sales or financial goals, ‘sandbagging’: underestimating sales so staff are rewarded for exceeding forecasts, and ‘spinning’: manipulating forecasts to garner favourable reactions from colleagues (Mello, 2009). Pennings, van Dalen, and Rook (2019) discuss schemes for correcting such intentional biases.

For a discussion of the forecasting of returned items in supply chains, see Section 3.2.9, while Section 3.9 offers a discussion of possible future developments in supply chain forecasting.

3.2.3. *Forecasting for inventories*[94]

Three aspects of the interaction between forecasting and inventory management have been studied in some depth and are the subject of this review: the bullwhip effect, forecast aggregation, and performance measurement.

The ‘bullwhip effect’ occurs whenever there is amplification of demand variability through the supply chain (Lee, Padmanabhan, & Whang, 2004), leading to excess inventories. This can be addressed by supply chain members sharing downstream demand information, at stock keeping unit level, to take advantage of less noisy data. Analytical results on the translation of ARIMA (see Section 2.3.4) demand processes have been established for order-up-to inventory systems (Gilbert, 2005). There would be no value in information sharing if the wholesaler can use such relationships to deduce the retailer’s demand process from their orders (see, for example, Graves, 1999). Such deductions assume that the retailer’s demand process and demand parameters are common knowledge to supply chain members. Ali and Boylan (2011) showed that, if such common knowledge is lacking, there is value in sharing the demand data itself and Ali, Boylan, and Syntetos (2012) established relationships between accuracy gains and inventory savings. Analytical research has tended to assume that demand parameters are known. Pastore, Alfieri, Zotteri, and Boylan (2020) investigated the impact of demand parameter uncertainty, showing how it exacerbates the bullwhip effect.

Forecasting approaches have been developed that are particularly suitable in an inventory context, even if not originally proposed to support inventory decisions. For example, Nikolopoulos et al. (2011) proposed that forecasts could be improved by aggregating higher frequency data into lower frequency data (see also Section 2.10.2; other approaches are reviewed in Section 3.2.1). Following this approach, Forecasts are generated at the lower frequency level and then disaggregated, if required, to the higher frequency level. For inventory replenishment decisions, the level of aggregation may conveniently be chosen to be the lead time, thereby taking advantage of the greater stability of data at the lower frequency level, with no need for disaggregation.

The variance of forecast errors over lead time is required to determine safety stock requirements for continuous review systems. The conventional approach is to take the variance of one-step-ahead errors and multiply it by the lead time. However, this estimator is unsound, even if demand is independent and identically distributed, as explained by Prak, Teunter, and Syntetos (2017). A more direct approach is to smooth the mean square errors over the lead time (Syntetos & Boylan, 2006).

Strijbosch and Moors (2005) showed that unbiased forecasts will not necessarily lead to achievement, on average, of target cycle service levels or fill rates. Wallström and Segerstedt (2010) proposed a ‘Periods in Stock’ measure, which may be interpreted, based on a ‘fictitious stock’, as the number of periods a unit of the forecasted item has been in stock or out of stock. Such measures may be complemented by a detailed examination of error-implication metrics (Boylan & Syntetos, 2006). For inventory management, these metrics will typically include inventory holdings and service level implications (e.g., cycle service level, fill rate). Comparisons may be based on total costs or via ‘exchange curves’, showing the trade-offs between service and inventory holding costs. Comparisons such as these are now regarded as standard in the literature on forecasting for inventories and align well with practice in industry.

3.2.4. *Forecasting in retail*[95]

Retail companies depend crucially on accurate demand forecasting to manage their supply chain and make decisions concerning planning, marketing, purchasing, distribution and labour force. Inaccurate forecasts lead to unnecessary costs and poor customer satisfaction. Inventories should be neither too high (to avoid waste and extra costs of storage and labour force), nor too low (to prevent stock-outs and lost sales, Ma & Fildes, 2017).

Forecasting retail demand happens in a three-dimensional space (Syntetos et al., 2016): the position in the supply chain hierarchy (store, distribution centre, or chain), the level in the product hierarchy (SKU, brand, category, or total) and the time granularity (day, week, month, quarter, or year). In general, the higher is the position in the supply chain, the lower is the time granularity required, e.g., retailers need daily forecasts for store replenishment and weekly forecasts for DC distribution/logistics activities at the SKU level (Fildes, Ma, & Kolassa, 2019b). Hierarchical forecasting (see Section 2.10.1) is a promising tool to generate coherent demand forecasts on multiple levels over different dimensions (Oliveira & Ramos, 2019).

Several factors affect retail sales, which often increase substantially during holidays, festivals, and other special events. Price reductions and promotions on own and competitors’ products, as well as weather conditions or pandemics, can also change sales considerably (Huang, Fildes, & Soopramanien, 2019).

[94] This subsection was written by John E. Boylan.

[95] This subsection was written by Stephan Kolassa & Patrícia Ramos.

Zero sales due to stock-outs or low demand occur very often at the SKU × store level, both at weekly and daily granularity. The most appropriate forecasting approaches for intermittent demand are Croston's method (Croston, 1972), the Syntetos-Boylan approximation (SBA; Syntetos & Boylan, 2005), and the TSB method (Teunter, Syntetos, & Zied Babai, 2011), all introduced in Section 2.8.1. These methods have been used to forecast sales of spare parts in automotive and aerospace industries but have not yet been evaluated in the retail context.

Univariate forecasting models are the most basic methods retailers may use to forecast demand. They range from simple methods such as simple moving averages or exponential smoothing to ARIMA and ETS models (discussed in Section 2.3). These are particularly appropriate to forecast demand at higher aggregation levels (Ramos & Oliveira, 2016; Ramos, Santos, & Rebelo, 2015). The main advantage of linear causal methods such as multiple linear regression is to allow the inclusion of external effects discussed above. There is no clear evidence yet that nonlinear models and novel machine learning methods can improve forecast accuracy (Fildes et al., 2019b).

To be effective, point estimates should be combined with quantile predictions or prediction intervals for determining safety stock amounts needed for replenishment. However, to the best of our knowledge this is an under-investigated aspect of retail forecasting (Kolassa, 2016; Taylor, 2007).

The online channel accounts for an ever-increasing proportion of retail sales and poses unique challenges to forecasting, beyond the characteristics of brick and mortar (B&M) retail stores. First, there are multiple *drivers or predictors* of demand that could be leveraged in online retail, but not in B&M:

- Online retailers can fine-tune customer interactions, e.g., through the landing page, product recommendations, or personalised promotions, leveraging the customer's purchasing, browsing or returns history, current shopping cart contents, or the retailer's stock position, in order to tailor the message to one specific customer in a way that is impossible in B&M.
- Conversely, product reviews are a type of interaction between the customer and the retailer and other customers which drives future demand.

Next, there are differences in forecast *use*:

- Forecast use strongly depends on the retailer's omnichannel strategy (Armstrong, 2017; Melacini, Perotti, Rasini, & Tappia, 2018; Sopadjieva, Dholakia, & Benjamin, 2017): e.g., for "order online, pick up in store" or "ship from store" fulfillment, we need separate but related forecasts for both total online demand and for the demand fulfilled at each separate store.
- Online retailers, especially in fashion, have a much bigger problem with product returns. They may need to forecast how many products are returned overall (e.g., Shang, McKie, Ferguson, & Galbreth, 2020), or whether a *specific* customer will return a *specific* product.

Finally, there are differences in the forecasting *process*:

- B&M retailers decouple pricing/promotion decisions and optimisation from the customer interaction, and therefore from forecasting. Online, this is not possible, because the customer has total transparency to competitors' offerings. Thus, online pricing needs to react much more quickly to competitive pressures – faster than the forecasting cycle.
- Thus, the specific value of predictors is often not known at the time of forecasting: we don't know yet which customer will log on, so we don't know yet how many people will see a particular product displayed on their personalised landing page. (Nor do we know today what remaining stock will be displayed.) Thus, changes in drivers need to be "baked into" the forecasting algorithm.
- Feedback loops between forecasting and other processes are thus even more important online: yesterday's forecasts drive today's stock position, driving today's personalised recommendations, driving demand, driving today's forecasts for tomorrow. Overall, online retail forecasting needs to be more agile and responsive to the latest interactional decisions taken in the web store, and more tightly integrated into the retailer's interactional tactics and omnichannel strategy.

Systematic research on demand forecasting in an online or omnnichannel context is only starting to appear (e.g. Omar, Klibi, Babai, & Ducq, 2021, who use basket data from online sales to improve omnichannel retail forecasts).

3.2.5. *Promotional forecasting*[96]

Promotional forecasting is central for retailing (see Section 3.2.4), but also relevant for many manufacturers, particularly of Fast Moving Consumer Goods (FMCG). In principle, the objective is to forecast sales, as in most business forecasting cases. However, what sets promotional forecasting apart is that we also make use of information about promotional plans, pricing, and sales of complementary and substitute products (Bandyopadhyay, 2009; Zhang, Chen, & Lee, 2008). Other relevant variables may include store location and format, variables that capture the presentation and location of a product in a store, proxies that characterise the competition, and so on (Andrews, Currim, Leeflang, & Lim, 2008; Van Heerde, Leeflang, & Wittink, 2002).

Three modelling considerations guide us in the choice of models. First, promotional (and associated effects) are proportional. For instance, we do not want to model the increase in sales as an absolute number of units, but instead, as a percentage uplift. We do this to not only make the model applicable to both smaller and larger applications, for example, small and large stores in a retailing chain, but also to gain a clearer insight into the behaviour of our customers. Second, it is common that there are synergy effects. For example, a promotion for a product

[96] This subsection was written by Nikolaos Kourentzes.

may be offset by promotions for substitute products. Both these considerations are easily resolved if we use multiplicative regression models. However, instead of working with the multiplicative models, we rely on the logarithmic transformation of the data (see Section 2.2.1) and proceed to construct the promotional model using the less cumbersome additive formulation (see Section 2.3.2). Third, the objective of promotional models does not end with providing accurate predictions. We are also interested in the effect of the various predictors: their elasticity. This can in turn provide the users with valuable information about the customers, but also be an input for constructing optimal promotional and pricing strategies (Zhang et al., 2008).

Promotional models have been widely used on brand-level data (for example, Divakar, Ratchford, & Shankar, 2005). However, they are increasingly used on Stock Keeping Unit (SKU) level data (Ma, Fildes, & Huang, 2016; Trapero, Kourentzes, & Fildes, 2015), given advances in modelling techniques. Especially at that level, limited sales history and potentially non-existing examples of past promotions can be a challenge. Trapero et al. (2015) consider this problem and propose using a promotional model that has two parts that are jointly estimated. The first part focuses on the time series dynamics and is modelled locally for each SKU. The second part tackles the promotional part, which pools examples of promotions across SKUs to enable providing reasonable estimates of uplifts even for new SKUs. To ensure the expected heterogeneity in the promotional effects, the model is provided with product group information. Another recent innovation is looking at modelling promotional effects both at the aggregate brand or total sales level, and disaggregate SKU level, relying on temporal aggregation (Kourentzes & Petropoulos, 2016, and Section 2.10.2). Ma et al. (2016) concern themselves with the intra-and inter-category promotional information. The challenge now is the number of variables to be considered for the promotional model, which they address by using sequential LASSO (see also Section 2.5.3). Although the aforementioned models have shown very promising results, one has to recognise that in practice promotions are often forecasted using judgmental adjustments, with inconsistent performance (Trapero, Pedregal, Fildes, & Kourentzes, 2013); see also Section 2.11.2 and Section 3.7.3.

3.2.6. New product forecasting[97]

Forecasting the demand for a new product accurately has even more consequence with regards to well-being of the companies than that for a product already in the market. However, this task is one of the most difficult tasks managers must deal with simply because of non-availability of past data (Wind, 1981). Much work has been going on for the last five decades in this field. Despite his Herculean attempt to collate the methods reported, Assmus (1984) could not list all even at that time. The methods used before and since could be categorised into three broad approaches (Goodwin, Dyussekeneva & Meeran, 2013) namely management judgment, consumer judgment and diffusion/formal mathematical models. In general, the hybrid methods combining different approaches have been found to be more useful (Hyndman & Athanasopoulos, 2018; Peres et al., 2010). Most of the attempts in New product Forecasting (NPF) have been about forecasting 'adoption' (i.e., enumerating the customers who bought at least one time) rather than 'sales', which accounts for repeat purchases also. In general, these attempts dealt with point forecast although there have been some attempts in interval and density forecasting (Meade & Islam, 2001).

Out of the three approaches in NPF, management judgment is the most used approach (Gartner & Thomas, 1993; Kahn, 2002; Lynn, Schnaars, & Skov, 1999) which is reported to have been carried out by either individual managers or group of them. Ozer (2011) and Surowiecki (2005) articulated their contrasting benefits and deficits. The Delphi method (see Section 2.11.4) has combined the benefits of these two modes of operation (Rowe & Wright, 1999) which has been effective in NPF. Prediction markets in the recent past offered an alternative way to aggregate forecasts from a group of Managers (Meeran, Dyussekeneva, & Goodwin, 2013; Wolfers & Zitzewitz, 2004) and some successful application of prediction markets for NPF have been reported by Karniouchina (2011) and Plott and Chen (2002).

In the second category, customer surveys, among other methods, are used to directly ask the customers the likelihood of them purchasing the product. Such surveys are found to be not very reliable (Morwitz, 1997). An alternative method to avoid implicit bias associated with such surveys in extracting inherent customer preference is conjoint analysis, which makes implicit trade off customers make between features explicit by analysing the customers' preference for different variants of the product. One analysis technique that attempts to mirror real life experience more is Choice Based Conjoint analysis (CBC) in which customers choose the most preferred product among available choices. Such CBC models used together with the analysis tools such as Logit (McFadden, 1977) have been successful in different NPF applications (Meeran, Jahanbin, Goodwin, & Quariguasi Frota Neto, 2017).

In the third approach, mathematical/formal models known as growth or diffusion curves (see Section 2.3.18 and Section 2.3.19) have been used successfully to do NPF (Hu et al., 2019). The non-availability of past data is mitigated by growth curves by capturing the generic pattern of the demand growth of a class of products, which could be defined by a limited number of parameters such as saturation level, inflexion point, etc. For a new product a growth curve can be constituted from well-estimated parameters using analogous products, market intelligence or regression methods. Most extensively used family of growth curves for NPF has started with Bass model (Bass, 1969) that has been extended extensively (Bass, Gordon, Ferguson, & Githens, 2001; Easingwood, Mahajan, & Muller, 1983; Islam & Meade, 2000; Peres et al., 2010; Simon & Sebastian, 1987). A recent applications of NPF focused on consumer electronic goods using analogous products (Goodwin, Dyussekeneva et al., 2013).

[97] This subsection was written by Sheik Meeran.

3.2.7. Spare parts forecasting[98]

Spare parts are ubiquitous in modern societies. Their demand arises whenever a component fails or requires replacement. Demand for spare parts is typically intermittent, which means that it can be forecasted using the plethora of parametric and non-parametric methods presented in Section 2.8. In addition to the intermittence of demand, spare parts have two additional characteristics that make them different from Work-In-Progress and final products, namely: (*i*) they are generated by maintenance policies and part breakdowns, and (*ii*) they are subject to obsolescence (Bacchetti & Saccani, 2012; Kennedy, Wayne Patterson, & Fredendall, 2002).

The majority of forecasting methods do not link the demand to the generating factors, which are often related to maintenance activities. The demand for spare parts originates from the replacement of parts in the installed base of machines (i.e., the location and number of products in use), either preventively or upon breakdown of the part (Kim, Dekker, & Heij, 2017). Fortuin (1984) claims that using installed base information to forecast the spare part demand can lead to stock reductions of up to 25%. An overview of the literature that deals with spare parts forecasting with installed base information is given by Van der Auweraer and Boute (2019). Spare parts demand can be driven by the result of maintenance inspections and, in this case, a maintenance-based forecasting model should then be considered to deal with this issue. Such forecasting models include the Delay Time (DT) model analysed in Wang and Syntetos (2011). Using the fitted values of the distribution parameters of a data set related to a hospital pumps, Wang and Syntetos (2011) have shown that when the failure and fault arriving characteristics of the items can be captured, it is recommended to use the DT model to forecast the spare part demand with a higher forecast accuracy. However, when such information is not available, then time series forecasting methods, such as those presented in Section 2.8.1, are recommended. The maintenance based forecasting is further discussed in Section 3.2.8.

Given the life cycle of products, spare parts are associated with a risk of obsolescence. Molenaers, Baets, Pintelon, and Waeyenbergh (2012) discussed a case study where 54% of the parts stocked at a large petrochemical company had seen no demand for the last 5 years. Hinton (1999) reported that the US Department of Defence was holding 60% excess of spare parts, with 18% of the parts (with a total value of $1.5 billion) having no demand at all. To take into account the issue of obsolescence in spare parts demand forecasting, Teunter et al. (2011) have proposed the TSB method, which deals with linearly decreasing demand and sudden obsolescence cases. By means of an empirical investigation based on the individual demand histories of 8000 spare parts SKUs from the automotive industry and the Royal Air Force (RAF, UK), Babai, Syntetos, and Teunter (2014) have demonstrated the high forecast accuracy and inventory performance of the TSB method. Other variants of the Croston's method developed to deal with the risk of obsolescence in forecasting spare parts demand include the Hyperbolic-Exponential Smoothing method proposed by Prestwich, Tarim, Rossi, and Hnich (2014) and the modified Croston's method developed by Babai, Dallery, Boubaker, and Kalai (2019).

3.2.8. Predictive maintenance[99]

A common classification of industrial maintenance includes three types of maintenance (Montero Jimenez, Schwartz, Vingerhoeds, Grabot, & Salaün, 2020). Corrective maintenance refers to maintenance actions that occur after the failure of a component. Preventive maintenance consists of maintenance actions that are triggered after a scheduled number of units as cycles, kilometers, flights, etc. To schedule the fixed time between two preventive maintenance actions, the Weibull distribution is commonly used (Baptista, Sankararaman, de Medeiros, Nascimento, Prendinger, & Henriques, 2018). The drawbacks of preventive maintenance are related to the replacement of components that still have a remaining useful life; therefore, early interventions imply a waste of resources and too late actions could imply catastrophic failures. Additionally, the preventive intervention itself could be a source of failures too. Finally, predictive maintenance (PdM) complements the previous ones and, essentially, uses predictive tools to determine when actions are necessary (Carvalho, Soares, Vita, Francisco, Basto, & Alcalá, 2019). Within this predictive maintenance group, other terms are usually found in the literature as Condition-Based Maintenance and Prognostic and Health Management, (Montero Jimenez et al., 2020).

The role of forecasting in industrial maintenance is of paramount importance. One application is to forecast spare parts (see Section 3.2.7), whose demands are typically intermittent, usually required to carry out corrective and preventive maintenances (Van der Auweraer & Boute, 2019; Wang & Syntetos, 2011). On the other hand, it is crucial for PdM the forecast of the remaining useful time, which is the useful life left on an asset at a particular time of operation (Si, Wang, Hu, & Zhou, 2011). This work will be focused on the latter, which is usually found under the prognostic stage (Jardine, Lin, & Banjevic, 2006).

The typology of forecasting techniques employed is very ample. Montero Jimenez et al. (2020) classify them in three groups: physics-based models, knowledge-based models, and data-driven models. Physics-based models require high skills on the underlying physics of the application. Knowledge-based models are based on facts or cases collected over the years of operation and maintenance. Although, they are useful for diagnostics and provide explicative results, its performance on prognostics is more limited. In this sense, data-driven models are gaining popularity for the development of computational power, data acquisition, and big data platforms. In this case, data coming from vibration analysis, lubricant analysis, thermography, ultrasound, etc. are usually employed. Here, well-known forecasting models as VARIMAX/GARCH (see also Section 2.3) are successfully used (Baptista et al., 2018; Cheng, Yu, & Chen, 2012;

[98] This subsection was written by Mohamed Zied Babai.

[99] This subsection was written by Juan Ramón Trapero Arenas.

García, Pedregal, & Roberts, 2010; Gomez Munoz, De la Hermosa Gonzalez-Carrato, Trapero Arenas, & Garcia Marquez, 2014). State Space models based on the Kalman Filter are also employed (Pedregal & Carmen Carnero, 2006; Pedregal, García, & Roberts, 2009, and Section 2.3.6). Recently, given the irruption of the Industry 4.0, physical and digital systems are getting more integrated and Machine Learning/Artificial Intelligence are drawing the attention of practitioners and academics alike (Carvalho et al., 2019). In that same reference, it is found that the most frequently used Machine Learning methods in PdM applications were Random Forest, Artificial Neural Networks, Support Vector Machines and K-means.

3.2.9. Reverse logistics[100]

As logistics and supply chain operations rely upon accurate demand forecasts (see also Section 3.2.2), reverse logistics and closed loop supply chain operations rely upon accurate forecasts of returned items. Such items (usually referred as cores) can be anything from reusable shipping or product containers to used laptops, mobile phones or car engines. If some (re)manufacturing activity is involved in supply chains, it is both demand and returned items forecasts that are needed since it is net demand requirements (demand – returns) that drive remanufacturing operations.

Forecasting methods that are known to work well when applied to demand forecasting, such as SES for example (see Section 2.3.1), do not perform well when applied to time-series of returns because they assume returns to be a process independent of sales. There are some cases when this independence might hold, such as when a recycler receives items sold by various companies and supply chains (Goltsos & Syntetos, 2020). In these cases, simple methods like SES applied on the time series of returns might prove sufficient. Typically though, returns are strongly correlated with past sales and the installed base (number of products with customers). After all, there cannot be a product return if a product has not first been sold. This lagged relationship between sales and returns is key to the effective characterisation of the returns process.

Despite the increasing importance of circular economy and research on closed loop supply chains, returns forecasting has not received sufficient attention in the academic literature (notable contributions in this area include Clottey, Benton Jr., & Srivastava, 2012; de Brito & van der Laan, 2009; Goh & Varaprasad, 1986; Toktay, 2003; Toktay, Wein, & Zenios, 2000). The seminal work by Kelle and Silver (1989) offers a useful framework to forecasting that is based on the degree of available information about the relationship between demand and returns. Product level (PL) information consists of the time series of sales and returns, alongside information on the time each product spends with a customer. The question then is how to derive this time to return distribution. This can be done through managerial experience, by investigating the correlation of the demand and the returns time series, or by serialising and tracking a subset (sample) of items. Past sales can then be used in conjunction with this distribution to create forecasts of returns. Serial number level (SL) information, is more detailed and consists of the time matching of an individual unit item's issues and returns and thus exactly the time each individual unit, on a serial number basis, spent with the customer. Serialisation allows for a complete characterisation of the time to return distribution. Very importantly, it also enables tracking exactly how many items previously sold remain with customers, providing time series of unreturned past sales. Unreturned past sales can then be extrapolated — along with a time to return distribution — to create forecasts of returns.

Goltsos, Syntetos, and van der Laan (2019) offered empirical evidence in the area of returns forecasting by analysing a serialised data set from a remanufacturing company in North Wales. They found the Beta probability distribution to best fit times-to-return. Their research suggests that serialisation is something worthwhile pursuing for low volume products, especially if they are expensive. This makes a lot of sense from an investment perspective, since the relevant serial numbers are very few. However, they also provided evidence that such benefits expand in the case of high volume items. Importantly, the benefits of serialisation not only enable the implementation of the more complex SL method, but also the accurate characterisation of the returns process, thus also benefiting the PL method (which has been shown to be very robust).

3.3. Economics and finance

3.3.1. Macroeconomic survey expectations[101]

Macroeconomic survey expectations allow tests of theories of how agents form their expectations. Expectations play a central role in modern macroeconomic research (Gali, 2008). Survey expectations have been used to test theories of expectations formation for the last 50 years. Initially the Livingston survey data on inflationary expectations was used to test extrapolative or adaptive hypothesis, but the focus soon turned to testing whether expectations are formed rationally (see Turnovsky and Wachter, 1972, for an early contribution). According to Muth (1961, p. 316), rational expectations is the hypothesis that: 'expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory.' This assumes all agents have access to all relevant information. Instead, one can test whether agents make efficient use of the information they possess. This is the notion of forecast efficiency (Mincer & Zarnowitz, 1969), and can be tested by regressing the outturns on a constant and the forecasts of those outturns. Under forecast efficiency, the constant should be zero and the coefficient on the forecasts should be one. When the slope coefficient is not equal to one, the forecast errors will be systematically related to information available at the forecast origin, namely, the forecasts, and cannot be optimal. The exchange between Figlewski and Wachtel (1981, 1983) and Dietrich and Joines (1983)

[100] This subsection was written by Aris A. Syntetos.

[101] This subsection was written by Michael P. Clements.

clarifies the role of partial information in testing forecast efficiency (that is, full information is not necessary), and shows that the use of the aggregate or consensus forecast in the individual realisation-forecast regression outlined above will give rise to a slope parameter less than one when forecasters are efficient but possess partial information. Zarnowitz (1985), Keane and Runkle (1990) and Bonham and Cohen (2001) consider pooling across individuals in the realisation-forecast regression, and the role of correlated shocks across individuals.

Recently, researchers considered why forecasters might not possess full-information, stressing informational rigidities: sticky information (see, *inter alia*, Mankiw & Reis, 2002; Mankiw, Reis, & Wolfers, 2003), and noisy information (see, *inter alia*, Sims, 2003; Woodford, 2002). Coibion and Gorodnichenko (2012, 2015) test these models using aggregate quantities, such as mean errors and revisions.

Forecaster behaviour can be characterised by the response to new information (see also Section 2.11.1). Over or under-reaction would constitute inefficiency. Broer and Kohlhas (2018) and Bordalo, Gennaioli, Ma, and Shleifer (2018) find that agents over-react, generating a negative correlation between their forecast revision and error. The forecast is revised by more than is warranted by the new information (over-confidence regarding the value of the new information). Bordalo et al. (2018) explain the over-reaction with a model of 'diagnostic' expectations, whereas Fuhrer (2018) finds 'intrinsic inflation persistence': individuals under-react to new information, smoothing their responses to news.

The empirical evidence is often equivocal, and might reflect: the vintage of data assumed for the outturns; whether allowance is made for 'instabilities' such as alternating over- and under-prediction (Rossi & Sekhposyan, 2016) and the assumption of squared-error loss (see, for example, Clements, 2014b; Patton & Timmermann, 2007).

Research has also focused on the histogram forecasts produced by a number of macro-surveys. Density forecast evaluation techniques such as the probability integral transform[102] have been applied to histogram forecasts, and survey histograms have been compared to benchmark forecasts (see, for example, Bao, Lee, & Saltoglu, 2007; Clements, 2018; Hall & Mitchell, 2009). Research has also considered uncertainty measures based on the histograms (Clements, 2014a). Sections 2.12.4 and 2.12.5 also discuss the evaluation and reliability of probabilistic forecasts.

Clements (2009, 2010) and Engelberg, Manski, and Williams (2009) considered the consistency between the point predictions and histogram forecasts. Reporting practices such as 'rounding' have also been considered (Binder, 2017; Clements, 2011; Manski & Molinari, 2010).

Clements (2019) reviews macroeconomic survey expectations.

[102] See, for example, Rosenblatt (1952), Shephard (1994), Kim, Shephard, and Chib (1998), Diebold et al. (1998) and Berkowitz (2001).

3.3.2. Forecasting GDP and inflation[103]

As soon as Bayesian estimation of DSGEs became popular, these models have been employed in forecasting horseraces to predict the key macro variables, for example, Gross Domestic Product (GDP) and inflation, as discussed in Del Negro and Schorfheide (2013). The forecasting performance is evaluated using rolling or recursive (expanded) prediction windows (for a discussion, see Cardani, Paccagnini, & Villa, 2015). DSGEs are usually estimated using revised data, but several studies propose better results estimating the models using real-time data (see, for example, Cardani et al., 2019; Del Negro & Schorfheide, 2013; Kolasa & Rubaszek, 2015b; Wolters, 2015).

The current DSGE model forecasting compares DSGE models to competitors (see Section 2.3.15 for an introduction to DSGE models). Among them, we can include the Random Walk (the naive model which assumes a stochastic trend), the Bayesian VAR models (Minnesota Prior à la Doan, Litterman, and Sims, 1984; and Large Bayesian VAR à la Bańbura, Giannone, and Reichlin, 2010), the Hybrid-Models (the DSGE-VAR à la Del Negro and Schorfheide, 2004; and the DSGE-Factor Augmented VAR à la Consolo, Favero, and Paccagnini, 2009), and the institutional forecasts (Greenbook, Survey Professional Forecasts, and the Blue Chip, as illustrated in Edge and Gürkaynak, 2010).

Table 2 summarises the current DSGE forecasting literature mainly for the US and Euro Area and provided by estimating medium-scale models. As general findings, DSGEs can outperform other competitors, with the exception for the Hybrid-Models, in the medium and long-run to forecast GDP and inflation. In particular, Smets and Wouters (2007) was the first empirical evidence of how DSGEs can be competitive with forecasts from Bayesian VARs, convincing researchers and policymakers in adopting DSGEs for prediction evaluations. As discussed in Del Negro and Schorfheide (2013), the accuracy of DSGE forecasts depends on how the model is able to capture low-frequency trends in the data. To explain the macro-finance linkages during the Great Recession, the Smets and Wouters model was also compared to other DSGE specifications including the financial sector. For example, Cardani et al. (2019), Del Negro and Schorfheide (2013), Galvão, Giraitis, Kapetanios, and Petrova (2016), and Kolasa and Rubaszek (2015a) provide forecasting performance for DSGEs with financial frictions. This strand of the literature shows how this feature can improve the baseline Smets and Wouters predictions for the business cycle, in particular during the recent Great Recession.

However, the Hybrid-Models always outperform the DSGEs thanks to the combination of the theory-based model (DSGE) and the statistical representation (VAR or Factor Augmented VAR), as illustrated by Del Negro and Schorfheide (2004) and Consolo et al. (2009).

Moreover, several studies discuss how prediction performance could depend on the parameters' estimation. Kolasa and Rubaszek (2015b) suggest that updating DSGE model parameters only once a year is enough to have accurate and efficient predictions about the main macro variables.

[103] This subsection was written by Alessia Paccagnini.

Table 2
Alternative competitors to DSGE models.

Competitor	Reference
Hybrid Models	US: Del Negro and Schorfheide (2004), Consolo et al. (2009)
Random Walk	US: Gürkaynak, Kısacıkoğlu, and Rossi (2013), Euro Area: Warne et al. (2010), Smets, Warne, and Wouters (2014)
Bayesian VAR	US: Smets and Wouters (2007), Gürkaynak et al. (2013), Wolters (2015), Bekiros and Paccagnini (2014), Bekiros and Paccagnini (2015a), Bekiros and Paccagnini (2015b), Euro Area: (Warne et al., 2010)
Time-Varying VAR and Markov-Switching	US: Bekiros, Cardani, Paccagnini, and Villa (2016), Euro Area: Bekiros and Paccagnini (2016)
Institutional Forecasts	US: Edge and Gürkaynak (2010), Kolasa et al. (2012), Del Negro and Schorfheide (2013), Wolters (2015)

3.3.3. *Forecasting unemployment*[104]

Unemployment has significant implications at both the micro and macro levels, influencing individual living standards, health and well-being, as well as imposing direct costs on the economy. Given its importance, policy-makers put unemployment at the heart of their economic plans, and as such require accurate forecasts to feed into economic policy decisions. Unemployment is described as a lagging indicator of the economy, with characteristics including business cycles and persistence. Despite this, forecasting the unemployment rate is difficult, because the data are highly non-stationary with abrupt distributional shifts, but persistence within regimes. In this section we focus on methods used to forecast the aggregate unemployment rate.

Unemployment is the outcome of supply and demand for labour, aggregated across all prospective workers, with labour demand derived from demand for goods and services. This implies a highly complex data generating process. Empirical forecasting models tend to simplify this relationship, with two approaches dominating the literature. The first is based on the Phillips (1958) curve capturing a non-linear relationship between nominal wage inflation and the unemployment rate, or the relation between unemployment and output described as Okun's 1962 Law. The second uses the time-series properties of the data to produce statistical forecasts, such as univariate linear models (for example, ARIMA or unobserved component models; see Sections 2.3.4 and 2.3.6), multivariate linear models (for example, VARMA or CVAR; see Section 2.3.9), various threshold autoregressive models (see Section 2.3.13), Markov Switching models (see Section 2.3.12) and Artificial Neural Networks (see Section 2.7.8).

The empirical literature is inconclusive as to the 'best' forecasting models for unemployment, which varies by country, time period and forecast horizon. There is some evidence that non-linear statistical models tend to outperform within business cycle contractions or expansions, but perform worse across business cycles (see, for example, Koop & Potter, 1999; Montgomery, Zarnowitz, Tsay, & Tiao, 1998; Rothman, 1998), whereas Proietti (2003) finds that linear models characterised by higher persistence perform significantly better. Evidence of non-linearities is found by Johnes (1999), Milas and Rothman (2008) and Peel and Speight (2000), and Gil-Alana (2001) finds evidence of long-memory. Barnichon and Garda (2016) applies a flow approach to unemployment forecasting and finds improvements, as does Smith (2011).

One approach that does yield accurate forecasts is to use a measure of profitability as the explanatory variable, assuming that unemployment will fall when hiring is profitable. Hendry (2001) proxies profitability (π) by the gap between the real interest rate (reflecting costs) and the real growth rate (reflecting the demand side), such that the unemployment rate rises when the real interest rate exceeds the real growth rate, and vice versa:

$$\pi_t = (R_L - \Delta p - \Delta y)_t$$

where R_L is the long-term interest rate, Δp is a measure of inflation and Δy is a measure of output growth. This is then embedded within a dynamic equilibrium correction model, using impulse indicator saturation (IIS: Hendry, Johansen, & Santos, 2008b; Johansen & Nielsen, 2009) and step indicator saturation (SIS: Castle et al., 2015c) to capture outliers, breaks and regime shifts, as well as allowing for any non-linearities using Taylor expansions for the regressors. The resulting forecasts perform well over the business cycle relative to alternative statistical models (also see Hendry, 2015, and Castle, Hendry, & Martinez, 2020c).

Forecasts from models of unemployment could be improved with either better economic theories of aggregate unemployment,[105] or more general empirical models that tackle stochastic trends, breaks, dynamics, non-linearities and interdependence,[106] or better still, both. The COVID-19 pandemic and subsequent lockdown policies highlight just how important forecasts of unemployment are (Castle, Doornik, & Hendry, 2021).

3.3.4. *Forecasting productivity*[107]

The growth of labour productivity, measured by the percent change in output per hours worked, has varied dramatically over the last 260 years. In the UK it ranged from −5.8% at the onset of the 1920 Depression to just over 7% in 1971; see panel A in Fig. 6. Productivity growth is very volatile and has undergone large historical shifts with productivity growth averaging around 1% between

[104] This subsection was written by Jennifer L. Castle.

[105] There are many relevant theories based on microfoundations, including search and matching, loss of skills, efficiency wages, and insider-outsider models, see Layard, Nickell, and Jackman (1991) for a summary.

[106] See Hendry and Doornik (2014) for an approach to jointly tackling all of these issues.

[107] This subsection was written by Andrew B. Martinez.

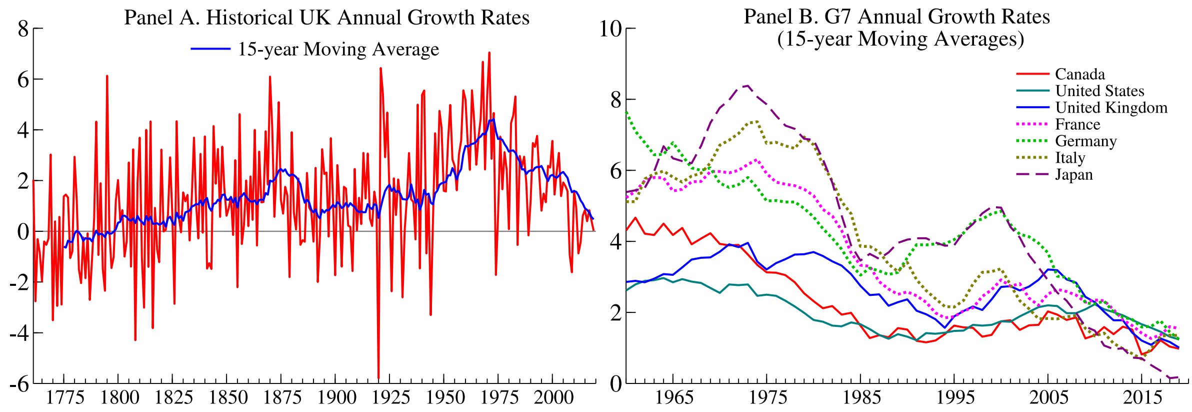

Fig. 6. Productivity Growth (Output per total hours worked).
Source: Bank of England and Penn World Table Version 10.0.

1800–1950 followed by an increase in the average annual growth to 3% between 1950–1975. Since the mid-1970's productivity growth has gradually declined in many developed economies; see panel B of Fig. 6. In the decade since 2009, 2% annual productivity growth was an upper bound for most G7 countries.

The most common approach for forecasting productivity is to estimate the trend growth in productivity using aggregate data. For example, Gordon (2003) considers three separate approaches for calculating trend labor productivity in the United States based on (*i*) average historical growth rates outside of the business cycle, (*ii*) filtering the data using the HP filter (Hodrick & Prescott, 1997), and (*iii*) filtering the data using the Kalman filter (see Kalman, 1960). The Office for Budget Responsibility (OBR) in the UK and the Congressional Budget Office (CBO) in the US follow similar approaches for generating its forecasts of productivity based on average historical growth rates as well as judgments about factors that may cause productivity to deviate from its historical trend in the short-term.[108] Alternative approaches include forecasting aggregate productivity using disaggregated firm-level data (see Bartelsman, Kurz, & Wolf, 2011; Bartelsman & Wolf, 2014, and Section 2.10.1) and using time-series models (see Žmuk, Dumičić, & Palić, 2018, and Section 2.3.4).

In the last few decades there have been several attempts to test for time-varying trends in productivity and to allow for them. However, the focus of these approaches has been primarily on the United States (Hansen, 2001; Roberts, 2001), which saw a sharp rise in productivity growth in the 1990's that was not mirrored in other countries (Basu, Fernald, Oulton, & Srinivasan, 2003). Test for shifts in productivity growth rates in other advanced economies did not find evidence of a changes in productivity growth until well after the financial crisis in 2007 (Benati, 2007; Glocker & Wegmüller, 2018; Turner & Boulhol, 2011).

A more recent approach by Martinez et al. (2021) allows for a time-varying long-run trend in UK productivity. They show that are able to broadly replicate the OBR's forecasts using a quasi-transformed autoregressive model with one lag, a constant, and a trend. The estimated long-run trend is just over 2% per year through 2007 Q4 which is consistent with the OBR's assumptions about the long-run growth rate of productivity (OBR, 2019). However, it is possible to dramatically improve upon OBR's forecasts in real-time by allowing for the long-term trend forecast to adjust based on more recent historical patterns. By taking a local average of the last four years of growth rates, Martinez et al. (2021) generate productivity forecasts whose RMSE is on average more than 75% smaller than OBR's forecasts extending five-years-ahead and is 84% smaller at the longest forecast horizon.

3.3.5. *Fiscal forecasting for government budget surveillance*[109]

Recent economic recessions have led to a renewed interest in fiscal forecasting, mainly for deficit and debt surveillance. This was certainly true in the case of the 2008 recession, and looks to become even more important in the current economic crisis brought on by the COVID-19 pandemic. This is particularly important in Europe, where countries are subject to strong fiscal monitoring mechanisms. Two main themes can be detected in the fiscal forecasting literature (Leal, Pérez, Tujula, & Vidal, 2008). First, investigate the properties of forecasts in terms of bias, efficiency and accuracy. Second, check the adequacy of forecasting procedures.

The first topic has its own interest for long, mainly restricted to international institutions (Artis & Marcellino, 2001). Part of the literature, however, argue that fiscal forecasts are politically biased, mainly because there is usually no clear distinction between political targets and rigorous forecasts (Frankel & Schreger, 2013; Strauch, Hallerberg, & Hagen, 2004). In this sense, the availability of forecasts from independent sources is of great value (Jonung & Larch, 2006). But it is not as easy as saying that independent forecasters would improve forecasts due to the absence of political bias, because forecasting accuracy is compromised by complexities of data,

[108] See https://obr.uk/forecasts-in-depth/the-economy-forecast/potential-output-and-the-output-gap. (Accessed: 2020-09-05)

[109] This subsection was written by Diego J. Pedregal.

country-specific factors, outliers, changes in the definition of fiscal variables, etc. Very often some of these issues are known by the staff of organisations in charge of making the official statistics and forecasts long before the general public, and some information never leaves such institutions. So this insider information is actually a valuable asset to improve forecasting accuracy (Leal et al., 2008).

As for the second issue, namely the accuracy of forecasting methods, the literature can be divided into two parts, one based on macroeconomic models with specific fiscal modules that allows to analyse the effects of fiscal policy on macro variables and vice versa (see Favero and Marcellino (2005) and references therein), and the other based on pure forecasting methods and comparisons among them. This last stream of research basically resembles closely what is seen in other forecasting areas: (*i*) there is no single method outperforming the rest generally, (*ii*) judgmental forecasting is especially important due to data problems (see Section 2.11), and (*iii*) combination of methods tends to outperform individual ones, see Leal et al. (2008) and Section 2.6.

Part of the recent literature focused on the generation of very short-term public finance monitoring systems using models that combine annual information with intra-annual fiscal data (Pedregal & Pérez, 2010) by time aggregation techniques (see Section 2.10.2), often set up in a SS framework (see Section 2.3.6). The idea is to produce global annual end-of-year forecasts of budgetary variables based on the most frequently available fiscal indicators, so that changes throughout the year in the indicators can be used as early warnings to infer the changes in the annual forecasts and deviations from fiscal targets (Pedregal, Pérez, & Sánchez, 2014).

The level of disaggregation of the indicator variables are established according to the information available and the particular objectives. The simplest options are the accrual National Accounts annual or quarterly fiscal balances running on their cash monthly counterparts. A somewhat more complex version is the previous one with all the variables broken down into revenues and expenditures. Other disaggregation schemes have been applied, namely by region, by administrative level (regional, municipal, social security, etc.), or by items within revenue and/or expenditure (VAT, income taxes, etc. Asimakopoulos, Paredes, & Warmedinger, 2020; Paredes, Pedregal, & Pérez, 2014).

Unfortunately, what is missing is a comprehensive and transparent forecasting system, independent of Member States, capable of producing consistent forecasts over time and across countries. This is certainly a challenge that no one has yet dared to take up.

3.3.6. Interest rate prediction[110]

The (spot) rate on a (riskless) bond represents the ex-ante return (yield) to maturity which equates its market price to a theoretical valuation. Modelling and predicting default-free, short-term interest rates are crucial tasks in asset pricing and risk management. Indeed, the value of interest rate–sensitive securities depends on the value of the riskless rate. Besides, the short interest rate is a fundamental ingredient in the formulation and transmission of the monetary policy (see, for example, Section 2.3.15). However, many popular models of the short rate (for instance, continuous time, diffusion models) fail to deliver accurate out-of-sample forecasts. Their poor predictive performance may depend on the fact that the stochastic behaviour of short interest rates may be time-varying (for instance, it may depend on the business cycle and on the stance of monetary policy).

Notably, the presence of nonlinearities in the conditional mean and variance of the short-term yield influences the behaviour of the entire term structure of spot rates implicit in riskless bond prices. For instance, the level of the short-term rate directly affects the slope of the yield curve. More generally, nonlinear rate dynamics imply a nonlinear equilibrium relationship between short and long-term yields. Accordingly, recent research has reported that dynamic econometric models with regime shifts in parameters, such as Markov switching (MS; see Section 2.3.12) and threshold models (see Section 2.3.13), are useful at forecasting rates.

The usefulness of MS VAR models with term structure data had been established since Hamilton (1988) and Garcia and Perron (1996): single-state, VARMA models are overwhelmingly rejected in favour of multi-state models. Subsequently, a literature has emerged that has documented that MS models are required to successfully forecast the yield curve. Lanne and Saikkonen (2003) showed that a mixture of autoregressions with two regimes improves the predictions of US T-bill rates. Ang and Bekaert (2002) found support for MS dynamics in the short-term rates for the US, the UK, and Germany. Cai (1994) developed a MS ARCH model to examine volatility persistence, reflecting a concern that it may be inflated by regimes. Gray (1996) generalised this attempt to MS GARCH and reported improvements in pseudo out-of-sample predictions. Further advances in the methods and applications of MS GARCH are in Haas, Mittnik, and Paolella (2004) and Smith (2002). A number of papers have also investigated the presence of regimes in the typical factors (level, slope, and convexity) that characterise the no-arbitrage dynamics of the term structure, showing the predictive benefits of incorporating MS (see, for example, Guidolin & Pedio, 2019; Hevia, Gonzalez-Rozada, Sola, & Spagnolo, 2015).

Alternatively, a few studies have tried to capture the time-varying, nonlinear dynamics of interest rates using threshold models. As discussed by Pai and Pedersen (1999), threshold models have an advantage compared to MS ones: the regimes are not determined by an unobserved latent variable, thus fostering interpretability. In most of the applications to interest rates, the regimes are determined by the lagged level of the short rate itself, in a self-exciting fashion. For instance, Pfann, Schotman, and Tschernig (1996) explored nonlinear dynamics of the US short-term interest rate using a (self-exciting) threshold autoregressive model augmented by conditional heteroskedasticity (namely, a TAR-GARCH model) and found

110 This subsection was written by Massimo Guidolin & Manuela Pedio.

strong evidence of the presence of two regimes. More recently, also Gospodinov (2005) used a TAR-GARCH to predict the short-term rate and showed that this model can capture some well-documented features of the data, such as high persistence and conditional heteroskedasticity.

Another advantage of nonlinear models is that they can reproduce the empirical puzzles that plague the expectations hypothesis of interest rates (EH), according to which it is a weighted average of short-term rates to drive longer-term rates (see, for example, Bansal, Tauchen, & Zhou, 2004; Dai, Singleton, & Yang, 2007). For instance, while Bekaert, Hodrick, and Marshall (2001) show single-state VARs cannot generate distributions consistent with the EH, Guidolin and Timmermann (2009) find that the optimal combinations of lagged short and forward rates depend on regimes so that the EH holds only in some states.

As widely documented (see, for instance, Guidolin & Thornton, 2018), the predictable component in mean rates is hardly significant. As a result, the random walk remains a hard benchmark to outperform as far as the prediction of the mean is concerned. However, density forecasts reflect all moments and the models that capture the dynamics of higher-order moments tend to perform best. MS models appear at the forefront of a class of non-linear models that produce accurate density predictions (see, for example, Hong, Li, & Zhao, 2004; Maheu & Yang, 2016). Alternatively, Pfann et al. (1996) and more recently Dellaportas, Denison, and Holmes (2007) estimated TAR models to also forecast conditional higher order moments and all report reasonable accuracy.

Finally, a literature has strived to fit rates not only under the physical measure, i.e., in time series, but to predict rates when MS enters the pricing kernel, the fundamental pricing operator. A few papers have assumed that regimes represent a new risk factor (see, for instance, Dai & Singleton, 2003). This literature reports that MS models lead to a range of shapes for nominal and real term structures (see, for instance, Veronesi & Yared, 1999). Often the model specifications that are not rejected by formal tests include regimes (Ang, Bekaert, & Wei, 2008; Bansal & Zhou, 2002).

To conclude, it is worthwhile noting that, while threshold models are more interpretable, MS remain a more popular alternative for the prediction of interest rates. This is mainly due to the fact that statistical inference for threshold regime switching models poses some challenges, because the likelihood function is discontinuous with respect to the threshold parameters.

3.3.7. *House price forecasting*[111]

The boom and bust in housing markets in the early and mid 2000s and its decisive role in the Great Recession has generated a vast interest in the dynamics of house prices and emphasised the importance of accurately forecasting property price movements during turbulent times. International organisations, central banks and research institutes have become increasingly engaged in monitoring the property price developments across the world.[112] At the same time, a substantial empirical literature has developed that deals with predicting future house price movements (for a comprehensive survey see Ghysels, Plazzi, Valkanov, & Torous, 2013). Although this literature concentrates almost entirely on the US (see, for example, Bork & Møller, 2015; Rapach & Strauss, 2009), there are many other countries, such as the UK, where house price forecastability is of prime importance. Similarly to the US, in the UK, housing activities account for a large fraction of GDP and of households' expenditures; real estate property comprises a significant component of private wealth and mortgage debt constitutes a main liability of households (Office for National Statistics, 2019).

The appropriate forecasting model has to reflect the dynamics of the specific real estate market and take into account its particular characteristics. In the UK, for instance, there is a substantial empirical literature that documents the existence of strong spatial linkages between regional markets, whereby the house price shocks emanating from southern regions of the country, and in particular Greater London, have a tendency to spread out and affect neighbouring regions with a time lag (see, for example, Antonakakis, Chatziantoniou, Floros, & Gabauer, 2018; Cook & Thomas, 2003; Holly, Pesaran, & Yamagata, 2010, *inter alia*); see also Section 2.3.10 on forecasting functional data.

Recent evidence also suggests that the relationship between real estate valuations and conditioning macro and financial variables displayed a complex of time-varying patterns over the previous decades (Aizenman & Jinjarak, 2013). Hence, predictive methods that do not allow for time-variation in both predictors and their marginal effects may not be able to capture the complex house price dynamics in the UK (see Yusupova, Pavlidis, & Pavlidis, 2019, for a comparison of forecasting accuracy of a battery of static and dynamic econometric methods).

An important recent trend is to attempt to incorporate information from novel data sources (such as newspaper articles, social media, etc.) in forecasting models as a measure of expectations and perceptions of economic agents (see also Section 2.9.3). It has been shown that changes in uncertainty about house prices impact on housing investment and real estate construction decisions (Banks, Blundell, Oldfield, & Smith, 2015; Cunningham, 2006; Oh & Yoon, 2020), and thus incorporating a measure of uncertainty in the forecasting model can improve the forecastability of real estate prices. For instance in the UK, the House Price Uncertainty (HPU) index (Yusupova, Pavlidis, Paya, & Peel, 2020), constructed using the methodology outlined in Baker, Bloom, and Davis (2016),[113] was found to be important in predicting

[111] This subsection was written by Alisa Yusupova.

[112] For instance, the International Monetary Fund recently established the Global Housing Watch, the Globalisation and Monetary Policy Institute of the Federal Reserve Bank of Dallas initiated a project on monitoring international property price dynamics, and the UK Housing Observatory initiated a similar project for the UK national and regional housing markets.

[113] For a comparison of alternative text-based measures of economic uncertainty see Kalamara, Turrell, Redl, Kapetanios, and Kapadia (2020).

property price inflation ahead of the house price collapse of the third quarter of 2008 and during the bust phase (Yusupova et al., 2019). Along with capturing the two recent recessions (in the early 1990s and middle 2000s) this index also reflects the uncertainly related to the EU Referendum, Brexit negotiations and COVID-19 pandemic.

3.3.8. Exchange rate forecasting[114]

Exchange rates have long fascinated and puzzled researchers in international finance. The reason is that following the seminal paper of Meese and Rogoff (1983), the common wisdom is that macroeconomic models cannot outperform the random walk in exchange rate forecasting (see Rossi, 2013, for a survey). This view is difficult to reconcile with the strong belief that exchange rates are driven by fundamentals, such as relative productivity, external imbalances, terms of trade, fiscal policy or interest rate disparity (Couharde, Delatte, Grekou, Mignon, & Morvillier, 2018; Lee, Milesi-Ferretti, & Ricci, 2013; MacDonald, 1998). These two contradicting assertions by the academic literature is referred to as "exchange rate disconnect puzzle".

The literature provides several explanations for this puzzle. First, it can be related to the forecast estimation error (see Section 2.5.2). The studies in which models are estimated with a large panels of data (Engel, Mark, & West, 2008; Ince, 2014; Mark & Sul, 2001), long time series (Lothian & Taylor, 1996) or calibrated (Ca' Zorzi & Rubaszek, 2020) deliver positive results on exchange rate forecastability. Second, there is ample evidence that the adjustment of exchange rates to equilibrium is nonlinear (Curran & Velic, 2019; Taylor & Peel, 2000), which might diminish the out-of-sample performance of macroeconomic models (Kilian & Taylor, 2003; Lopez-Suarez & Rodriguez-Lopez, 2011). Third, few economists argue that the role of macroeconomic fundamentals may be varying over time and this should be accounted for in a forecasting setting (Beckmann & Schussler, 2016; Byrne, Korobilis, & Ribeiro, 2016).

The dominant part of the exchange rate forecasting literature investigates which macroeconomic model performs best out-of-sample. The initial studies explored the role of monetary fundamentals to find that these models deliver inaccurate short-term and not so bad long-term predictions in comparison to the random walk (Mark, 1995; Meese & Rogoff, 1983). In a comprehensive study from mid-2000s, Cheung, Chinn, and Pascual (2005) showed that neither monetary, uncovered interest parity (UIP) nor behavioural equilibrium exchange rate (BEER) model are able to outperform the no-change forecast. A step forward was made by Molodtsova and Papell (2009), who proposed a model combining the UIP and Taylor rule equations and showed that it delivers competitive exchange rate forecasts. This result, however, has not been confirmed by more recent studies (Cheung, Chinn, Pascual, & Zhang, 2019; Engel, Lee, Liu, Liu, & Wu, 2019). In turn, Ca' Zorzi and Rubaszek (2020) argue that a simple method assuming gradual adjustment of the exchange rate towards the level implied by the Purchasing Power Parity (PPP) performs well over shorter as well as longer horizon. This result is consistent with the results of Ca' Zorzi et al. (2017) and Eichenbaum, Johannsen, and Rebelo (2017), who showed that exchange rates are predictable within a general equilibrium DSGE framework (see Section 2.3.15), which encompasses an adjustment of the exchange rate to a PPP equilibrium. Finally, Ca' Zorzi, Cap, Mijakovic, and Rubaszek (2020) discuss how extending the PPP framework for other fundamentals within the BEER framework is not helping in exchange rate forecasting. Overall, at the current juncture it might be claimed that "exchange rate disconnect puzzle" is still puzzling, with some evidence that methods based on PPP and controlling the estimation forecast error can deliver more accurate forecast than the random walk benchmark. A way forward to account for macroeconomic variables in exchange rate forecasting could be to use variable selection methods that allow to control for the estimation error (see Section 2.5.3).

3.3.9. Financial time series forecasting with range-based volatility models[115]

The range-based (RB) volatility models is a general term for the models constructed with high and low prices, and most often with their difference i.e., the price range. A short review and classification of such models is contained in Section 2.3.14. From practical point of view, it is important that low and high prices are almost always available with daily closing prices for financial series. The price range (or its logarithm) is a significantly more efficient estimator of volatility than the estimator based on closing prices (Alizadeh et al., 2002). Similarly the co-range (the covariance based on price ranges) is a significantly more efficient estimator of the covariance of returns than the estimator based on closing prices (Brunetti & Lildholdt, 2002). For these reasons models based on the price range and the co-range better describe variances and covariances of financial returns than the ones based on closing prices.

Forecasts of volatility from simple models like moving average, EWMA, AR, ARMA based on the RB variance estimators are more accurate than the forecasts from the same models based on squared returns of closing prices (Rajvanshi, 2015; Vipul & Jacob, 2007). Forecasts of volatility from the AR model based on the Parkinson estimator are more precise even than the forecasts from the standard GARCH models (see Section 2.3.11) based on closing prices (Li & Hong, 2011).

In plenty of studies it was shown that forecasts of volatility of financial returns from the univariate RB models are more accurate than the forecasts from standard GARCH models based on closing prices (see, for example, Mapa, 2003 for the GARCH-PARK-R model; Chou, 2005 for the CARR model; Fiszeder, 2005 for the GARCH-TR model; Brandt and Jones, 2006 for the RE-GARCH model; Chen et al., 2008 for the TARR model; Lin et al., 2012 for the STARR model; Fiszeder and Perczak, 2016

[114] This subsection was written by Michał Rubaszek.

[115] This subsection was written by Piotr Fiszeder.

for the GARCH model estimated with low, high and closing prices during crisis periods; Molnár, 2016 for the RGARCH model).

The use of daily low and high prices in the multivariate volatility models leads to more accurate forecasts of covariance or covariance matrix of financial returns than the forecasts from the models based on closing prices (see, for example, Chou et al., 2009 for the RB DCC model; Harris and Yilmaz, 2010 for the hybrid EWMA model; Fiszeder, 2018 for the BEKK-HL model; Fiszeder and Fałdziński, 2019 for the co-range DCC model; Fiszeder et al., 2019 for the DCC-RGARCH model).

The RB models were used in many financial applications. They lead for example to more precise forecasts of value-at-risk measures in comparison to the application of only closing prices (see, for example, Chen, Gerlach, Hwang, and McAleer, 2012 for the threshold CAViaR model; Asai and Brugal, 2013 for the HVAR model; Fiszeder et al., 2019 for the DCC-RGARCH model; Meng and Taylor, 2020 for scoring functions). The application of the multivariate RB models provides also the increase in the efficiency of hedging strategies (see, for example, Chou et al., 2009 for the RB DCC model; Harris and Yilmaz, 2010 for the hybrid EWMA model; Su and Wu, 2014 for the RB-MS-DCC model). Moreover, the RB volatility models have more significant economic value than the return-based ones in the portfolio construction (Chou and Liu, 2010 for the RB DCC model; Wu and Liang, 2011 for the RB-copula model). Some studies show that based on the forecasts from the volatility models with low and high prices it is possible to construct profitable investment strategies (He, Kwok, and Wan, 2010 for the VECM model; Kumar, 2015 for the CARRS model).

3.3.10. *Copula forecasting with multivariate dependent financial times series*[116]

In this section, we focus on the practical advances on jointly forecasting multivariate financial time series with copulas. In the copula framework (see Section 2.4.3), because marginal models and copula models are separable, point forecasts are straightforward with marginal models, but dependence information is ignored. A joint probabilistic forecast with copulas involves both estimations of the copula distribution and marginal models.

In financial time series, an emerging interest is to model and forecast the asymmetric dependence. A typical asymmetric dependence phenomenon is that two stock returns exhibit greater correlation during market downturns than market upturns. Patton (2006) employs the asymmetric dependence between exchange rates with a time-varying copula construction with AR and GARCH margins. A similar study for measuring financial contagion with copulas allows the parameters of the copula to change with the states of the variance to identify shifts in the dependence structure in times of crisis (Rodriguez, 2007).

In stock forecasting, Almeida and Czado (2012) employ a stochastic copula autoregressive model to model DJI and Nasdaq, and the dependence at the time is modelled by a real-valued latent variable, which corresponds to the Fisher transformation of Kendall's τ. Li and Kang (2018) use a covariate-dependent copula framework to forecast the time varying dependence that improves both the probabilistic forecasting performance and the forecasting interpretability. Liquidity risk is another focus in finance. Weiß and Supper (2013) forecast three types of liquidity-adjusted intraday Value-at-Risk (L-IVaR) with a vine copula structure. The liquidity-adjusted intraday VaR is based on simulated portfolio values, and the results are compared with the realised portfolio profits and losses.

In macroeconomic forecasting, most existing reduced-form models for multivariate time series produce symmetric forecast densities. Gaussian copulas with skew Student's-t margins depict asymmetries in the predictive distributions of GDP growth and inflation (Smith & Vahey, 2016). Real-time macroeconomic variables are forecasted with heteroscedastic inversion copulas (Smith & Maneesoonthorn, 2018) that allow for asymmetry in the density forecasts, and both serial and cross-sectional dependence could be captured by the copula function (Loaiza-Maya & Smith, 2020).

Copulas are also widely used to detect and forecast default correlation, which is a random variable called *time-until-default* to denote the survival time of each defaultable entity or financial instrument (Li, 2000). Then copulas are used in modelling the dependent defaults (Li, 2000), forecasting credit risk (Bielecki & Rutkowski, 2013), and credit derivatives market forecasting (Schönbucher, 2003). A much large volume of literature is available for this specific area. See the aforementioned references therein. For particular applications in default risk and credit default swap (CDS) forecasting, see Li and He (2019) and Oh and Patton (2018) respectively.

In energy economics, Aloui, Hammoudeh, and Nguyen (2013) employ the time-varying copula approach, where the marginal models are from ARMA(p, q)–GARCH(1,1) to investigate the conditional dependence between the Brent crude oil price and stock markets in the Central and Eastern European transition economies. Bessa, Miranda, Botterud, Zhou, and Wang (2012) propose a time-adaptive quantile-copula where the copula density is estimated with a kernel density forecast method. The method is applied to wind power probabilistic forecasting (see also Section 3.4.6) and shows its advantages for both system operators and wind power producers. Vine copula models are also used to forecast wind power farms' uncertainty in power system operation scheduling. Wang, Wang, Liu, Wang, and Hou (2017) shows vine copulas have advantages of providing reliable and sharp forecast intervals, especially in the case with limited observations available.

3.3.11. *Financial forecasting with neural networks*[117]

Neural Networks (NNs; see Section 2.7.8) are capable of successfully modelling non-stationary and non-linear series. This property has made them one of the most popular (if not the most) non-linear specification used by practitioners and academics in Finance. For example, 89% of European banks use NNs to their operations (European

[116] This subsection was written by Feng Li.

[117] This subsection was written by Georgios Sermpinis.

Banking Federation, 2019) while 25.4% of the NNs applications in total is in Finance (Wong, Bodnovich, & Selvi, 1995).

The first applications of NNs in Finance and currently the most widespread, is in financial trading. In the mid-80s when computational power became cheaper and more accessible, hedge fund managers started to experiment with NNs in trading. Their initial success led to even more practitioners to apply NNs and nowadays 67% of hedge fund managers use NNs to generate trading ideas (BarclayHedge, 2018). A broad measure of the success of NNs in financial trading is provided by the Eurekahedge AI Hedge Fund Index[118] where it is noteworthy the 13.02% annualised return of the selected AI hedge funds over the last 10 years.

In academia, financial trading with NNs is the focus of numerous papers. Notable applications of NNs in trading financial series were provided by Dunis, Laws, and Sermpinis (2010), Kaastra and Boyd (1996), Panda and Narasimhan (2007), Tenti (1996), and Zhang and Ming (2008). The aim of these studies is to forecast the sign or the return of financial trading series and based on these forecasts to generate profitable trading strategies. These studies are closely related to the ones presented in Section 3.3.13 but the focus is now in profitability. The second major field of applications of NNs in Finance is in derivatives pricing and financial risk management. The growth of the financial industry and the provided financial services have made NNs and other machine learning algorithms a necessity for tasks such as fraud detection, information extraction and credit risk assessment (Buchanan, 2019). In derivatives pricing, NNs try to fill the limitations of the Black–Scholes model and are being used in options pricing and hedging. In academia notable applications of NNs in risk management are provided by Liu (2005) and Locarek-Junge and Prinzler (1998) and in derivatives by Bennell and Sutcliffe (2004) and Psaradellis and Sermpinis (2016).

As discussed before, financial series due to their non-linear nature and their wide applications in practice seems the perfect forecasting data set for researchers that want to test their NN topologies. As a result, there are thousands of forecasting papers in the field of NNs in financial forecasting. However, caution is needed in interpretation of their results. NNs are sensitive to the choice of their hyperparameters. For a simple MLP, a practitioner needs to set (among others) the number and type of inputs, the number of hidden nodes, the momentum, the learning rate, the number of epochs and the batch size. This complexity in NN modelling leads inadvertently to the data snooping bias (see also Section 2.12.6). In other words, a researcher that experiments long enough with the parameters of a NN topology can have excellent in-sample and out-of-sample results for a series. However, this does not mean that the results of his NN can be generalised. This issue has led the related literature to be stained by studies cannot be extended in different samples.

[118] https://www.eurekahedge.com/Indices/IndexView/Eurekahedge/683/Eurekahedge-AI-Hedge-fund-Index (Accessed: 2020-09-01)

3.3.12. Forecasting returns to investment style[119]

Investment style or factor portfolios are constructed from constituent securities on the basis of a variety of a-priori observable characteristics, thought to affect future returns. For example a 'Momentum' portfolio might be constructed with positive ('long') exposures to stocks with positive trailing 12-month returns, and negative ('short') exposure to stocks with negative trailing 12-month returns (for full background and context, see, for example Bernstein, 1995; Haugen, 2010).[120] Explanations as to why such characteristics seem to predict returns fall in to two main camps: firstly that the returns represent a risk premium, earned by the investor in return for taking on some kind of (undiversifiable) risk, and secondly that such returns are the result of behavioural biases on the part of investors. In practice, both explanations are likely to drive style returns to a greater or lesser extent. Several such strategies have generated reasonably consistent positive risk-adjusted returns over many decades, but as with many financial return series, return volatility is large relative to the mean, and there can be periods of months or even years when returns deviate significantly from their long-run averages. The idea of timing exposure to styles is therefore at least superficially attractive, although the feasibility of doing so is a matter of some debate (Arnott, Beck, Kalesnik, & West, 2016; Asness, 2016; Bender, Sun, Thomas, & Zdorovtsov, 2018). Overconfidence in timing ability has a direct cost in terms of trading frictions and opportunity cost in terms of potential expected returns and diversification forgone.

A number of authors write on the general topic of style timing (recent examples include Dichtl, Drobetz, Lohre, Rother, & Vosskamp, 2019; Hodges, Hogan, Peterson, & Ang, 2017), and several forecasting methodologies have been suggested, falling in to three main camps:

1. Serial Correlation: Perhaps the most promising approach is exploiting serial correlation in style returns. Babu, Levine, Ooi, Pedersen, and Stamelos (2020) and Tarun and Bryan (2019) outline two such approaches and Ehsani and Linnainmaa (2020) explore the relationship between momentum in factor portfolios and momentum in underlying stock returns. As with valuation spreads mentioned below, there is a risk that using momentum signals to time exposure to momentum factor portfolios risks unwittingly compounding exposure. A related strand of research relates (own) factor volatility to future returns, in particular for momentum factors (Barroso, 2015; Daniel & Moskowitz, 2016).
2. Valuation Spreads: Using value signals (aggregated from individual stock value exposures) to time exposure to various fundamental-based strategies is a popular and intuitively appealing approach (Asness, 2016); however evidence of value added from doing so is mixed, and the technique seems to compound risk exposure to value factors.

[119] This subsection was written by Ross Hollyman.

[120] The website of Kenneth French is an excellent source of data on investment style factor data and research. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

3. Economic & Financial Conditions: Polk, Haghbin, and de Longis (2020) explore how economic and financial conditions affect style returns (an idea that dates back at least to Bernstein (1995) and references therein).

Style returns exhibit distinctly non-normal distributions. On a univariate basis, most styles display returns which are highly negatively skewed and demonstrate significant kurtosis. The long-run low correlation between investment styles is often put forward as a benefit of style-based strategies, but more careful analysis reveals that non-normality extends to the co-movements of investment style returns; factors exhibit significant tail dependence. Christoffersen and Langlois (2013) explores this issue, also giving details of the skew and kurtosis of weekly style returns. These features of the data mean that focusing solely on forecasting the mean may not be sufficient, and building distributional forecasts becomes important for proper risk management. Jondeau (2007) writes extensively on modelling non-gaussian distributions.

3.3.13. *Forecasting stock returns*[121]

Theory and intuition suggest a plethora of potentially relevant predictors of stock returns. Financial statement data (e.g., Chan & Genovese, 2001; Yan & Zheng, 2017) provide a wealth of information, and variables relating to liquidity, price trends, and sentiment, among numerous other concepts, have been used extensively by academics and practitioners alike to predict stock returns. The era of big data further increases the data available for forecasting returns. When forecasting with large numbers of predictors, conventional ordinary least squares (OLS) estimation is highly susceptible to overfitting, which is exacerbated by the substantial noise in stock return data (reflecting the intrinsically large unpredictable component in returns); see Section 2.7.11.

Over the last decade or so, researchers have explored methods for forecasting returns with large numbers of predictors. Principal component regression extracts the first few principal components (or factors) from the set of predictors; the factors then serve as predictors in a low-dimensional predictive regression, which is estimated via OLS (see Section 2.7.1). Intuitively, the factors combine the information in the individual predictors to reduce the dimension of the regression, which helps to guard against overfitting. Ludvigson and Ng (2007) find that a few factors extracted from hundreds of macroeconomic and financial variables improve out-of-sample forecasts of the US market return. Kelly and Pruitt (2013) and Huang, Jiang, Tu and Zhou (2015) use partial least squares (Wold, 1966) to construct target-relevant factors from a cross section of valuation ratios and a variety of sentiment measures, respectively, to improve market return forecasts.

Since Bates and Granger (1969), it has been known that combinations of individual forecasts often perform better than the individual forecasts themselves (Timmermann, 2006, and Section 2.6.1). Rapach, Strauss, and Zhou (2010) show that forecast combination can significantly improve out-of-sample market return forecasts. They first construct return forecasts via individual univariate predictive regressions based on numerous popular predictors from the literature (Goyal & Welch, 2008). They then compute a simple combination forecast by taking the average of the individual forecasts. Rapach et al. (2010) demonstrate that forecast combination exerts a strong shrinkage effect, thereby helping to guard against overfitting.

An emerging literature uses machine-learning techniques to construct forecasts of stock returns based on large sets of predictors. In an investigation of lead–lag relationships among developed equity markets, Rapach, Strauss, and Zhou (2013) appear to be the first to employ machine-learning tools to predict market returns. They use the elastic net (ENet: Zou & Hastie, 2005), a generalisation of the popular least absolute shrinkage and selection operator (LASSO: Tibshirani, 1996). The LASSO and ENet employ penalised regression to guard against overfitting in high-dimensional settings by shrinking the parameter estimates toward zero. Chinco, Clark-Joseph, and Ye (2019) use the LASSO to forecast high-frequency (one-minute-ahead) individual stock returns and report improvements in out-of-sample fit, while Rapach et al. (2019) use the LASSO to improve monthly forecasts of industry returns.

Incorporating insights from Diebold and Shin (2019), Han, He, Rapach, and Zhou (2021) use the LASSO to form combination forecasts of cross-sectional stock returns based on a large number of firm characteristics from the cross-sectional literature (e.g., Harvey, Liu, & Zhu, 2016; Hou, Xue, & Zhang, 2020; McLean & Pontiff, 2016), extending the conventional OLS approach of Green, Hand, and Zhang (2017), Haugen and Baker (1996), and Lewellen (2015). Dong, Li, Rapach, and Zhou (2021) and Rapach and Zhou (2020) use the ENet to compute combination forecasts of the market return based on popular predictors from the time-series literature and numerous anomalies from the cross-sectional literature, respectively. Forecasting individual stock returns on the basis of firm characteristics in a panel framework, Freyberger, Neuhierl, and Weber (2020) and Gu, Kelly, and Xiu (2020) employ machine-learning techniques – such as the nonparametric additive LASSO (Huang, Horowitz, & Wei, 2010), random forests (Breiman, 2001), and artificial neural networks – that allow for nonlinear predictive relationships.

3.3.14. *Forecasting crashes in stock markets*[122]

Time series data on financial asset returns have special features. Returns themselves are hard to forecast, while it seems that volatility of returns can be predicted. Empirical distributions of asset returns show occasional clusters of large positive and large negative returns. Large negative returns, that is, crashes seem to occur more frequently than large positive returns. Forecasting upcoming increases or decreases in volatility can be achieved by using variants of the Autoregressive Conditional Heteroskedasticity (ARCH) model (Bollerslev, 1986; Engle, 1982, and Section 2.3.11) or realized volatility models

121 This subsection was written by David E. Rapach.

122 This subsection was written by Philip Hans Franses.

(Taylor, 1986a). These models take (functions of) past volatility and past returns as volatility predictors, although also other explanatory variables can be incorporated in the regression.

An important challenge that remains is to predict crashes. Sornette (2003) summarises potential causes for crashes and these are computer trading, increased trading in derivatives, illiquidity, trade and budget deficits, and especially, herding behaviour of investors. Yet, forecasting the exact timing of crashes may seem impossible, but on the other hand, it may be possible to forecast the probability that a crash may occur within a foreseeable future. Given the herding behaviour, any model to use for prediction should include some self-exciting behaviour. For that purpose, Aït-Sahalia, Cacho-Diaz, and Laeven (2015) propose mutually exciting jump processes, where jumps can excite new jumps, also across assets or markets (see also Chavez-Demoulin, Davison, & McNeil, 2005). Another successful approach is the Autoregressive Conditional Duration (ACD) model (Engle & Russell, 1997, 1998), which refers to a time series model for durations between (negative) events.

An alternative view on returns' volatility and the potential occurrence of crashes draws upon the earthquake literature (Ogata, 1978, 1988). The idea is that tensions in and across tectonic plates build up, until an eruption, and after that, tension starts to build up again until the next eruption. By modelling the tension-building-up process using so-called Hawkes processes (Hawkes, 1971, 2018; Hawkes & Oakes, 1974; Ozaki, 1979), one can exploit the similarities between earthquakes and financial crashes (see also Section 2.8.4). Gresnigt, Kole, and Franses (2015) take Hawkes processes to daily S&P 500 data and show that it is possible to create reliable probability predictions of a crash occurrence within the next five days. Gresnigt, Kole, and Franses (2017a, 2017b) further develop a specification strategy for any type of asset returns, and document that there are spillovers across assets and markets.

Given investor behaviour, past crashes can ignite future crashes. Hawkes processes are particularly useful to describe this feature and can usefully be implemented to predict the probability of nearby crashes. By the way, these processes can also be useful to predict social conflicts, as also there one may discern earthquake-like patterns. van den Hengel and Franses (2020) document their forecasting power for social conflicts in Africa.

3.4. Energy

3.4.1. Building energy consumption forecasting and optimisation[123]

In Europe, buildings account for 40% of total energy consumed and 36% of total CO_2 emissions (Patti, Acquaviva, Jahn, Pramudianto, Tomasi, Rabourdin, Virgone, & Macii, 2016). Given that energy consumption of buildings is expected to increase in the coming years, forecasting electricity consumption becomes critical for improving energy management and planning by supporting a large variety of optimisation procedures.

The main challenge in electricity consumption forecasting is that building energy systems are complex in nature, with their behaviour depending on various factors related to the type (e.g., residential, office, entertainment, business, and industrial) and the end-uses (e.g., heating, cooling, hot water, and lighting) of the building, its construction, its occupancy, the occupants' behaviour and schedule, the efficiency of the installed equipment, and the weather conditions (Zhao & Magoulès, 2012). Special events, holidays, and calendar effects can also affect the behaviour of the systems and further complicate the consumption patterns, especially when forecasting at hourly or daily level (see Section 2.3.5). As a result, producing accurate forecasts typically requires developing tailored, building-specific methods.

To deal with this task, the literature focuses on three main classes of forecasting methods, namely engineering, statistical, and ML (Mat Daut, Hassan, Abdullah, Rahman, Abdullah, & Hussin, 2017). Engineering methods, typically utilised through software tools such as DOE-2, EnergyPlus, BLAST, and ESP-r, build on physical models that forecast consumption through detailed equations which account for the particularities of the building (Al-Homoud, 2001; Foucquier, Robert, Suard, Stéphan, & Jay, 2013; Zhao & Magoulès, 2012). Statistical methods usually involve linear regression (see Section 2.3.2), ARIMA/ARIMAX (see Section 2.3.4), and exponential smoothing (see Section 2.3.1) models that forecast consumption using past consumption data or additional explanatory variables, such as weather or occupancy and calendar related information (Deb, Zhang, Yang, Lee, & Shah, 2017). Finally, ML methods (see Section 2.7.10) typically involve neural networks (see Section 2.7.8), support vector machines, and grey models that account for multiple non-linear dependencies between the electricity consumed and the factors influencing its value (Ahmad, Hassan, Abdullah, Rahman, Hussin, Abdullah, & Saidur, 2014). Till present, the literature has been inconclusive about which class of methods is the most appropriate, with the conclusions drawn being subject to the examined building type, data set used, forecasting horizon considered, and data frequency at which the forecasts are produced (Wei, Li, Peng, Zeng, & Lu, 2019). To mitigate this problem, combinations of methods (see Section 2.6) and hybrids (see Section 2.7.13) have been proposed, reporting encouraging results (Mohandes, Zhang, & Mahdiyar, 2019; Zhao & Magoulès, 2012).

Other practical issues refer to data pre-processing. Electricity consumption data is typically collected at high frequencies through smart meters and therefore display noise and missing or extreme values due to monitoring issues (see Section 2.7.11). As a result, verifying the quality of the input data through diagnostics and data cleansing techniques (see Section 2.2.3 and Section 2.2.4), as well as optimising the selected time frames, are important for improving forecasting performance (Bourdeau, qiang Zhai, Nefzaoui, Guo, & Chatellier, 2019). Similarly, it is critical to engineer (see Section 2.2.5) and select (see Section 2.5.3) appropriate regressor variables which

[123] This subsection was written by Christoph Bergmeir & Evangelos Spiliotis.

are of high quality and possible to accurately predict to assist electricity consumption forecasting. Finally, it must be carefully decided whether the bottom-up, the top-down or a combination method (see Section 2.10.1) will be used for producing reconciled forecasts at both building and end-use level (Kuster, Rezgui, & Mourshed, 2017), being also possibly mixed with temporal aggregation approaches (Spiliotis, Petropoulos et al., 2020, but also Section 2.10.3).

Provided that accurate forecasts are available, effective energy optimisation can take place at a building level or across blocks of buildings (see Section 3.4.10) to reduce energy cost, improve network stability, and support efforts towards a carbon-free future, by exploiting smart grid, internet of things (IoT), and big data technologies along with recommendation systems (Marinakis et al., 2020).

An example for a typical application in this area is the optimisation of heating, ventilation, and air conditioning (HVAC) systems. The goal is to minimise the energy use of the HVAC system under the constraints of maintaining certain comfort levels in the building (Marinakis, Doukas, Spiliotis, & Papastamatiou, 2017). Though this is predominantly an optimisation exercise, forecasting comes in at different points of the system as input into the optimisation, and many problems in this space involve forecasting as a sub-problem, including energy consumption forecasting, room occupancy forecasting, inside temperature forecasting, (hyper-local) forecasts of outside temperature, and air pressure forecasting for ventilation, among others. For instance, Krüger and Givoni (2004) use a linear regression approach to predict inside temperatures in 3 houses in Brazil, and Ruano, Crispim, Conceiçao, and Lúcio (2006) propose the use of a neural network to predict temperatures in a school building. Madaus, McDermott, Hacker, and Pullen (2020) predict hyper-local extreme heat events, combining global climate models and machine learning models. Jing, Cai, Chen, Zhai, Cui, and Yin (2018) predict air pressure to tackle the air balancing problem in ventilation systems, using a support vector machine.

Predicting energy demand on a building/household level from smart meter data is an important research topic not only for energy savings. In the building space, Ahmad, Mourshed, and Rezgui (2017), Touzani, Granderson, and Fernandes (2018), and Wang, Wang, Zeng, Srinivasan, and Ahrentzen (2018) predict building energy consumption of residential and commercial buildings using decision tree-based algorithms (random forests and gradient boosted trees) and neural networks to improve energy efficiency.

A recent trend in forecasting are global forecasting models, built across sets of time series (Januschowski et al., 2020). (Recurrent) neural networks (Bandara et al., 2020a; Hewamalage et al., 2021) are particularly suitable for this type of processing due to their capabilities to deal with external inputs and cold-start problems. Such capabilities are necessary if there are different regimes in the simulations under which to predict, an example of such a system for HVAC optimisation is presented by Godahewa, Deng, Prouzeau, and Bergmeir (2020).

More generally, many challenges in the space of building energy optimisation are classical examples of so-called "predict then optimise" problems (Demirovic et al., 2019; Elmachtoub & Grigas, 2017). Here, different possible scenario predictions are obtained from different assumptions in the form of input parameters. These input parameters are then optimised to achieve a desired predicted outcome. As both prediction and optimisation are difficult problems, they are usually treated separately (Elmachtoub & Grigas, 2017), though there are now recent works where they are considered together (Demirovic et al., 2019; El Balghiti, Elmachtoub, Grigas, & Tewari, 2019), and this will certainly be an interesting avenue for future research.

3.4.2. Electricity price forecasting[124]

Forecasting electricity prices has various challenges that are highlighted in the detailed review papers by Weron (2014). Even though there are economically well motivated fundamental electricity price models, forecasting models based on evaluating historic price data are the dominating the academic literature. In recent years the focus on probabilistic forecasting grew rapidly, as they are highly relevant for many applications in energy trading and risk management, storage optimisation and predictive maintenance, (Nowotarski & Weron, 2018; Ziel & Steinert, 2018). Electricity price data is highly complex and is influenced by regulation. However, there is electricity trading based on auctions and on continuous trading. Many markets like the US and European markets organise day-ahead auctions for electricity prices, see Fig. 7. Thus, we have to predict multivariate time series type data (Ziel & Weron, 2018). In contrast, intraday markets usually apply continuous trading to manage short term variations due to changes in forecasts of renewable energy and demand, and outages (Kiesel & Paraschiv, 2017).

The key challenge in electricity price forecasting is to address all potential characteristics of the considered market, most notably (some of them visible in Fig. 7):

1. (time-varying) autoregressive effects and (in)stationarity,
2. calendar effects (daily, weekly and annual seasonality, holiday effects, clock-change),
3. (time-varying) volatility and higher moment effects,
4. price spikes (positive and negative), and
5. price clustering.

Some of those impacts can be explained by external inputs, that partially have to be predicted in advance:

1. load/demand/consumption (see Section 3.4.3),
2. power generation, especially from the renewable energy sources (RES) of wind and solar (see Sections 3.4.6 and 3.4.8),
3. relevant fuel prices (especially oil, coal, natural gas; see also Section 3.4.4),
4. prices of emission allowances (CO_2e costs),

[124] This subsection was written by Luigi Grossi & Florian Ziel.

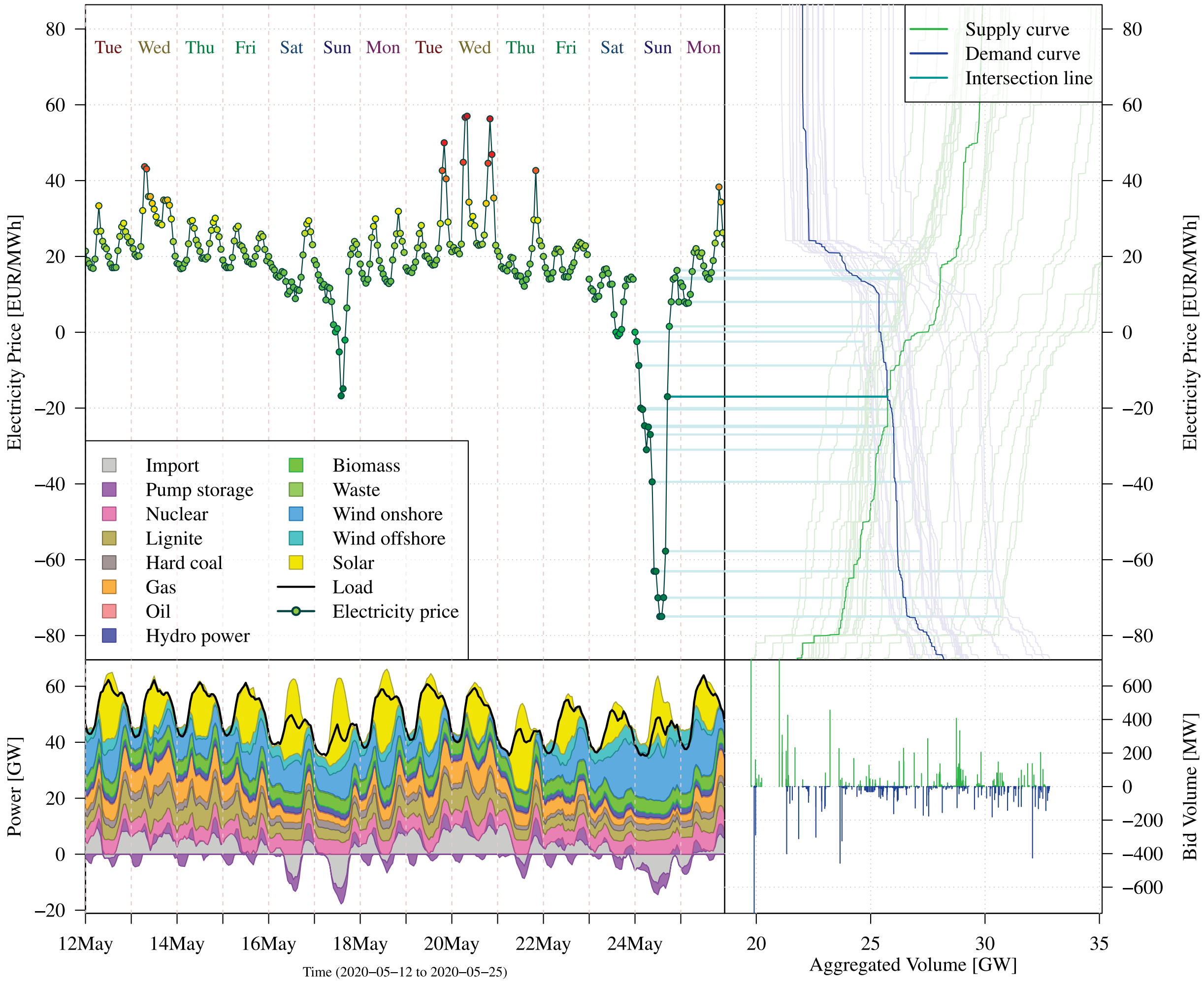

Fig. 7. Hourly German day-ahead electricity price data resulting from a two-sided auction (top left) with corresponding 24 sale/supply and purchase/demand curves for 24 May 2020 and highlighted curves for 17:00 (top right), power generation and consumption time series (bottom left), and bid structure of 24 May 2020 17:00 (bottom right).

5. related power market prices (future, balancing and neighboring markets),
6. availabilities of power plants and interconnectors,
7. import/export flow related data, and
8. weather effects (e.g. temperature due to cooling and heating and combined heat and power (CHP) effects; see also Section 3.5.2).

Note that other weather effects might be relevant as well, but should be covered from the fundamental point of view by the listed external inputs. Obvious examples are wind speed for the wind power prediction, cloud cover for the solar power production and illumination effects in the electricity consumption.

Many of those external effects may be explained by standard economic theory from fundamental electricity price models (Cludius, Hermann, Matthes, & Graichen, 2014; Kulakov & Ziel, 2021). Even the simple supply stack model (merit order model), see Fig. 8, explains many features and should be kept in mind when designing an appropriate electricity price forecasting model.

In recent years, statistical and machine learning methods gained a lot of attraction in day-ahead electricity price forecasting. Even though the majority of effects is linear there are specific non-linear dependencies that can be explored by using non-linear models, especially neural networks (Dudek, 2016; Lago, De Ridder, & De Schutter, 2018; Marcjasz, Uniejewski, & Weron, 2019; Ugurlu, Oksuz, & Tas, 2018). Of course this comes along with higher computational costs compared to linear models. Fezzi and Mosetti (2020) illustrate that even simple linear models can give highly accurate forecasts, if correctly calibrated. However, there seems to be consensus that forecast combination is appropriate, particularly for models that have different structures or different calibration window length (Gaillard, Goude, & Nedellec, 2016; Hubicka, Marcjasz, & Weron, 2018; Mirakyan, Meyer-Renschhausen, & Koch, 2017).

Another increasing stream of electricity price forecasting models do not focus on the electricity price itself, but the bid/sale/sell/supply and ask/sell/purchase/demand curves of the underlying auctions (see Fig. 7, but also

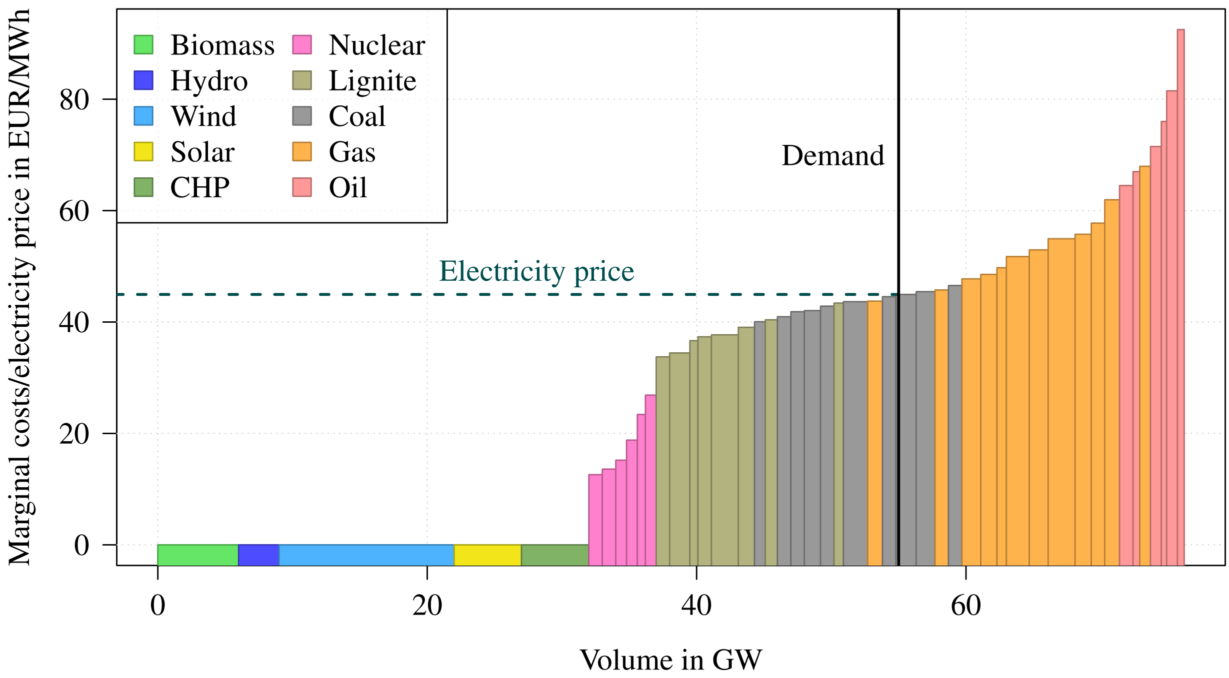

Fig. 8. Illustrative example of a supply stack model with inelastic demand for different power plant types, roughly covering the situation in Germany 2020.

Kulakov, 2020; Mestre, Portela, San Roque, & Alonso, 2020; Shah & Lisi, 2020; Ziel & Steinert, 2016). This sophisticated forecasting problem allows more insights for trading applications and the capturing of price clusters.

In forecasting intraday markets the literature just started to grow quickly. As the aforementioned market characteristics get less distinct if information from day-ahead markets is taken into account appropriately. However, intraday prices are usually more volatile and exhibit stronger price spikes. Thus, probabilistic forecasting is even more relevant (Janke & Steinke, 2019; Narajewski & Ziel, 2020b). Recent studies showed that European markets are close to weak-form efficiency. Thus naive point forecasting benchmarks perform remarkably well (Marcjasz, Uniejewski, & Weron, 2020; Narajewski & Ziel, 2020a; Oksuz & Ugurlu, 2019).

As pointed out above, predicting price spikes is particularly important in practice, due to the high impact in decision making problems which occur usually in extreme situations, see Fig. 8. Very high electricity prices are usually observed in connection to high demand and low renewable energy generation, sometimes together with sudden power plant failures. In contrast, negative price spikes occur in oversupply situation, when there is low demand but high penetration from wind and solar power. The presence of spikes is explored in two main streams in literature: spike forecasting and prediction of prices under normal regime through robust estimators.

Within the first set of papers, spikes are often modelled as one regime of non-linear models for time series. This approach is followed by Mount, Ning, and Cai (2006) focusing on regime-switching models with parameters driven by time-varying variables and by Becker, Hurn, and Pavlov (2008) who adopt Markov switching models for spikes prediction. Christensen, Hurn, and Lindsay (2009, 2012) suggest treating and forecasting price spikes through Poisson autoregressive and discrete-time processes, respectively. Herrera and González (2014) use a Hawkes model combined with extreme events theory. Interregional links among different electricity markets are used by Clements, Herrera, and Hurn (2015) and Manner, Türk, and Eichler (2016) to forecast electricity price spikes. A new procedure for the simulation of electricity spikes has been recently proposed by Muniain and Ziel (2020) utilising bivariate jump components in a mean reverting jump diffusion model in the residuals.

The second stream of literature includes papers developing outlier detection methods or robust estimators to improve the forecasting performance of the models. Martínez-Álvarez, Troncoso, Riquelme, and Aguilar-Ruiz (2011) tackle the issue of outlier detection and prediction defining "motifs", that is patches of units preceding observations marked as anomalous in a training set. Janczura, Trück, Weron, and Wolff (2013) focus on the detection and treatment of outliers in electricity prices. A very similar approach, based on seasonal autoregressive models and outlier filtering, is followed by Afanasyev and Fedorova (2019). Grossi and Nan (2019) introduced a procedure for the robust statistical prediction of electricity prices. The econometric framework is represented by the robust estimation of non-linear SETAR processes. A similar approach has been followed by Wang, Yang, Du, and Niu (2020) using an outlier-robust machine learning algorithm.

3.4.3. Load forecasting[125]

Load forecasting forms the basis where power system operation and planning builds upon. Based on the time horizon of the forecasts, load forecasting can be classified into very short-term (VSTLF), that refers to horizon from several minutes ahead up to 1 h, short-term (STLF), that spans from 1 h to 168 h ahead, medium-term (MTLF), that spans from 168 h to 1 year ahead and finally, and long-term (LTLF) that concerns predictions from 1 year to several years ahead. In VSTLF and STLF applications, the focus is on the sub-hourly or hourly load. In MTLF and LTLF, the variables of interest can be either monthly electricity peak load and total demand for energy.

Inputs differ in the various horizons. In VSTLF and STLF, apart from meteorological data, day type identification codes are used. In LTLF, macroeconomic data are used since total demand of energy is influenced by the long-term modifications of the social and economic environments. Among the horizons, special attention is placed at STLF. This is reflected by the research momentum that have been placed in the load forecasting related literature by other researchers (Hong & Fan, 2016). Processes like unit commitment and optimal power flow rely on STLF (Bo & Li, 2012; Saksornchai, Lee, Methaprayoon, Liao, & Ross, 2005). Additionally, since competitive energy markets continually evolve, STLF becomes vital for new market entities such as retailers, aggregators, and prosumers for applications such as strategic bidding, portfolio optimisation, and tariff design (Ahmad, Javaid, Mateen, Awais, & Khan, 2019; Danti & Magnani, 2017).

The models that can be found in the load forecasting related literature can in general categorised into three types: time-series, machine learning, and hybrid. Time-series models historically precede the others. Typical examples of this family are ARMA, ARIMA, and others (see also Section 2.3.4). In the machine learning models, the structure is usually determined via the training process. NNs are commonly used. Once a NN is sufficiently trained, it can provide forecasts for all types of forecasting horizons (Hippert, Pedreira, & Souza, 2001). The third category of models refers to the integration of two or more individual forecasting approaches (see also see Section 2.7.13). For instance, a NN can be combined with time series methods, with unsupervised machine learning algorithms, data transformation, and with meta-heuristics algorithms (Bozkurt, Biricik, & Tayşi, 2017; El-Hendawi & Wang, 2020; López, Zhong, & Zheng, 2017; Lu, Azimi, & Iseley, 2019).

Hybrid systems has been tested on validation data (through forecasting competitions), power system aggregated load, and application oriented tasks. Ma (2021) proposed an ensemble method based on a combination of various single forecasters on GEFCom2012 forecasting competition data that outperformed benchmark forecasters such as Theta method, NN, ARIMA, and others (see Section 2.12.7 for further discussions on forecasting competitions). For aggregated load cases, researchers focus on different countries and energy markets. Zhang, Wei, Li, Tan, and Zhou (2018) combined empirical mode decomposition (EMD), ARIMA, and wavelet neural networks (WNN) optimised by the fruit fly algorithm on Australian Market data and New York City data. Their approach was to separate the linear and nonlinear components from original electricity load; ARIMA is used for linear part while the WNN for the non-linear one.

Sideratos, Ikonomopoulos, and Hatziargyriou (2020) proposed that a radial basis network that performs the initial forecasting could serve as input to a convolutional neural network that performs the final forecasting. The proposed model led to lower error compared to the persistence model, NN, and SVM. Semero, Zhang, and Zheng (2020) focused on the energy management of a microgrid located in China using EMD to decompose the load, adaptive neuro-fuzzy inference system (ANFIS) for forecasting and particle swarm intelligence (PSO) to optimize ANFIS parameters. The results show that the proposed approach yielded superior performance over four other methods. Faraji, Ketabi, Hashemi-Dezaki, Shafie-Khah, and Catalão (2020) proposed a hybrid system for the scheduling of a prosumer microgrid in Iran. Various machine learning algorithms provided load and weather forecasts. Through an optimisation routine, the best individual forecast is selected. The hybrid system displayed better accuracy from the sole application of the individual forecasters.

3.4.4. Crude oil price forecasting[126]

Crude oil, one of the leading energy resources, has contributed to over one-third of the world's energy consumption (Alvarez-Ramirez, Soriano, Cisneros, & Suarez, 2003). The fluctuations of the crude oil price have a significant impact on industries, governments as well as individuals, with substantial up-and-downs of the crude oil price bringing dramatic uncertainty for the economic and political development (Cunado & De Gracia, 2005; Kaboudan, 2001). Thus, it is critical to develop reliable methods to accurately forecast crude oil price movement, so as to guard against the crude oil market extreme risks and improve macroeconomic policy responses. However, the crude oil price movement suffers from complex features such as nonlinearity, irregularities, dynamics and high volatility (Alquist, Kilian, & Vigfusson, 2013; Herrera, Hu, & Pastor, 2018; Kang, Kang, & Yoon, 2009, and also Section 2.3.11), making the crude oil price forecasting still one of the most challenging forecasting problems.

Some prior studies have suggested that the crude oil price movement is inherently unpredictable, and it would be pointless and futile to attempt to forecast future prices, see Miao, Ramchander, Wang, and Yang (2017) for a detailed summary. These agnostics consider the naive no-change forecast as the best available forecast value of future prices. In recent years, however, numerous studies result in forecasts that are more accurate than naive no-change forecasts, making the forecasting activities of crude oil prices promising (Alquist et al., 2013; Baumeister, Guérin, & Kilian, 2015).

[125] This subsection was written by Ioannis Panapakidis.

[126] This subsection was written by Xiaoqian Wang.

Extensive research on crude oil price forecasting has focused predominantly on the econometric models, such as VAR, ARCH-type, ARIMA, and Markov models (see, for example, Agnolucci, 2009; e Silva, Legey, & e Silva, 2010; Mirmirani & Li, 2004; Mohammadi & Su, 2010, and Section 2.3). In the forecasting literature, unit root tests (see Section 2.3.4) are commonly applied to examine the stationarity of crude oil prices prior to econometric modelling (Rahman & Serletis, 2012; Serletis & Rangel-Ruiz, 2004; Silvapulle & Moosa, 1999). It is well-documented that crude oil prices are driven by a large set of external components, which are themselves hard to predict, including supply and demand forces, stock market activities, oil-related events (e.g., war, weather conditions), political factors, etc. In this context, researchers have frequently considered structural models (see Section 2.3.9), which relate the oil price movements to a set of economic factors. With so many econometric models, is there an optimal one? Recently, de Albuquerquemello, de Medeiros, da Nóbrega Besarria, and Maia (2018) proposed a SETAR model, allowing for predictive regimes changing after a detected threshold, and achieved performance improvements over six widely used econometric models. Despite their high computational efficiency, the econometric models are generally limited in the ability to nonlinear time series modelling.

On the other hand, artificial intelligence and machine learning techniques, such as belief networks, support vector machines (SVMs), recurrent neural networks (RNNs), and extreme gradient boosting (XGBoost), provided powerful solutions to recognise the nonlinear and irregular patterns of the crude oil price movement with high automation (see, for example, Abramson & Finizza, 1991; Gumus & Kiran, 2017; Mingming & Jinliang, 2012; Xie, Yu, Xu, & Wang, 2006). However, challenges also exist in these techniques, such as computational cost and overfitting. In addition, a large number of studies have increasingly focused on the hybrid forecasting models (see also Section 2.7.13) based on econometrics models and machine learning techniques (Baumeister & Kilian, 2015; Chiroma, Abdulkareem, & Herawan, 2015; He, Yu, & Lai, 2012; Jammazi & Aloui, 2012), achieving improved performance. Notably, the vast majority of the literature has focused primarily on the deterministic prediction, with much less attention paid to the probabilistic prediction and uncertainty analysis. However, the high volatility of crude oil prices makes probabilistic prediction more crucial to reduce the risk in decision-making (Abramson & Finizza, 1995; Sun, Sun, Wang, & Wei, 2018).

3.4.5. *Forecasting renewable energy technologies*[127]

The widespread adoption of renewable energy technologies, RETs, plays a driving role in the transition to low-carbon energy systems, a key challenge to face climate change and energy security problems. Forecasting the diffusion of RETs is critical for planning a suitable energy agenda and setting achievable targets in terms of electricity generation, although the available time series are often very short and pose difficulties in modelling. According to Rao and Kishore (2010), renewables' typical characteristics such as low load factor, need for energy storage, small size, high upfront costs create a competitive disadvantage, while Meade and Islam (2015b) suggested that renewable technologies are different from other industrial technological innovations because, in the absence of focused support, they are not convenient from a financial point of view. In this sense, policy measures and incentive mechanisms, such as feed-in tariffs, have been used to stimulate the market. As highlighted in Lee and Huh (2017b), forecasting RETs requires to capture different socio-economic aspects, such as policy choices by governments, carbon emissions, macroeconomic factors, economic and financial development of a country, competitive strength of traditional energy technologies.

The complex and uncertain environment concerning RETs deployment has been faced in literature in several ways, in order to account for various determinants of the transition process. A first stream of research employed a bottom-up approach, where forecasts at a lower level are aggregated to higher levels within the forecasting hierarchy. For instance Park, Yun, Yun, Lee, and Choi (2016) realised a bottom-up analysis to study the optimum renewable energy portfolio, while Lee and Huh (2017a) performed a three-step forecasting analysis, to reflect the specificities of renewable sources, by using different forecasting methods for each of the sources considered. A similar bottom-up perspective was adopted in Zhang, Bauer, Yin, and Xie (2020), by conducting a multi-region study, to understand how multi-level learning may affect RETs dynamics, with the regionalised model of investment and technological development, a general equilibrium model linking a macro-economic growth with a bottom-up engineering-based energy system model.

The relative newness of RETs has posed the challenge of forecasting with a limited amount of data: in this perspective, several contributions applied the 'Grey System' theory, a popular methodology for dealing with systems with partially unknown parameters (Kayacan, Ulutas, & Kaynak, 2010). Grey prediction models for RETs forecasting were proposed in Liu and Wu (2021), Lu (2019), Moonchai and Chutsagulprom (2020), Tsai, Xue, Zhang, Chen, Liu, Zhou, and Dong (2017) and Wu, Ma, Zeng, Wang, and Cai (2019).

Other studies developed forecasting procedures based on growth curves and innovation diffusion models (see Sections 2.3.18–2.3.20): from the seminal work by Marchetti and Nakicenovic (1979), contributions on the diffusion of RETs were proposed by Bunea, Della Posta, Guidolin, and Manfredi (2020), Dalla Valle and Furlan (2011), Guidolin and Mortarino (2010), Lee and Huh (2017b) and Meade and Islam (2015b). Forecasting the diffusion of renewable energy technologies was also considered within a competitive environment in Furlan and Mortarino (2018), Guidolin and Alpcan (2019), Guidolin and Guseo (2016) and Huh and Lee (2014).

3.4.6. *Wind power forecasting*[128]

Wind energy is a leading source of renewable energy, meeting 4.8% of global electricity demand in 2018, more

[127] This subsection was written by Mariangela Guidolin.

[128] This subsection was written by Jethro Browell.

than twice that of solar energy (IEA, 2020). Kinetic energy in the wind is converted into electrical energy by wind turbines according to a characteristic 'power curve'. Power production is proportion to the cube of the wind speed at low-to-moderate speeds, and above this is constant at the turbine's rated power. At very high or low wind speeds no power is generated. Furthermore, the power curve is influenced by additional factors including air density, icing, and degradation of the turbine's blades.

Forecasts of wind energy production are required from minutes to days-ahead to inform the operation of wind farms, participation in energy markets and power systems operations. However, the limited predictability of the weather (see also Section 3.5.2) and the complexity of the power curve make this challenging. For this reason, probabilistic forecasts are increasing used in practice (Bessa et al., 2017). Their value for energy trading is clear (Pinson, Chevallier, & Kariniotakis, 2007), but quantifying value for power system operation is extremely complex. Wind power forecasting may be considered a mature technology as many competing commercial offerings exist, but research and development efforts to produce novel and enhanced products is ongoing (see also Section 3.4.5).

Short-term forecasts (hours to days ahead) of wind power production are generally produced by combining numerical weather predictions (NWP) with a model of the wind turbine, farm or even regional power curve, depending on the objective. The power curve may be modelled using physical information, e.g. provided by the turbine manufacturer, in which case it is also necessary to post-process NWP wind speeds to match the same height-above-ground as the turbine's rotor. More accurate forecasts can be produced by learning the NWP-to-energy relationship from historic data when it is available. State-of-the-art methods for producing wind power forecasts leverage large quantities of NWP data to produce a single forecast (Andrade, Filipe, Reis, & Bessa, 2017) and detailed information about the target wind farm (Gilbert, Browell, & McMillan, 2020a). A number of practical aspects may also need to be considered by users, such as maintenance outages and requirements to reduce output for other reasons, such as noise control or electricity network issues.

Very short-term forecast (minutes to a few hours ahead) are also of value, and on these time scales recent observations are the most significant input to forecasting models and more relevant than NWP. Classical time series methods perform well (see Section 2.3), and those which are able to capture spatial dependency between multiple wind farms are state-of-the-art, notably vector autoregressive models and variants (Cavalcante, Bessa, Reis, & Browell, 2016; Messner & Pinson, 2018). Care must be taken when implementing these models as wind power time series are bounded by zero and the wind farm's rated power meaning that errors may not be assumed to be normally distributed. The use of transformations is recommended (see also Section 2.2.1), though the choice of transformation depends on the nature of individual time series (Pinson, 2012).

Wind power forecasting is reviewed in detail in Giebel and Kariniotakis (2017), Hong, Pinson, Wang, Weron, Yang, and Zareipour (2020) and Zhang, Wang, and Wang (2014), and research is ongoing in a range of directions including: improving accuracy and reducing uncertainty in short-term forecasting, extending forecast horizons to weeks and months ahead, and improving very short-term forecast with remote sensing and data sharing (Sweeney, Bessa, Browell, & Pinson, 2019, and Section 3.4.10).

3.4.7. Wave forecasting[129]

Ocean waves are primarily generated by persistent winds in one direction. The energy thus propagated by the wind is referred to as wave energy flux and follows a linear function of wave height squared and wave period. Wave height is typically measured as significant wave height, the average height of the highest third of the waves. The mean wave period, typically measured in seconds, is the average time between the arrival of consecutive crests, whereas the peak wave period is the wave period at which the highest energy occurs at a specific point.

The benefit of wave energy is that it requires significantly less reserve compared to those from wind (see Section 3.4.6) and solar (see Section 3.4.8) renewable energy sources (Hong et al., 2016). For example, the forecast error at one hour ahead for the simulated wave farms is typically in the range of 5%–7%, while the forecast errors for solar and wind are 17 and 22% respectively (Reikard, Pinson, & Bidlot, 2011). Solar power is dominated by diurnal and annual cycles but also exhibits nonlinear variability due to factors such as cloud cover, temperature and precipitation. Wind power is dominated by large ramp events such as irregular transitions between states of high and low power. Wave energy exhibits annual cycles and is generally smoother although there are still some large transitions, particularly during the winter months. In the first few hours of forecasting wave energy, time series models are known to be more accurate than numerical wave prediction. Beyond these forecast horizons, numerical wave prediction models such as SWAN (Simulating WAves Nearshore, Booij, Ris, & Holthuijsen, 1999) and WAVEWATCH III® (Tolman, 2008) are widely used. As there is as yet no consensus on the most efficient model for harnessing wave energy, potential wave energy is primarily measured with energy flux, but the wave energy harnessed typically follows non-linear functions of wave height and wave period in the observations of the six different types of wave energy converters (Reikard, Robertson, Buckham, Bidlot, & Hiles, 2015).

To model the dependencies of wind speed, wave height, wave period and their lags, Reikard et al. (2011) uses linear regressions, which were then converted to forecasts of energy flux. Pinson, Reikard, and Bidlot (2012) use Reikard et al.'s (2011) regression model and log-normal distribution assumptions to produce probabilistic forecasts. López-Ruiz, Bergillos, and Ortega-Sánchez (2016) model the temporal dependencies of significant wave heights, peak wave periods and mean wave direction using a vector autoregressive model, and used them to produce medium to long term wave energy

[129] This subsection was written by Jooyoung Jeon.

forecasts. Jeon and Taylor (2016) model the temporal dependencies of significant wave heights and peak wave periods using a bivariate VARMA-GARCH (see also Section 2.3.11) to convert the two probabilistic forecasts into a probabilistic forecast of wave energy flux, finding this approach worked better than either univariate modelling of wave energy flux or bivariate modelling of wave energy flux and wind speed. Taylor and Jeon (2018) produce probabilistic forecasts for wave heights using a bivariate VARMA-GARCH model of wave heights and wind speeds, and using forecasts so as to optimise decision making for scheduling offshore wind farm maintenance vessels dispatched under stochastic uncertainty. On the same subject, Gilbert, Browell, and McMillan (2020b) use statistical post-processing of numerical wave predictions to produce probabilistic forecasts of wave heights, wave periods and wave direction and a logistic regression to determine the regime of the variables. They further applied the Gaussian copula to model temporal dependency but this did not improve their probabilistic forecasts of wave heights and periods.

3.4.8. Solar power forecasting[130]

Over the past few years, a number of forecasting techniques for photovoltaic (PV) power systems has been developed and presented in the literature. In general, the quantitative comparison among different forecast techniques is challenging, as the factors influencing the performance are numerous: the historical data, the weather forecast, the temporal horizon and resolution, and the installation conditions. A recent review by Sobri, Koohi-Kamali, and Rahim (2018) presents a comparative analysis of previous works, also including statistical errors. However, since the conditions and metrics used in each work were different, the comparison is not very meaningful. Dolara, Grimaccia, Leva, Mussetta, and Ogliari (2018) present relevant evaluation metrics for PV forecasting accuracy, while Leva, Mussetta, & Ogliari, 2019 compare their effectiveness and immediate comprehension. In term of forecast horizon for PV power systems, intraday (Nespoli et al., 2019) and the 24 h of the next day (Mellit et al., 2020) are considered the most important.

Nespoli et al. (2019) compared two of the most widely used and effective methods for the forecasting of the PV production: a method based on Multi-Layer Perceptron (MLP) and a hybrid method using artificial neural network combined with clear sky solar radiation (see also Sections 2.7.8 and 2.7.13). In the second case, the simulations are based on a feed-forward neural network (FFNN) but, among the inputs, the irradiation in clear sky conditions is provided. This method is called Physical Hybrid Artificial Neural Network (PHANN) and is graphically depicted in Fig. 9 (Dolara, Grimaccia, Leva, Mussetta, & Ogliari, 2015). PHANN method demonstrates better performance than classical NN methods. Fig. 10 shows a comparison between the measured and forecasted hourly output power of the PV plant for both sunny and cloudy days. The PHANN method shows good forecasting performance, especially for sunny days.

Ogliari, Dolara, Manzolini, and Leva (2017) compared the PV output power day-ahead forecasts performed by deterministic (based on three and five parameters electric equivalent circuit) and stochastic hybrid (based on artificial neural network models) methods aiming to find the best performance conditions. In general, there is no significant difference between the two deterministic models, with the three-parameter approach being slightly more accurate. Fig. 11 shows the daily value of normalised mean absolute error (NMAE%) for 216 days evaluated by using PHANN and three parameters electric circuit. The PHANN hybrid method achieves the best forecasting results, and only a few days of training can provide accurate forecasts.

Dolara et al. (2018) analysed the effect of different approaches in the composition of a training data-set for the day-ahead forecasting of PV power production based on NN. In particular, the influence of different data-set compositions on the forecast outcome has been investigated by increasing the size of the training set size and by varying the lengths of the training and validation sets, in order to assess the most effective training method of this machine learning approach. As a general comment on the reported results, it can be stated that a method that employs the same chronologically consecutive samples for training is best suited when the availability of historical data is limited (for example, in newly deployed PV plant), while training based on randomly mixed samples method, appears to be most effective in the case of a greater data availability. Generally speaking, ensembles composed of independent trials are most effective.

3.4.9. Long-term simulation for large electrical power systems[131]

In large electrical power systems with renewable energy dependence, the power generators need to be scheduled to supply the system demand (de Queiroz, 2016). In general, for modelling long-term renewables future behaviour, such as hydro, wind and solar photovoltaics (PV), stochastic scenarios should be included in the scheduling, usually in a dispatch optimisation problem under uncertainty – like described, for small systems, in Section 3.4.1 and, for wave forecasting, in Section 3.4.7. Due to the complexity and uncertainly associated, this problem is, in general, modelled with time series scenarios and multi-stage stochastic approaches. de Queiroz (2016) presented a review for hydrothermal systems, with a focus on the optimisation algorithms. Sections 3.4.6 and 3.4.8 explore the up-to-date methods for wind and PV solar power forecasting.

Here, we emphasise the importance of forecasting with simulation in the long-term renewable energy planning, especially in hydroelectric systems. In this context, due to the data spatial and temporal dependence structure, time series models are useful for future scenarios generation. Although the proposal could be forecasting for short-term planning and scheduling (as described in Section 3.4.6, 3.4.7, and Section 3.4.8), simulation strategies

130 This subsection was written by Sonia Leva.

131 This subsection was written by Fernando Luiz Cyrino Oliveira.

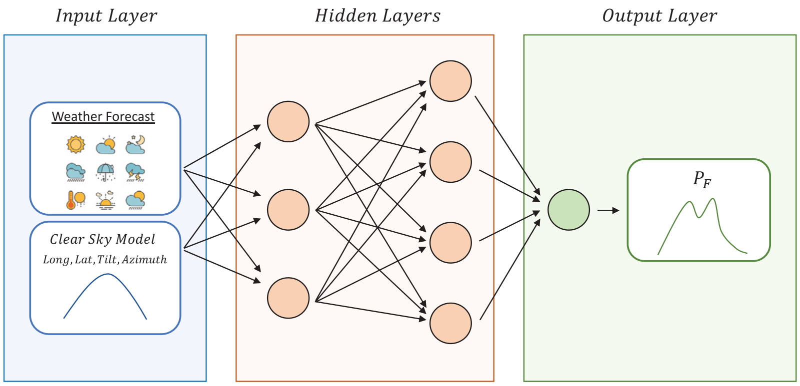

Fig. 9. Physical Hybrid Artificial Neural Network (PHANN) for PV power forecasting.

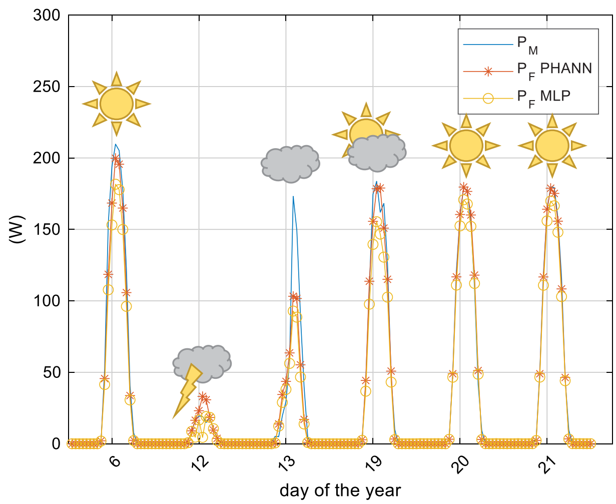

Fig. 10. Measured versus forecasted output power by MLP and PHANN methods.

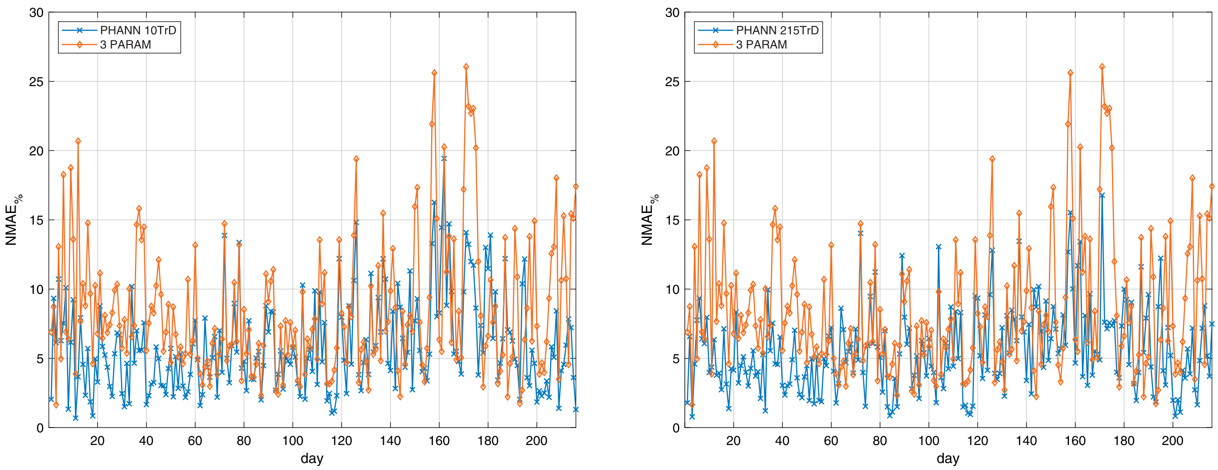

Fig. 11. Daily NMAE% of the PHANN method trained with 10 days (left) and with 215 days (right) compared with the three-parameters model.

are explored for considering and estimating uncertainty in medium and/or long-term horizons.

According to Hipel and McLeod (1994), stochastic processes of natural phenomena, such as the renewables ones, are, in general, stationary. One of the main features of hydroelectric generation systems is the strong dependence on hydrological regimes. To deal with this task, the literature focuses on two main classes for forecasting/

simulation streamflow data: physical and data-driven models (Zhang, Peng, Zhang, & Wang, 2015). Water resources management for hydropower generation and energy planning is one of the main challenges for decision-makers. At large, the hydrological data are transformed into the so-called affluent natural energy, that is used for scenarios simulation and serve as input for the optimisation algorithms (Oliveira, Souza, & Marcato, 2015). The current state-of-the-art models for this proposal are the periodic ones. Hipel and McLeod (1994) presented a wide range of possibilities, but the univariate periodic autoregressive (PAR, a periodic extension version of the ones presented in Section 2.3.4) is still the benchmark, with several enhanced versions. The approach fits a model to each period of the historical data and the residuals are simulated to generate new future versions of the time series, considered stationary. Among many others, important variations and alternative proposals to PAR with bootstrap procedures (see bootstrap details in Section 2.7.5), Bayesian dynamic linear models, spatial information and copulas versions (for copulas references, see Section 2.4.3) are detailed in Souza, Marcato, Dias, and Oliveira (2012), Marangon Lima, Popova, and Damien (2014), Lohmann, Hering, and Rebennack (2016), and de Almeida Pereira and Veiga (2019), respectively.

It is worth considering the need for renewables portfolio simulation. This led Pinheiro Neto, Domingues, Coimbra, de Almeida, Alves, and Calixto (2017) to propose a model to integrate hydro, wind and solar power scenarios for Brazilian data. For the Eastern United States, Shahriari and Blumsack (2018) add to the literature on the wind, solar and blended portfolios over several spatial and temporal scales. For China, Liu et al. (2020) proposed a multi-variable model, with a unified framework, to simulate wind and PV scenarios to compensate hydropower generation. However, in light of the aforementioned, one of the key challenges and trends for renewable electrical power systems portfolio simulation are still related to the inclusion of exogenous variables, such as climate, meteorological, calendar and economic ones, as mentioned in Section 3.4.2.

3.4.10. Collaborative forecasting in the energy sector[132]

As mentioned in Section 3.4.6, the combination of geographically distributed time series data, in a collaborative forecasting (or data sharing) framework, can deliver significant improvements in the forecasting accuracy of each individual renewable energy power plant. The same is valid for hierarchical load forecasting (Hong et al., 2019) and energy price forecasting (see Section 3.4.2). A review of multivariate time series forecasting methods can be found in Section 2.3.9 2.3.11 and Section 2.4.3. However, this data might have different owners, which are unwilling to share their data due to the following reasons: (*i*) personal or business sensitive information, (*ii*) lack of understanding about which data can and cannot be shared, and (*iii*) lack of information about economic (and technical) benefits from data sharing.

In order to tackle these limitations, recent research in energy time series forecasting is exploring two alternative (and potentially complementary) pathways: (*i*) privacy-preserving analytics, and (*ii*) data markets.

The role of privacy-preserving techniques applied collaborative forecasting is to combine time series data from multiple data owners in order to improve forecasting accuracy and keep data private at the same time. For solar energy forecasting, Berdugo, Chaussin, Dubus, Hebrail, and Leboucher (2011) described a method based on local and global analog-search that uses solar power time series from neighbouring sites, where only the timestamps and normalised weights (based on similarity) are exchanged and not the time series data. Zhang and Wang (2018) proposed, for wind energy forecasting with spatia-temporal data, a combination of ridge linear quantile regression and Alternating Direction Method of Multipliers (ADMM) that enables each data owner to autonomously solve its forecasting problem, while collaborating with the others to improve forecasting accuracy. However, as demonstrated by Gonçalves, Bessa, and Pinson (2021a), the mathematical properties of these algorithms should be carefully analysed in order to avoid privacy breaches (i.e., when a third party recovers the original data without consent).

An alternative approach is to design a market (or auction) mechanism for time series or forecasting data where the data owners are willing to sell their private (or confidential) data in exchange for an economic compensation (Agarwal, Dahleh, & Sarkar, 2019). The basic concept consists in pricing data as a function of privacy loss, but it can be also pricing data as a function of tangible benefits such as electricity market profit maximization. Gonçalves, Pinson, and Bessa (2021b) adapted for renewable energy forecasting the model described in Agarwal et al. (2019), by considering the temporal nature of the data and relating data price with the extra revenue obtained in the electricity market due to forecasting accuracy improvement. The results showed a benefit in terms of higher revenue resulting from the combination of electricity and data markets. With the advent of peer-to-peer energy markets at the domestic consumer level (Parag & Sovacool, 2016), smart meter data exchange between peers is also expected to increase and enable collaborative forecasting schemes. For this scenario, Yassine, Shirehjini, and Shirmohammadi (2015) proposed a game theory mechanism where a energy consumer maximizes its reward by sharing consumption data and a data aggregator can this data with a data analyst (which seeks data with the lowest possible price).

Finally, promoting data sharing via privacy-preserving or data monetisation can also solve data scarcity problems in some use cases of the energy sector, such as forecasting the condition of electrical grid assets (Fan, Nowaczyk, & Röognvaldsson, 2020). Moreover, combination of heterogeneous data sources (e.g., numerical, textual, categorical) is a challenging and promising avenue of future research in collaborative forecasting (Obst, Ghattas, Claudel, Cugliari, Goude, & Oppenheim, 2019).

132 This subsection was written by Ricardo Bessa.

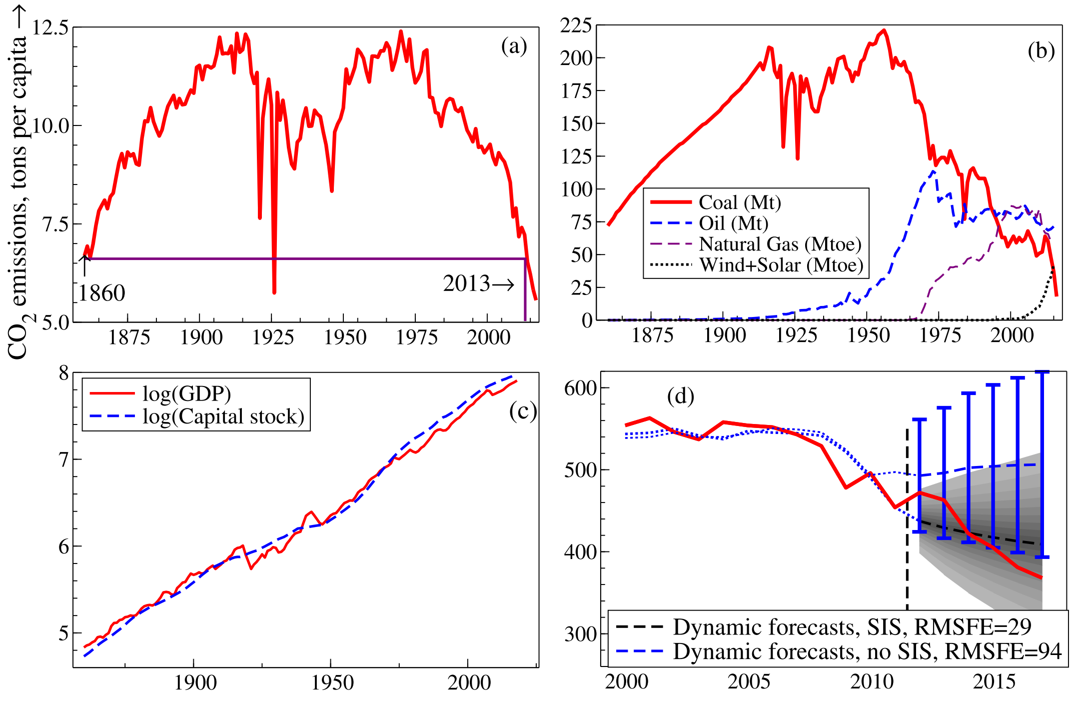

Fig. 12. (a) UK emissions, (b) energy sources in megatonnes (Mt) and megatonnes of oil equivalent (Mtoe), (c) economic variables, and (d) multi-step forecasts of CO2 emissions in Mt.

3.5. Environmental applications

3.5.1. Forecasting two aspects of climate change[133]

First into the Industrial Revolution, the UK is one of the first out: in 2013 its per capita CO_2 emissions dropped below their 1860 level, despite per capita real incomes being around 7-fold higher (Hendry, 2020). The model for forecasting UK CO_2 emissions was selected from annual data 1860–2011 on CO_2 emissions, coal and oil usage, capital and GDP, their lags and non-linearities (see Section 3.5.2 for higher frequency weather forecasts). Figs. 12(a) to 12(c) show the non-stationary time series with strong upward then downward trends, punctuated by large outliers from world wars, miners strikes plus shifts from legislation and technological change: (Castle & Hendry, 2020a). Saturation estimation at 0.1% using *Autometrics* (Doornik, 2018) retaining all other regressors, detected 4 step shifts coinciding with major policy interventions like the 2008 Climate Change Act, plus numerous outliers, revealing a cointegrated relation. The multi-step forecasts over 2012—2017 from a VAR in panel (d) of Fig. 12 show the advantage of using step-indicator saturation (SIS: Castle et al., 2015b).

We formulated a 3-equation simultaneous model of atmospheric CO_2 and Antarctic Temperature and Ice volume over 800,000 years of Ice Ages in 1000-year frequency (Kaufmann & Juselius, 2013; Paillard, 2001). Driven by non-linear functions of eccentricity, obliquity, and precession (see panels (a), (b), and (c) of Fig. 13 respectively), the model was selected with saturation estimation. Earth's orbital path is calculable into the future (Croll, 1875 and Milankovitch, 1969), allowing 100,000 years of multi-step forecasts at endogenous emissions. Humanity has affected climate since 10 thousand years ago (kya: Ruddiman, 2005), so we commence forecasts there. Forecasts over −10 to 100 with time series from 400kya in panels (d) to (f) of Fig. 13 show paths within the ranges of past data ±2.2SE (Pretis & Kaufmann, 2018).

Atmospheric CO_2 already exceeds 400 ppm (parts per million), dramatically outside the Ice-Age range (Sundquist & Keeling, 2009). Consequently, we conditionally forecast the next 100,000 years, simulating the potential climate for anthropogenic CO_2 (Castle & Hendry, 2020b) noting the 'greenhouse' temperature is proportional to the logarithm of CO_2 (Arrhenius, 1896). The orbital drivers will continue to influence all three variables but that relation is switched off in the scenario for 'exogenised' CO_2. The 110 dynamic forecasts conditional on 400 ppm and 560 ppm with ±2SE bands are shown in Fig. 14, panels (a) and (b) for Ice and Temperature respectively. The resulting global temperature rises inferred from these Antarctic temperatures would be dangerous, at more than 5 °C, with Antarctic temperatures positive for thousands of years (Pretis & Kaufmann, 2020; Vaks, Mason, Breitenbach, et al., 2019).

3.5.2. Weather forecasting[134]

The weather has a huge impact on our lives, affecting health, transport, agriculture (see also Section 3.8.10), energy use (see also Section 3.4), and leisure. Since Bjerknes (1904) introduced hydrodynamics and thermodynamics

[133] This section was written by David F. Hendry.

[134] This subsection was written by Thordis Thorarinsdottir.

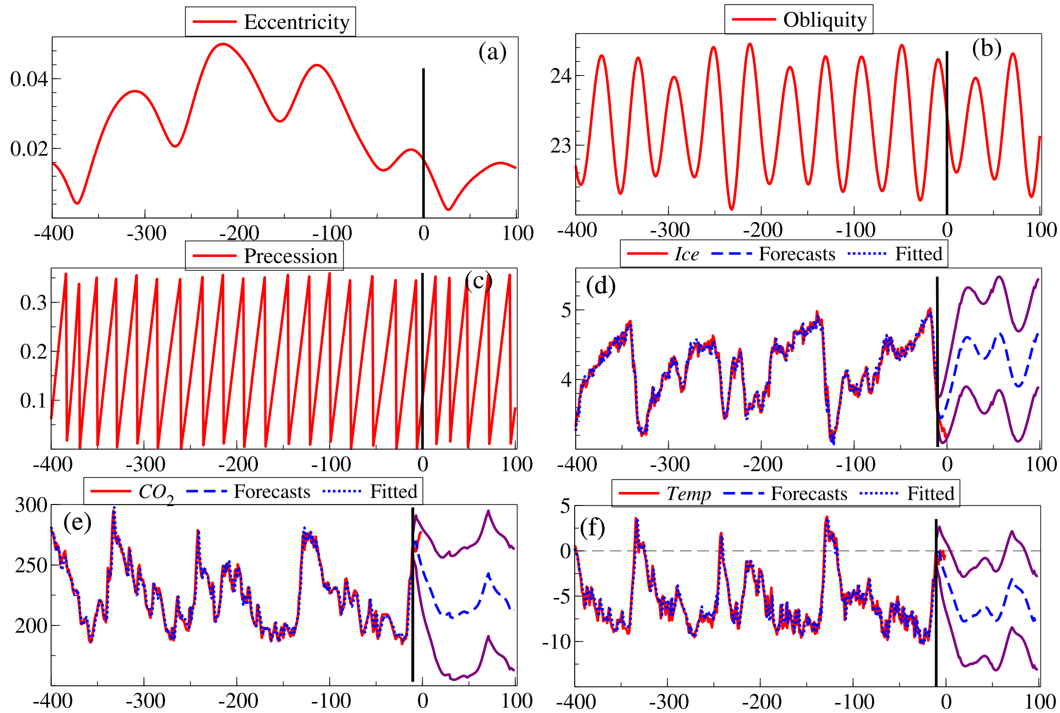

Fig. 13. Ice-Age data, model fits, and forecasts with endogenous CO_2.

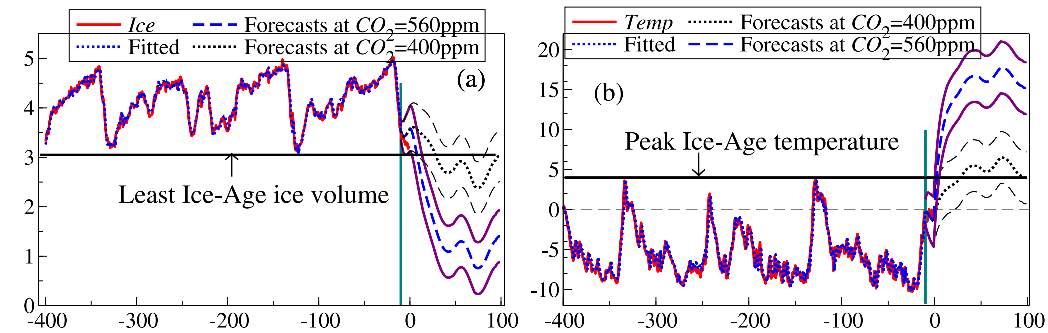

Fig. 14. Ice-Age simulations with exogenous CO_2.

into meteorology, weather prediction has been based on merging physical principles and observational information. Modern weather forecasting is based on numerical weather prediction (NWP) models that rely on accurate estimates of the current state of the climate system, including ocean, atmosphere and land surface. Uncertainty in these estimates is propagated through the NWP model by running the model for an ensemble of perturbed initial states, creating a weather forecast ensemble (Buizza, 2018; Toth & Buizza, 2019).

One principal concern in NWP modelling is that small-scale phenomena such as clouds and convective precipitation are on too small a scale to be represented directly in the models and must, instead, be represented by approximations known as parameterisations. Current NWP model development aims at improving both the grid resolution and the observational information that enters the models (Bannister, Chipilski, & Martinez-Alvarado, 2020; Leuenberger et al., 2020). However, for fixed computational resources, there is a trade-off between grid resolution and ensemble size, with a larger ensemble generally providing a better estimate of the prediction uncertainty. Recent advances furthermore include machine learning approaches (see Section 2.7.10) to directly model the small-scale processes, in particular cloud processes (see, for example, Gentine, Pritchard, Rasp, Reinaudi, & Yacalis, 2018; Rasp, Pritchard, & Gentine, 2018).

Despite rapid progress in NWP modelling, the raw ensemble forecasts exhibit systematic errors in both magnitude and spread (Buizza, 2018). Statistical post-processing is thus routinely used to correct systematic errors in calibration and accuracy before a weather forecast is issued; see Vannitsem, Wilks, and Messner (2018) for a recent review but also Sections 2.12.4 and 2.12.5. A fundamental challenge here is to preserve physical consistency across space, time and variables (see, for example, Heinrich, Hellton, Lenkoski, & Thorarinsdottir, 2020; Möller, Lenkoski, & Thorarinsdottir, 2013; Schefzik, Thorarinsdottir, & Gneiting, 2013). This is particularly important when the weather forecast is used as input for further prediction modelling, e.g., in hydrometeorology (Hemri, 2018; Hemri, Lisniak, & Klein, 2015).

At time scales beyond two weeks, the weather noise that arises from the growth of the initial uncertainty, becomes large (Royer, 1993). Sources of long-range predictability are usually associated with the existence of slowly evolving components of the earth system, including the El Niño Southern Oscillation (ENSO), monsoon rains, the Madden Julian Oscillation (MJO), the Indian Ocean dipole, and the North Atlantic Oscillation (NAO), spanning a wide range of time scales from months to decades (Hoskins, 2013; Vitart, Robertson, & Anderson, 2012). It is expected that, if a forecasting system is capable of reproducing these slowly evolving components, they may also be able to forecast them (Van Schaeybroeck & Vannitsem, 2018). The next step is then to find relationships between modes of low-frequency variability and the information needed by forecast users such as predictions of surface temperature and precipitation (Roulin & Vannitsem, 2019; Smith, Scaife, Eade, Athanasiadis, Bellucci, Bethke, Bilbao, Borchert, Caron, Counillon, Danabasoglu, Delowrth, Doblas-Reyes, Dunstone, Estella-Perez, Flavoni, Hermanson, Keenlyside, Kharin, Kimoto, Merryfield, Mignot, Mochizuki, Modali, Moneri, Müller, Nicolí, Ortega, Pankatz, Pholman, Robson, Ruggieri, Sospedra-Alfonso, Swingedouw, Wang, Wild, Yeager, Yang, & Zhang, 2020).

3.5.3. Air quality forecasting[135]

To preserve human health, European Commission stated in the Directive (2008/50/EC) that member states have to promptly inform the population when the particulate matter (PM) daily mean value exceeds (or is expected to exceed) the threshold of $50\mu g/m^3$. Therefore, systems have been designed in order to produce forecasts for up to three days in advance using as input the measured value of concentration and meteorological conditions. These systems can be classified in (*i*) data-driven models (Carnevale, Finzi, Pisoni, & Volta, 2016; Corani, 2005; Stadlober, Hormann, & Pfeiler, 2018, and Section 2.7), and (*ii*) deterministic chemical and transport models (Honoré, Menut, Bessagnet, Meleux, & Rou, 2007; Manders, Schaap, & Hoogerbrugge, 2009). In this section, a brief overview of the application of these systems to the high polluted area of Lombardy region, in Italy, will be presented.

Carnevale, Finzi, Pederzoli, Turrini, and Volta (2018) compared the results of three different forecasting systems based on neural networks, lazy learning models, and regression trees respectively. A single model has been identified for each monitoring station. In the initial configuration, only the last three PM measurements available were used to produce the forecast. In this configuration, the systems offered reasonable performance, with correlation coefficients ranging from 0.6 (lazy learning method) to 0.75 (neural network). The work also demonstrated that the performance of the ensemble of the three systems was better than the best model for each monitoring station (see also Section 2.6 for further discussions on forecast combinations).

Starting from the results of this work, a second configuration was implemented, using as input also the wind speed measured in the meteorological monitoring station closest to the measurement point of PM. The researchers observed an improvement in all performance indices, with the median of the correlation for the best model (neural networks) increasing from 0.75 to 0.82 and the RMSE dropping from $15\mu g/m^3$ to $7\mu g/m^3$.

One of the main drawbacks of data-driven models for air quality is that they provide information only in the point where the measurements are available. To overcome this limitation, recent literature has presented mixed deterministic and data-driven approaches (see, for example, Carnevale, Angelis, Finzi, Turrini, & Volta, 2020) which use the data assimilation procedure and offer promising forecasting performance.

From a practical point of view, critical issues regarding forecasting air quality include:

- Information collection and data access: even if regional authorities have to publicly provide data and information related to air quality and meteorology, the measured data are not usually available in real-time and the interfaces are sometimes not automated;
- Data quantity: the amount of information required by air quality forecasting systems is usually large, in particular towards the definition of the training and validation sets;
- Non-linear relationships: the phenomenon of accumulation of pollutants in atmosphere is usually affected by strong nonlinearities, which significantly impact the selection of the models and their performance;
- Unknown factors: it is a matter of fact that the dynamic of pollutants in atmosphere is affected by a large number of non-measurable variables (such as meteorological variables or the interaction with other non-measurable pollutants), largely affecting the capability of the models to reproduce the state of the atmosphere.

3.5.4. Forecasting and decision making for floods and water resources management[136]

In Water Resources and Flood Risk Management, decision makers are frequently confronted with the need of taking the most appropriate decisions not knowing what will occur in the future. To support their decision-making under uncertainty, decision theory (Berger, 1985; Bernardo, 1994; DeGroot, 2004) invokes Bayesian informed decision approaches, which find the most appropriate

[135] This subsection was written by Claudio Carnevale.

[136] This subsection was written by Ezio Todini.

decision by maximising (or minimising) the expected value of a "utility function", thus requiring its definition, together with the estimation of a "predictive probability" density (Berger, 1985) due to the fact that utility functions are rarely linear or continuous. Consequently, their expected value does not coincide with the value assumed on the predicted "deterministic" expected value. Accordingly, overcoming the classical 18th century "mechanistic" view by resorting into probabilistic forecasting approaches becomes essential (see also Section 2.6.2).

The failure of decision-making based on deterministic forecasts in the case of Flood Risk Management is easily shown through a simple example. At a river section, the future water level provided by a forecast is uncertain and can be described by a Normal distribution with mean 10 meters and standard deviation of 5 m. Given a dike elevation of 10.5 m, damages may be expected as zero if water level falls below the dike elevation and linearly growing when level exceeds it with a factor of 10^6 dollars. If one assumes the expect value of forecast as the deterministic prediction to compute the damage the latter will result equal to zero, while if one correctly integrates the damage function times the predictive density the estimated expected damage will results into 6.59 millions of dollars and educated decisions on alerting or not the population or evacuating or not a flood-prone area can be appropriately taken (see also Section 3.6).

Water resources management, and in particular reservoirs management, aim at deriving appropriate operating rules via long term expected benefits maximisation. Nonetheless, during flood events decision makers must decide how much to preventively release from multipurpose reservoirs in order to reduce dam failure and downstream flooding risks the optimal choice descending from trading-off between loosing future water resource vs the reduction of short term expected losses.

This is obtained by setting up an objective function based on the linear combination of long and short term "expected losses", once again based on the available probabilistic forecast. This Bayesian adaptive reservoir management approach incorporating into the decision mechanism the forecasting information described by the short-term predictive probability density, was implemented on the lake Como since 1997 (Todini, 1999, 2017) as an extension of an earlier original idea (Todini, 1991). This resulted into:

- a reduction of over 30% of the city of Como frequency;
- an average reduction of 12% of the water deficit;
- an increase of 3% in the electricity production.

Lake Como example clearly shows that instead of basing decisions on the deterministic prediction, the use of a Bayesian decision scheme, in which model forecasts describe the predictive probability density, increases the reliability of the management scheme by essentially reducing the probability of wrong decisions (Todini, 2017, 2018).

3.6. Social good and demographic forecasting

3.6.1. Healthcare[137]

There are many decisions that depend on the quality of forecasts in the health care system, from capacity planning to layout decisions to the daily schedules. In general, the role of forecasting in health care is to inform both clinical and non-clinical decisions. While the former concerns decisions related to patients and their treatments (Makridakis, Kirkham, Wakefield, Papadaki, Kirkham, & Long, 2019), the latter involves policy/management, and supply chain decisions that support the delivery of high-quality care for patients.

A number of studies refer to the use of forecasting methods to inform clinical decision making. These methods are used to screen high risk patients for preventative health care (Chen, Wang, & Hung, 2015; Santos, Abreu, Garca-Laencina, Simão, & Carvalho, 2015; Uematsu, Kunisawa, Sasaki, Ikai, & Imanaka, 2014; van der Mark, van Wonderen, Mohrs, van Aalderen, ter Riet, & Bindels, 2014), to predict mental health issues (Shen, Jia, Nie, Feng, Zhang, Hu, Chua, & Zhu, 2017; Tran, Phung, Luo, Harvey, Berk, & Venkatesh, 2013), to assist diagnosis and disease progression (Ghassemi et al., 2015; Ma, Chitta, Zhou, You, Sun, & Gao, 2017; Pierce, Hess, Kline, Shah, Breslin, Branda, Pencille, Asplin, Nestler, Sadosty, Stiell, Ting, & Montori, 2010; Qiao, Wu, Ge, & Fan, 2019), to determine prognosis (Dietzel et al., 2010; Ng, Stein, Ning, & Black-Schaffer, 2007), and to recommend treatments for patients (Kedia & Williams, 2003; Scerri, De Goumoens, Fritsch, Van Melle, Stiefel, & So, 2006; Shang, Ma, Xiao, & Sun, 2019). Common forecasting methods to inform clinical decisions include time series (see Sections 2.3.1, 2.3.4 and 2.3.5), regression models (see Section 2.3.2), classification trees (see Section 2.7.12), neural networks (see Section 2.7.8), Markov models (see Section 2.3.12) and Bayesian networks. These models utilise structured and unstructured data including clinician notes (Austin & Kusumoto, 2016; Labarere, Bertrand, & Fine, 2014) which makes the data pre-processing a crucial part of the forecasting process in clinical health care.

One of the aspects of the non-clinical forecasting that has received the most attention in both research and application is the policy and management. Demand forecasting is regularly used in Emergency Departments (Arora, Taylor, & Mak, 2020; Choudhury & Urena, 2020; Khaldi, El Afia, & Chiheb, 2019; Rostami-Tabar & Ziel, 2020), ambulance services (Al-Azzani, Davari, & England, 2020; Setzler, Saydam, & Park, 2009; Vile, Gillard, Harper, & Knight, 2012; Zhou & Matteson, 2016) and hospitals with several different specialities (McCoy, Pellegrini, & Perlis, 2018; Ordu, Demir, & Tofallis, 2019; Zhou, Zhao, Wu, Cheng, & Huang, 2018) to inform operational, tactical and strategic planning. The common methods used for this purpose include classical ARIMA and exponential smoothing methods, regression, singular spectrum analysis, Prophet, Double-Seasonal Holt-Winter, TBATS and Neural Networks. In public health, forecasting can guide policy and planning. Although it has a wider definition,

[137] This section was written by Bahman Rostami-Tabar.

the most attention is given to Epidemic forecasting (see also Section 3.6.2).

Forecasting is also used in both national and global health care supply chains, not only to ensure the availability of medical products for the population but also to avoid excessive inventory. Additionally, the lack of accurate demand forecast in a health supply chain may cost lives (Baicker, Chandra, & Skinner, 2012) and has exacerbated risks for suppliers (Levine, Pickett, Sekhri, & Yadav, 2008). Classical exponential smoothing, ARIMA, regression and Neural Network models have been applied to estimate the drug utilisation and expenditures (Dolgin, 2010; Linnér, Eriksson, Persson, & Wettermark, 2020), blood demand (Fortsch & Khapalova, 2016), hospital supplies (Gebicki, Mooney, Chen, & Mazur, 2014; Riahi, Hosseini-Motlagh, & Teimourpour, 2013) and demand for global medical items (Amarasinghe, Wichmann, Margolis, & Mahoney, 2010; Hecht & Gandhi, 2008; van der Laan, van Dalen, Rohrmoser, & Simpson, 2016). It is important to note that, while the demand in a health care supply chain has often grouped and hierarchical structures (Mircetica, Rostami-Tabar, Nikolicica, & Maslarica, 2020, see also Section 2.10.1), this has not been well investigated and needs more attention.

3.6.2. Epidemics and pandemics[138]

Pandemics and epidemics both refer to disease outbreaks. An epidemic is a disease outbreak that spreads across a particular region. A pandemic is defined as spread of a disease worldwide. Forecasting the evolution of a pandemic or an epidemic, the growth of cases and fatalities for various horizons and levels of granularity, is a complex task with raw and limited data – as each disease outbreak type has unique features with several factors affecting the severity and the contagiousness. Be that as it may, forecasting becomes an paramount task for the countries to prepare and plan their response (Nikolopoulos, 2020), both in healthcare and the supply chains (Beliën & Forcé, 2012, see also Section 3.6.1 and Section 3.2.2).

Successful forecasting methods for the task include time-series methods (see Section 2.3), epidemiological and agent-based models (see Section 2.7.3), metapopulation models, approaches in metrology (Nsoesie, Mararthe, & Brownstein, 2013), machine and deep learning methods (Yang, Zeng, Wang, Wong, Liang, Zanin, Liu, Cao, Gao, Mai, Liang, Liu, Li, Li, Ye, Guan, Yang, Li, Luo, Xie, Liu, Wang, Zhang, Wang, Zhong, & He, 2020). Andersson, Kühlmann-Berenzon, Linde, Schiöler, Rubinova, and Frisén (2008) used regression models for the prediction of the peak time and volume of cases for a pandemic with evidence from seven outbreaks in Sweden. Yaffee, Nikolopoulos, Reilly, Crone, Wagoner, Douglas, Amman, Ksiazek, and Mills (2011) forecasted the evolution of the Hantavirus epidemic in USA and compared causal and machine-learning methods with time-series methods and found that univariate methods quite successful. Soebiyanto, Adimi, and Kiang (2010) used ARIMA models for successfully short-term forecasting of influenza weekly cases. Shaman and Karspeck (2012) used Kalman filter based SIR epidemiological models to forecast the peak time of influenza 6–7 weeks ahead.

For COVID-19, Petropoulos and Makridakis (2020) applied a multiplicative exponential smoothing model (see also Section 2.3.1) for predicting global number of confirmed cases, with very successful results both for point forecasts and prediction intervals. This article got serious traction with 100,000 views and 300 citations in the first twelve months since its publication, thus evidencing the importance of such empirical investigations. There has been a series of studies focusing on predicting deaths in the USA and European countries for the first wave of the COVID-19 pandemic (IHME COVID-19 health service utilization forecasting team & Murray, 2020a, 2020b). Furthermore, Petropoulos, Makridakis, and Stylianou (2020) expanded their investigation to capture the continuation of both cases and deaths as well as their uncertainty, achieving high levels of forecasting accuracy for ten-days-ahead forecasts over a period of four months. Along the same lines, Doornik et al. (2020b) have been publishing real-time accurate forecasts of confirmed cases and deaths from mid-March 2020 onwards. Their approach is based on extraction of trends from the data using machine learning.

Pinson and Makridakis (2020) organised a debate between Taleb and Ioannidis on forecasting pandemics. Ioannidis, Cripps, and Tanner (2020) claim that forecasting for COVID-19 has by and large failed. However they give recommendations of how this can be averted. They suggest that the focus should be on predictive distributions and models should be continuously evaluated. Moreover, they emphasise the importance of multiple dimensions of the problem (and its impact). Taleb et al. (2020) discuss the dangers of using naive, empirical approaches for fat-tailed variables and tail risk management. They also reiterate the inefficiency of point forecasts for such phenomena.

Finally, Nikolopoulos, Punia, Schäfers, Tsinopoulos, and Vasilakis (2020) focused on forecast-driven planning, predicting the growth of COVID-19 cases and the respective disruptions across the supply chain at country level with data from the USA, India, UK, Germany, and Singapore. Their findings confirmed the excess demand for groceries and electronics, and reduced demand for automotive – but the model also proved that the earlier a lock-down is imposed, the higher the excess demand will be for groceries. Therefore, governments would need to secure high volumes of key products before imposing lock-downs; and, when this is not possible, seriously consider more radical interventions such as rationing.

Dengue is one of the most common epidemic diseases in tropical and sub-tropical regions of the world. Estimates of World Health Organisation reveals that about half of the world's population is now at risk for Dengue infection (Romero, Olivero, Real, & Guerrero, 2019). *Aedes aegypti* and *Aedes albopictus* are the principal vectors of dengue transmission and they are highly domesticated mosquitoes. Rainfall, temperature and relative humidity are thought of as important factors attributing towards the growth and dispersion of mosquito vectors and potential of dengue outbreaks (Banu, Hu, Hurst, & Tong, 2011).

[138] This subsection was written by Konstantinos Nikolopoulos & Thiyanga S. Talagala.

In reviewing the existing literature, two data types have been used to forecast dengue incidence: (*i*) spatio-temporal data: incidence of laboratory-confirmed dengue cases among the clinically suspected patients (Naish, Dale, Mackenzie, McBride, Mengersen, & Tong, 2014), (*ii*) web-based data: Google trends, tweets associated with Dengue cases (de Almeida Marques-Toledo et al., 2017).

SARIMA models (see also Section 2.3.4) have been quite popular in forecasting laboratory-confirmed dengue cases (Gharbi et al., 2011; Martinez & Silva, 2011; Promprou, Jaroensutasinee, & Jaroensutasinee, 2006). Chakraborty, Chattopadhyay, and Ghosh (2019) used a hybrid model combining ARIMA and neural network autoregressive (NNAR) to forecast dengue cases. In light of biological relationships between climate and transmission of *Aedes* mosquitoes, several studies have used additional covariates such as, rainfall, temperature, wind speed, and humidity to forecasts dengue incidence (Banu et al., 2011; Naish et al., 2014; Talagala, 2015). Poisson regression model has been widely used to forecast dengue incidence using climatic factors and lagged time between dengue incidence and weather variables (Hii, Zhu, Ng, Ng, & Rocklöv, 2012; Koh, Spindler, Sandgren, & Jiang, 2018). Several researchers looked at the use of Quasi-Poisson and negative binomial regression models to accommodate over dispersion in the counts (Lowe et al., 2011; Wang, Jiang, Fan, Wang, & Liu, 2014). Cazelles, Chavez, McMichael, and Hales (2005) used wavelet analysis to explore the dynamic of dengue incidence and wavelet coherence analyses was used to identify time and frequency specific association with climatic variables. de Almeida Marques-Toledo et al. (2017) took a different perspective and look at weekly tweets to forecast Dengue cases. Rangarajan, Mody, and Marathe (2019) used Google trend data to forecast Dengue cases. Authors hypothesised that web query search related to dengue disease correlated with the current level of dengue cases and thus may be helpful in forecasting dengue cases.

A direction for future research in this field is to explore the use of spatio-temporal hierarchical forecasting (see Section 2.10).

3.6.3. Forecasting mortality[139]

Actuarial, Demographic, and Health studies are some examples where mortality data are commonly used. A valuable source of mortality information is the Human Mortality Database (HMD), a database that provides mortality and population data for 41 mainly developed countries. Additionally, at least five country-specific databases are devoted to subnational data series: Australian, Canadian, and French Human Mortality Databases, United States and Japan Mortality Databases. In some situations, the lack of reliable mortality data can be a problem, especially in developing countries, due to delays in registering or miscounting deaths (Checchi & Roberts, 2005). Analysis of National Causes of Death for Action (ANACONDA) is a valuable tool that assesses the accuracy and completeness of data for mortality and cause of death by checking for potential errors and inconsistencies (Mikkelsen, Moesgaard, Hegnauer, & Lopez, 2020).

The analysis of mortality data is fundamental to public health authorities and policymakers to make decisions or evaluate the effectiveness of prevention and response strategies. When facing a new pandemic, mortality surveillance is essential for monitoring the overall impact on public health in terms of disease severity and mortality (Setel, AbouZahr, Atuheire, Bratschi, Cercone, Chinganya, Clapham, Clark, Congdon, de Savigny, Karpati, Nichols, Jakob, Mwanza, Muhwava, Nahmias, Ortiza, & Tshangelab, 2020; Vestergaard, Nielsen, Richter, Schmid, Bustos, Braeye, Denissov, Veideman, Luomala, Möttönen, Fouillet, Caserio-Schönemann, an der Heiden, Uphoff, Lytras, Gkolfinopoulou, Paldy, Domegan, O'Donnell, de' Donato, Noccioli, Hoffmann, Velez, England, van Asten, White, Tønnessen, da Silva, Rodrigues, Larrauri, Delgado-Sanz, Farah, Galanis, Junker, Perisa, Sinnathamby, Andrews, O'Doherty, Marquess, Kennedy, Olsen, Pebody, ECDC Public Health Emergency Team for COVID-19, Krause, & Mlbak, 2020). A useful metric is excess mortality and is the difference between the observed number of deaths and the expected number of deaths under "normal" conditions (Aron & Muellbauer, 2020; Checchi & Roberts, 2005). Thus, it can only be estimated with accurate and high-quality data from previous years. Excess mortality has been used to measure the impact of heat events (Limaye, Vargo, Harkey, Holloway, & Patz, 2018; Matte, Lane, & Ito, 2016), pandemic influenza (Nielsen, Mazick, Andrews, Detsis, Fenech, Flores, Foulliet, Gergonne, Green, Junker, Nunes, O'Donnell, Oza, Paldy, Pebody, Reynolds, Sideroglou, E, Simon-Sofia, Uphoff, Van Asten, Virtanen, Wuillaume, & Molbak, 2013; Nunes, Viboud, Machado, Ringholz, Rebelo-de Andrade, Nogueira, & Miller, 2011), and nowadays COVID-19 (Nogueira, de Araújo Nobre, Nicola, Furtado, & Carneiro, 2020; Ritchie, Ortiz-Ospina, Beltekian, Mathieu, Hasell, Macdonald, Giattino, & Roser, 2020; Shang & Xu, 2021; Sinnathamby, Whitaker, Coughlan, Bernal, Ramsay, & Andrews, 2020, and Section 3.6.2), among others. Excess mortality data have been making available by the media publications *The Economist*, *The New York Times* and *The Financial Times*. Moreover, a monitoring system of the weekly excess mortality in Europe has been performed by the EuroMOMO project (Vestergaard et al., 2020).

An essential use of mortality data for those individuals at age over 60 is in the pension and insurance industries, whose profitability and solvency crucially rely on accurate mortality forecasts to adequately hedge longevity risks (see, e.g., Shang & Haberman, 2020a, 2020b). Longevity risk is a potential systematic risk attached to the increasing life expectancy of annuitants, and it is an important factor to be considered when determining a sustainable government pension age (see, e.g., Hyndman, Zeng, & Shang, 2021, for Australia). The price of a fixed-term or lifelong annuity is a random variable, as it depends on the value of zero-coupon bond price and mortality forecasts. The zero-coupon bond price is a function of interest rate (see Section 3.3.6) and is comparably more stable than the retirees' mortality forecasts.

Several methodologies were developed for mortality modelling and forecasting (Booth & Tickle, 2008; Janssen,

[139] This subsection was written by Clara Cordeiro & Han Lin Shang.

2018). These methods can be grouped into three categories: expectation, explanation, and extrapolation (Booth & Tickle, 2008).

The expectation approach is based on the subjective opinion of experts (see also Section 2.11.4), who set a long-run mortality target. Methods based on expectation make use of experts' opinions concerning future mortality or life expectancy with a specified path and speed of progression towards the assumed value (Continuous Mortality Investigation, 2020). The advantage of this approach is that demographic, epidemiological, medical, and other relevant information may be incorporated into the forecasts. The disadvantages are that such information is subjective and biased towards experts' opinions, and it only produces scenario-based (see Section 2.11.5) deterministic forecasts (Ahlburg & Vaupel, 1990; Wong-Fupuy & Haberman, 2004).

The explanation approach captures the correlation between mortality and the underlying cause of death. Methods based on the explanation approach incorporate medical, social, environmental, and behavioural factors into mortality modelling. Example include smoking and disease-related mortality models. The benefit of this approach is that mortality change can be understood from changes in related explanatory variables; thus, it is attractive in terms of interpretability (Gutterman & Vanderhoof, 1998).

The extrapolative approach is considered more objective, easy to use and more likely to obtain better forecast accuracy than the other two approaches (Janssen, 2018). The extrapolation approach identifies age patterns and trends in time which can be then forecasted via univariate and multivariate time series models (see Section 2.3). In the extrapolation approach, many parametric and nonparametric methods have been proposed (see, e.g., Alho & Spencer, 2005; Hyndman & Ullah, 2007; Shang, Booth, & Hyndman, 2011). Among the parametric methods, the method of Heligman and Pollard (1980) is well-known. Among the nonparametric methods, the Lee–Carter model (Lee & Carter, 1992), Cairns-Blake-Dowd model (Cairns et al., 2009; Dowd, Cairns, Blake, Coughlan, Epstein, & Khalaf-Allah, 2010), and functional data model (Hyndman & Ullah, 2007, and Section 2.3.10), as well as their extensions and generalisations are dominant. The time-series extrapolation approach has the advantage of obtaining a forecast probability distribution rather than a deterministic point forecast and, also, enable the determination of forecast intervals (Booth & Tickle, 2008).

Janssen (2018) presents a review of the advances in mortality forecasting and possible future research challenges.

3.6.4. *Forecasting fertility*[140]

Aside from being a driver of population forecasts (see Section 2.3.7), fertility forecasts are vital for planning maternity services and anticipating demand for school places. The key challenge relates to the existence of, and interaction between, the quantum (how many?) and tempo (when?) components (Booth, 2006). This intrinsic dependence on human decisions means that childbearing behaviour is influenced by numerous factors acting at different levels, from individual characteristics to societal change (Balbo, Billari, & Mills, 2013). An important methodological challenge for many low- and middle-income countries is fertility estimation, due to deficiencies in vital statistics caused by inadequate birth registration systems (AbouZahr, de Savigny, Mikkelsen, Setel, Lozano, & Lopez, 2015; Moultrie, Dorrington, Hill, Hill, Timæ us, & Zaba, 2013; Phillips, Adair, & Lopez, 2018). Such countries are often also in the process of transitioning from high to low fertility, which induces greater forecast uncertainty compared to low-fertility countries (Programme, 2019).

A range of statistical models have been proposed to forecast fertility – see Booth (2006), Bohk-Ewald, Li, and Myrskylä (2018), and Shang and Booth (2020) for reviews. The Lee–Carter model (Lee & Carter, 1992) originally developed for mortality forecasting (see Section 3.6.3) has been applied to fertility (Lee, 1993), with extensions in functional data (Hyndman & Ullah, 2007) and Bayesian (Wiśniowski, Smith, Bijak, Raymer, & Forster, 2015) contexts. Other notable extrapolative methods include the cohort-ARIMA model of De Beer (1985, 1990) – see Section 2.3.4 – and the linear extrapolation method of Myrskylä, Goldstein, and Cheng (2013). Many parametric models have been specified to describe the shapes of fertility curves (Brass, 1974; Evans, 1986; Hoem, Madsen, Nielsen, Ohlsen, Hansen, & Rennermalm, 1981; Schmertmann, 2003), with forecasts obtained through time series extrapolations of the parameters (Congdon, 1990; De Iaco & Maggio, 2016; Knudsen, McNown, & Rogers, 1993). Bayesian methods have been used to borrow strength across countries (for example, Alkema et al., 2011; Schmertmann, Zagheni, Goldstein, & Myrskylä, 2014), with Ellison, Dodd, and Forster (2020) developing a hierarchical model in the spirit of the latter. The top-down approach (see Section 2.10.1) of the former, which is used by the United Nations, projects the aggregate Total Fertility Rate (TFR) measure probabilistically (also see Tuljapurkar & Boe, 1999) before decomposing it by age. Hyppöla, Tunkelo, and Törnqvist (1949) provide one of the earliest known examples of probabilistic fertility forecasting (Alho & Spencer, 2005).

Little work has been done to compare forecast performance across this broad spectrum of approaches. The study of Bohk-Ewald et al. (2018) is the most comprehensive to date. Most striking is their finding that few methods can better the naive freezing of age-specific rates, and those that can differ greatly in method complexity (see also Section 2.5.2). A recent survey of fertility forecasting practice in European statistical offices (Gleditsch & Syse, 2020) found that forecasts tend to be deterministic and make use of expert panels (see Section 2.11.4). Expert elicitation techniques are gaining in sophistication, highlighted by the protocol of Statistics Canada (Dion, Galbraith, & Sirag, 2020) which requests a full probability distribution of the TFR.

A promising avenue is the development of forecasting methods that incorporate birth order (parity) information, supported by evidence from individual-level analyses (for

140 This subsection was written by Joanne Ellison.

example, Fiori, Graham, & Feng, 2014). Another underexplored area is the integration of survey data into fertility forecasting models, which tend to use vital statistics alone when they are of sufficient quality (see Rendall, Handcock, & Jonsson, 2009; Zhang & Bryant, 2019, for Bayesian fertility estimation with imperfect census data). Alternative data sources also have great potential. For example, Wilde, Chen, and Lohmann (2020) use Google data to predict the effect of COVID-19 on US fertility in the absence of vital statistics. Lastly, investigation of the possible long-term impacts of delayed motherhood in high-income countries, alongside developments in assisted reproduction technology such as egg freezing, is required (see, for example, Sobotka & Beaujouan, 2018).

3.6.5. Forecasting migration[141]

Migration forecasts are needed both as crucial input into population projections (see Section 2.3.7), as well as standalone predictions, made for a range of users, chiefly in the areas of policy and planning. At the same time, migration processes are highly uncertain and complex, with many underlying and interacting drivers, which evade precise conceptualisation, definitions, measurement, and theoretical description (Bijak & Czaika, 2020). Given the high level of the predictive uncertainty, and the non-stationary character of many migration processes (Bijak & Wiśniowski, 2010), the current state of the art of forward-looking migration studies reflects therefore a shift from prediction to the use of forecasts as contingency planning tools (idem).

Reviews of migration forecasting methods are available in Bijak (2010) and Sohst, Tjaden, de Valk, and Melde (2020). The applications in official statistics, with a few exceptions, are typically based on various forms scenario-based forecasting with judgment (see Section 2.11.5), based on pre-set assumptions (for an example, see Abel, 2018). Such methods are particularly used for longer time horizons, of a decade or more, so typically in the context of applications in population projections, although even for such long time horizons calibrated probabilistic methods have been used as well (Azose et al., 2016).

The mainstream developments in migration forecasting methodology, however, include statistical and econometric methods discussed in Section 2.3, such as time series models, both uni- and multivariate (for example, Bijak, 2010; Bijak, Disney, Findlay, Forster, Smith, & Wiśniowski, 2019; Gorbey, James, & Poot, 1999), econometric models (for example, Brücker & Siliverstovs, 2006; Cappelen, Skjerpen, & Tønnessen, 2015), Bayesian hierarchical models (Azose & Raftery, 2015), and dedicated methods, for example for forecasting data structured by age (Raymer & Wiśniowski, 2018). In some cases, the methods additionally involve selection and combining forecasts through Bayesian model selection and averaging (Bijak, 2010, see also Section 2.5 and Section 2.6). Such models can be expected to produce reasonable forecasts (and errors) for up to a decade ahead (Bijak & Wiśniowski, 2010), although this depends on the migration flows being forecast, with some processes (e.g., family migration) more predictable than other (e.g., asylum). Another recognised problem with models using covariates is that those can be endogenous to migration (e.g., population) and also need predicting, which necessitates applying structured models to prevent uncertainty from exploding.

The methodological gaps and current work in migration forecasting concentrate in a few key areas, notably including causal (mechanistic) forecasting based on the process of migrant decision making (Willekens, 2018); as well as early warnings and 'nowcasting' of rapidly changing trends, for example in asylum migration (Napierała, Hilton, Forster, Carammia, & Bijak, 2021). In the context of early warnings, forays into data-driven methods for changepoint detection, possibly coupled with the digital trace and other high-frequency 'Big data', bear particular promise. At the same time, coherent uncertainty description across a range of time horizons, especially in the long range (Azose & Raftery, 2015), remains a challenge, which needs addressing for the sake of proper calibration of errors in the population forecasts, to which these migration components contribute.

3.6.6. Forecasting risk for violence and wars[142]

Can we predict the occurrence of WW3 in the next 20 years? Is there any trend in the severity of wars?

The study of armed conflicts and atrocities, both in terms of frequency over time and the number of casualties, has received quite some attention in the scientific literature and the media (e.g., Cederman, 2003; Friedman, 2015; Hayes, 2002; Norton-Taylor, 2015; Richardson, 1948, 1960), falling within the broader discussion about violence (Berlinski, 2009; Goldstein, 2011; Spagat, Mack, Cooper, & Kreutz, 2009), with the final goal of understanding whether humanity is becoming less belligerent (Pinker, 2011), or not (Braumoeller, 2019).

Regarding wars and atrocities, the public debate has focused its attention on the so-called *Long Peace Theory* (Gaddis, 1989), according to which, after WW2, humanity has experienced the most peaceful period in history, with a decline in the number and in the severity of bloody events. Scholars like Mueller (2009a, 2009b) and Pinker (2011, 2018) claim that sociological arguments and all statistics suggest we live in better times, while others like Gray (2015a, 2015b) and Mann (2018) maintain that those statistics are often partial and misused, the derived theories weak, and that war and violence are not declining but only being transformed. For Mann, the Long Peace proves to be ad-hoc, as it only deals with Western Europe and North America, neglecting the rest of the world, and the fact that countries like the US have been involved in many conflicts out of their territories after WW2.

Recent statistical analyses confirm Gray's and Mann's views: empirical data do not support the idea of a decline in human belligerence (no clear trend appears), and in its severity. Armed conflicts show long inter-arrival times, therefore a relative peace of a few decades means nothing statistically (Cirillo & Taleb, 2016b). Moreover, the distribution of war casualties is extremely fat-tailed (Clauset,

141 This subsection was written by Jakub Bijak.

142 This subsection was written by Pasquale Cirillo.

2018; Clauset & Gleditsch, 2018), often with a tail exponent $\xi = \frac{1}{\alpha} > 1$ (Cirillo & Taleb, 2016b), indicating a possibly infinite mean, i.e., a tremendously erratic and unforeseeable phenomenon (see Section 2.3.22). An only apparently infinite-mean phenomenon though (Cirillo & Taleb, 2019), because no single war can kill more than the entire world population, therefore a finite upper bound exists, and all moments are necessarily finite, even if difficult to estimate. Extreme value theory (Embrechts et al., 2013) can thus be used to correctly model tail risk and make prudential forecasts (with many caveats like in Scharpf, Schneider, Nöh, & Clauset, 2014), while avoiding naive extrapolations (Taleb et al., 2020).

As history teaches (Nye, 1990), humanity has already experienced periods of relative regional peace, like the famous Paces Romana and Sinica. The present Pax Americana is not enough to claim that we are structurally living in a more peaceful era. The Long Peace risks to be another apophenia, another example of Texan sharpshooter fallacy (Carroll, 2003).

Similar mistakes have been made in the past. Buckle (1858) wrote: "that [war] is, in the progress of society, steadily declining, must be evident, even to the most hasty reader of European history. If we compare one country with another, we shall find that for a very long period wars have been becoming less frequent; and now so clearly is the movement marked, that, until the late commencement of hostilities, we had remained at peace for nearly forty years: a circumstance unparalleled [...] in the affairs of the world". Sadly, Buckle was victim of the illusion coming from the Pax Britannica (Johnston, 2008): the century following his prose turned out to be the most murderous in human history.

3.7. Systems and humans

3.7.1. Support systems[143]

Forecasting in businesses is a complicated procedure, especially when predicting numerous, diverse series (see Section 2.7.4), dealing with unstructured data of multiple sources (see Section 2.7.1), and incorporating human judgment (Lim & O'Connor, 1996a, but also Section 2.11). In this respect, since the early 80's, various Forecasting Support Systems (FSSs) have been developed to facilitate forecasting and support decision making (Kusters, McCullough, & Bell, 2006). Rycroft (1993) provides an early comparative review of such systems, while many studies strongly support their utilisation over other forecasting alternatives (Sanders & Manrodt, 2003; Tashman & Leach, 1991).

In a typical use-case scenario, the FSSs will retrieve the data required for producing the forecasts, will provide some visualisations and summary statistics to the user, allow for data pre-processing, and then produce forecasts that may be adjusted according to the preferences of the user. However, according to Ord and Fildes (2013), effective FSS should be able to produce forecasts by combining relevant information, analytical models, judgment, visualisations, and feedback. To that end, FSSs must (*i*) elaborate accurate, efficient, and automatic statistical forecasting methods, (*ii*) enable users to effectively incorporate their judgment, (*iii*) allow the users to track and interact with the whole forecasting procedure, and (*iv*) be easily customised based on the context of the company.

Indeed, nowadays, most off-the-self solutions, such as SAP, SAS, JDEdwards, and ForecastPro, offer a variety of both standard and advanced statistical forecasting methods (see Section 2.3), as well as data pre-processing (see Section 2.2) and performance evaluation algorithms (see Section 2.12). On the other hand, many of them still struggle to incorporate state-of-the-art methods that can further improve forecasting accuracy, such as automatic model selection algorithms and temporal aggregation (see also Section 2.10.2), thus limiting the options of the users (Petropoulos, 2015). Similarly, although many FSSs support judgmental forecasts (see Section 2.11.1) and judgmental adjustments of statistical forecasts (see Section 2.11.2), this is not done as suggested by the literature, i.e., in a guided way under a well-organised framework. As a result, the capabilities of the users are restrained and methods that could be used to mitigate biases, overshooting, anchoring, and unreasonable or insignificant changes that do not rationalise the time wasted, are largely ignored (Fildes & Goodwin, 2013; Fildes et al., 2006).

Other practical issues of FSSs are related with their engine and interfaces which are typically designed so that they are generic and capable to serve different companies and organisations of diverse needs (Kusters et al., 2006). From a developing and economic perspective, this is a reasonable choice. However, the lack of flexibility and customisability can lead to interfaces with needless options, models, tools, and features that may confuse inexperienced users and undermine their performance (Fildes et al., 2006). Thus, simple, yet exhaustive interfaces should be designed in the future to better serve the needs of each company and fit its particular requirements (Spiliotis, Raptis, & Assimakopoulos, 2015). Ideally, the interfaces should be adapted to the strengths and weaknesses of the user, providing useful feedback when possible (Goodwin, Fildes, Lawrence, & Nikolopoulos, 2007). Finally, web-based FSSs could replace windows-based ones that are locally installed and therefore of limited accessibility, availability, and compatibility (Asimakopoulos & Dix, 2013). Cloud computing and web-services could be exploited in that direction.

3.7.2. Cloud resource capacity forecasting[144]

One of the central promises in cloud computing is that of elasticity. Customers of cloud computing services can add compute resources in real-time to meet and satisfy increasing demand and, when demand for a cloud-hosted application goes down, it is possible for cloud computing customers to down-scale. The benefit of the latter is particularly economically interesting during the

[143] This subsection was written by Vassilios Assimakopoulos.

[144] This subsection was written by Tim Januschowski.

current pandemic. Popular recent cloud computing offerings take this elasticity concept one step further. They abstract away the computational resources completely from developers, so that developers can build serverless applications. In order for this to work, the cloud provider handles the addition and removal of compute resources "behind the scenes".

To keep the promise of elasticity, a cloud provider must address a number of forecasting problems at varying scales along the operational, tactical and strategic problem dimensions (Januschowski & Kolassa, 2019). As an example for a strategic forecasting problems: where should data centres be placed? In what region of a country and in what geographic region? As an example for tactical forecasting problems, these must take into account energy prices (see Section 3.4.2) and also, classic supply chain problems (Larson, Simchi-Levi, Kaminsky, & Simchi-Levi, 2001). After all, physical servers and data centres are what enables the cloud and these must be ordered and have a lead-time. The careful incorporation of life cycles of compute types is important (e.g., both the popularity of certain machine types and the duration of a hard disk). Analogous to the retail sector, cloud resource providers have tactical cold-start forecasting problems. For example, while GPU or TPU instances are still relatively recent but already well estabilished, the demand for quantum computing is still to be decided. In the class of operational forecasting problems, cloud provider can choose to address short-term resource forecasting problems for applications such as adding resources to applications predictively and make this available to customers (Barr, 2018). The forecasting of the customer's spend for cloud computing is another example. For serverless infrastructure, a number of servers is often maintained in a ready state (Gias & Casale, 2020) and the forecasting of the size of this 'warmpool' is another example. We note that cloud computing customers have forecasting problems that mirror the forecasting challenges of the cloud providers. Interestingly, forecasting itself has become a software service that cloud computing companies offer (Januschowski, Arpin, Salinas, Flunkert, Gasthaus & Lorenzo, 2018; Liberty et al., 2020; Poccia, 2019)

Many challenges in this application area are not unique to cloud computing. Cold start problems exist elsewhere for example. What potentially stands out in cloud computing forecasting problems may be the scale (e.g., there are a lot of physical servers available), the demands on the response time and granularity of a forecast and the degree of automation. Consider the operational forecasting problem of predictive scaling. Unlike in retail demand forecasting, no human operator will be able to control this and response times to forecasts are in seconds. It will be interesting to see whether approaches based on reinforcement learning (Dempster, Payne, Romahi, & Thompson, 2001; Gamble & Gao, 2018) can partially replace the need to have forecasting models (Januschowski, Gasthaus, Wang, Rangapuram & Callot, 2018).

3.7.3. *Judgmental forecasting in practice*[145]

Surveys of forecasting practice (De Baets, 2019) have shown that the use of pure judgmental forecasting by practitioners has become less common. About 40 years ago, Sparkes and McHugh (1984) found that company action was more likely to be influenced by judgmental forecasts than by any other type of forecast. In contrast, Fildes and Petropoulos (2015) found that only 15.6% of forecasts in the surveyed companies were made by judgment alone. The majority of forecasts (55.6%) were made using a combination of statistical and judgmental methods. In this section, we discuss forecasting using unaided judgment (pure judgmental forecasting; see also Section 2.11.1), judgmental adjustments (judgment in combination with statistical models; see also Section 2.11.2), and the role of judgment in forecasting support systems.

On the first theme, the survey results discussed above beg the question of whether pure judgmental forecasting is still relevant and reliable. Answers here depend on the type of information on which the judgmental forecasts are based (Harvey, 2007, see also Section 2.11.1). For instance, people have difficulty making cross-series forecasts, as they have difficulty learning the correlation between variables and using it to make their forecasts (Harvey, Bolger, & McClelland, 1994; Lim & O'Connor, 1996b, 1996c). Additionally, they appear to take account of the noise as well as the pattern when learning the relation between variables; hence, when later using one of the variables to forecast the other, they add noise to their forecasts (Gray, Barnes, & Wilkinson, 1965). Judgmental extrapolation from a single time series is subject to various effects. First, people are influenced by optimism. For example, they over-forecast time series labelled as 'profits' but under-forecast the same series labelled as 'losses' (Harvey & Reimers, 2013). Second, they add noise to their forecasts so that a sequence of forecasts looks similar to ('represents') the data series (Harvey, 1995). Third, they damp trends in the data (Eggleton, 1982; Harvey & Reimers, 2013; Lawrence & Makridakis, 1989). Fourth, forecasts from un-trended independent series do not lie on the series mean but between the last data point and the mean; this is what we would expect if people perceived a positive autocorrelation in the series (Reimers & Harvey, 2011). These last two effects can be explained in terms of the under-adjustment that characterises use of the anchor-and-adjust heuristic: forecasters anchor on the last data point and adjust towards the trend line or mean – but do so insufficiently. However, in practice, this under-adjustment may be appropriate because real linear trends do become damped and real series are more likely to contain a modest autocorrelation than be independent (Harvey, 2011). We should therefore be reluctant to characterise these last two effects as biases.

Given these inherent flaws in people's decision making, practitioners might be hesitant to base their predictions on judgment. However, the reality is that companies persist in incorporating judgment into their forecasting.

145 This subsection was written by Shari De Baets, M. Sinan Gönül, & Nigel Harvey.

Assertions that they are wrong to do so represent an over-simplified view of the reality in which businesses operate. Statistical models are generally not able to account for external events, events with low frequency, or a patchy and insufficient data history (Armstrong & Collopy, 1998; Goodwin, 2002; Hughes, 2001). Hence, a balance may be found in the combination of statistical models and judgment (see Section 2.11.2).

In this respect, judgmental adjustments to statistical model outputs are the most frequent form of judgmental forecasting in practice (Arvan et al., 2019; Eksoz et al., 2019; Lawrence et al., 2006; Petropoulos et al., 2016). Judgmental adjustments give practitioners a quick and convenient way to incorporate their insights, their experience and the additional information that they possess into a set of statistical baseline forecasts. Interestingly, Fildes et al. (2009) examined the judgmental adjustment applications in four large supply-chain companies and found evidence that the adjustments in a 'negative' direction improved the accuracy more than the adjustments in a 'positive' direction. This effect may be attributable to wishful thinking or optimism that may underlie positive adjustments. Adjustments that were 'larger' in magnitude were also more beneficial in terms of the final forecast accuracy than 'smaller' adjustments (Fildes et al., 2009). This may simply be because smaller adjustments are merely a sign of tweaking the numbers, but large adjustments are carried out when there is a highly valid reason to make them. These findings have been confirmed in other studies (see, for example, Franses & Legerstee, 2009b; Syntetos, Nikolopoulos, Boylan, Fildes, & Goodwin, 2009).

What are the main reasons behind judgmental adjustments? Önkal and Gönül (2005) conducted a series of interviews and a survey on forecasting practitioners (Gönül, Önkal, & Goodwin, 2009) to explore these. The main reasons given were (*i*) to incorporate the practitioners' intuition and experience about the predictions generated externally, (*ii*) to accommodate sporadic events and exceptional occasions, (*iii*) to integrate confidential/insider information that may have not been captured in the forecasts, (*iv*) to hold responsibility and to gain control of the forecasting process, (*v*) to incorporate the expectations and viewpoints of the practitioners, and (*vi*) to compensate for various judgmental biases that are believed to exist in the predictions. These studies also revealed that forecasting practitioners are very fond of judgmental adjustments and perceive them as a prominent way of 'completing' and 'owning' the predictions that are generated by others.

While the first three reasons represent the integration of an un-modelled component into the forecast, potentially improving accuracy, the other reasons tend to harm accuracy rather than improve it. In such cases, the forecast would be better off if left unadjusted. Önkal and Gönül (2005) and Gönül et al. (2009) report that the occasions when forecasters refrain from adjustments are (*i*) when the practitioners are adequately informed and knowledgeable about the forecasting method(s) that are used to generate the baseline forecasts, (*ii*) when there are accompanying explanations and convincing communications that provide the rationale behind forecast method selection, (*iii*) when baseline predictions are supplemented by additional supportive materials such as scenarios and alternative forecasts, (*iv*) when the forecasting source is believed to be trustworthy and reliable, and (*v*) when organisational policy or culture prohibits judgmental adjustments. In these circumstances, the baseline forecasts are more easily accepted by practitioners and their adjustments tend to be less frequent.

Ideally, a Forecast Support System (FSS; see Section 3.7.1) should be designed to ensure that it encourages adjustment or non-adjustment whichever is appropriate (Fildes et al., 2006). But how can this be achieved? The perceived quality and accessibility of a FSS can be influenced by its design. More on this can be found in the literature on the Technology Acceptance Model (Davis, Bagozzi, & Warshaw, 1989) and decision making (for instance, by means of framing, visual presentation or nudging; e.g., Gigerenzer, 1996; Kahneman & Tversky, 1996; Payne, 1982; Thaler & Sunstein, 2009). A number of studies have investigated the design aspects of FSS, with varying success. One of the more straightforward approaches is to change the look and feel of the FSS as well as its presentation style. Harvey and Bolger (1996) found that trends were more easily discernible when the data was displayed in graphical rather than tabular format. Additionally, simple variations in presentation such as line graphs versus point graphs can alter accuracy (Theocharis, Smith, & Harvey, 2018). The functionalities of the FSS can also be modified (see Section 2.11.2). Goodwin (2000b) investigated three ways of improving judgmental adjustment via changes in the FSS: a 'no adjustment' default, requesting forecasters specify the size of an adjustment rather than give a revised forecast, and requiring a mandatory explanation for the adjustment. Only the default option and the explanation feature were successful in increasing the acceptance of the statistical forecast and so improving forecast accuracy.

Goodwin et al. (2011) reported an experiment that investigated the effects of (*i*) 'guidance' in the form of providing information about when to make adjustments and (*ii*) 'restriction' of what the forecaster could do (e.g., prohibiting small adjustments). They found that neither restrictiveness nor guidance was successful in improving accuracy, and both were met with resistance by the forecasters. While these studies focused on voluntary integration, Goodwin (2000a, 2002) examined the effectiveness of various methods of mechanical integration and concluded that automatic correction for judgmental biases by the FSS was more effective than combining judgmental and statistical inputs automatically with equal or varying weights. Another approach to mechanical integration was investigated by Baecke et al. (2017). They compared ordinary judgmental adjustment with what they termed "integrative judgment". This takes the judgmental information into account as a predictive variable in the forecasting model and generates a new forecast. This approach improved accuracy. It also had the advantage that forecasters still had their input into the forecasting process and so the resistance found by Goodwin et al. (2011) should not occur. Finally, it is worth emphasising that an effective FSS should not only improve forecast accuracy but should also be easy to use, understandable, and acceptable (Fildes et al., 2006, see also Section 2.11.6 and Section 3.7.1).

3.7.4. Trust in forecasts[146]

Regardless of how much effort is poured into training forecasters and developing elaborate forecast support systems, decision-makers will either modify or discard the predictions if they do not trust them (see also Section 2.11.2, Section 2.11.6, Section 3.7.1, and Section 3.7.3). Hence, trust is essential for forecasts to be actually used in making decisions (Alvarado-Valencia & Barrero, 2014; Önkal et al., 2019).

Given that trust appears to be the most important attribute that promotes a forecast, what does it mean to practitioners? Past work suggests that trusting a forecast is often equated with trusting the forecaster, their expertise and skills so that predictions could be used without adjustment to make decisions (Önkal et al., 2019). It is argued that trust entails relying on credible forecasters that make the best use of available information while using correctly applied methods and realistic assumptions (Gönül et al., 2009) with no hidden agendas (Gönül, Önkal, & Goodwin, 2012). Research suggests that trust is not only about trusting forecaster's competence; users also need to be convinced that no manipulations are made for personal gains and/or to mislead decisions (Twyman, Harvey, & Harries, 2008).

Surveys with practitioners show that key determinants of trust revolve around (*i*) forecast support features and tools (e.g., graphical illustrations, rationale for forecasts), (*ii*) forecaster competence/credibility, (*iii*) forecast combinations (from multiple forecasters/methods), and (*iv*) forecast user's knowledge of forecasting methods (Önkal et al., 2019).

What can be done to enhance trust? If trust translates into accepting guidance for the future while acknowledging and tolerating potential forecast errors, then both the providers and users of forecasts need to work as partners towards shared goals and expectations. Important pathways to accomplish this include (*i*) honest communication of forecaster's track record and relevant accuracy targets (Önkal et al., 2019), (*ii*) knowledge sharing (Özer et al., 2011; Renzl, 2008) and transparency of forecasting methods, assumptions and data (Önkal et al., 2019), (*iii*) communicating forecasts in the correct tone and jargon-free language to appeal to the user audience (Taylor & Thomas, 1982), (*iv*) users to be supported with forecasting training (Merrick, Hardin, & Walker, 2006), (*v*) providing explanations/rationale behind forecasts (Gönül, Önkal, & Lawrence, 2006; Önkal, Gönül, & Lawrence, 2008), (*vi*) presenting alternative forecasts under different scenarios (see Section 2.11.5), and (*vii*) giving combined forecasts as benchmarks (Önkal et al., 2019).

Trust must be earned and deserved (Maister et al., 2012) and is based on building a relationship that benefits both the providers and users of forecasts. Take-aways for those who make forecasts and those who use them converge around clarity of communication as well as perceptions of competence and integrity. Key challenges for forecasters are to successfully engage with users throughout the forecasting process (rather than relying on a forecast statement at the end) and to convince them of their objectivity and expertise. In parallel, forecast users face challenges in openly communicating their expectations from forecasts (Gönül et al., 2009), as well as their needs for explanations and other informational addendum to gauge the uncertainties surrounding the forecasts. Organisational challenges include investing in forecast management and designing resilient systems for collaborative forecasting.

3.7.5. Communicating forecast uncertainty[147]

Communicating forecast uncertainty is a critical issue in forecasting practice. Effective communication allows forecasters to influence end-users to respond appropriately to forecasted uncertainties. Some frameworks for effective communication have been proposed by decomposing the communication process into its elements: the communicator, object of uncertainty, expression format, audience, and its effect (National Research Council, 2006; van der Bles, van der Linden, Freeman, Mitchell, Galvao, Zaval, & Spiegelhalter, 2019).

Forecasters have long studied part of this problem focusing mostly in the manner by which we express forecast uncertainties. Gneiting and Katzfuss (2014) provides a review of recent probabilistic forecasting methods (see also Section 2.12.4 and Section 2.12.5). Forecasting practice however revealed that numeracy skills and cognitive load can often inhibit end-users from correctly interpreting these uncertainties (Joslyn & Nichols, 2009; Raftery, 2016). Attempts to improve understanding through the use of less technical vocabulary also creates new challenges. Research in psychology show that wording and verbal representation play important roles in disseminating uncertainty (Joslyn, Nadav-Greenberg, Taing, & Nichols, 2009). Generally forecasters are found to be consistent in their use of terminology, but forecast end-users often have inconsistent interpretation of these terms even those commonly used (Budescu & Wallsten, 1985; Clark, 1990; Ülkümen, Fox, & Malle, 2016). Pretesting verbal expressions and avoiding commonly misinterpreted terms are some easy ways to significantly reduce biases and improve comprehension.

Visualisations can also be powerful in communicating uncertainty. Johnson and Slovic (1995) and Spiegelhalter, Pearson, and Short (2011) propose several suggestions for effective communication (e.g., multiple-format use, avoiding framing bias, and acknowledging limitations), but also recognise the limited amount of existing empirical evidence. Some domain-specific studies do exist. For example, Riveiro, Helldin, Falkman, and Lebram (2014) showed uncertainty visualisation helped forecast comprehension in a homeland security context.

With respect to the forecaster and her audience, issues such as competence, trust, respect, and optimism have been recently examined as a means to improve uncertainty communication. Fiske and Dupree (2014) discusses how forecast recipients often infer apparent intent and competence from the uncertainty provided and use these to judge trust and respect (see also Sections 2.11.6 and 3.7.4 for discussion on trust and forecasting). This

[146] This subsection was written by Dilek Önkal.

[147] This subsection was written by Victor Richmond R. Jose.

suggests that the amount of uncertainty information provided should be audience dependent (Han, Klein, Lehman, Massett, Lee, & Freedman, 2009; Politi, Han, & Col, 2007). Raftery (2016) acknowledges this by using strategies depending on the audience type (e.g., low-stakes user, risk avoider, etc.). Fischhoff and Davis (2014) suggests a similar approach by examining how people are likely to use the information (e.g., finding a signal, generating new options, etc.)

When dealing with the public, experts assert that communicating uncertainty helps users understand forecasts better and avoid a false sense of certainty (Morss, Demuth, & Lazo, 2008). Research however shows that hesitation to include forecast uncertainty exists among experts because it provides an opening for criticism and the possibility of misinterpration by the public (Fischhoff, 2012). This is more challenging when the public has prior beliefs on a topic or trust has not been established. Uncertainty can be used by individuals to reinforce a motivated-reasoning bias that allows them to "see what they want to see" (Dieckmann, Gregory, Peters, & Hartman, 2017). Recent work however suggests that increasing transparency for uncertainty does not necessarily affect trust in some settings. van der Bles, van der Linden, Freeman, and Spiegelhalter (2020) recently showed in a series of experiments that people recognise greater uncertainty with more information but expressed only a small decrease in trust in the report and trustworthiness of the source.

3.8. Other applications

3.8.1. Tourism demand forecasting[148]

As seen throughout 2020, (leisure) tourism demand is very sensitive to external shocks such as natural and human-made disasters, making tourism products and services extremely perishable (Frechtling, 2001). As the majority of business decisions in the tourism industry require reliable demand forecasts (Song, Witt, & Li, 2008), improving their accuracy has continuously been on the agenda of tourism researchers and practitioners alike. This continuous interest has resulted in two tourism demand forecasting competitions to date (Athanasopoulos et al., 2011; Song & Li, 2021), the current one with a particular focus on tourism demand forecasting during the COVID-19 pandemic (for forecasting competitions, see Section 2.12.7). Depending on data availability, as well as on geographical aggregation level, tourism demand is typically measured in terms of arrivals, bed-nights, visitors, export receipts, import expenditures, etc.

Since there are no specific tourism demand forecast models, standard univariate and multivariate statistical models, including common aggregation and combination techniques, etc., have been used in quantitative tourism demand forecasting (see, for example, Jiao & Chen, 2019; Song, Qiu, & Park, 2019, for recent reviews). Machine learning and other artificial intelligence methods, as well as hybrids of statistical and machine learning models, have recently been employed more frequently.

Traditionally, typical micro-economic demand drivers (own price, competitors' prices, and income) and some more tourism-specific demand drivers (source-market population, marketing expenditures, consumer tastes, habit persistence, and dummy variables capturing one-off events or qualitative characteristics) have been employed as predictors in tourism demand forecasting (Song et al., 2008). One caveat of some of these economic demand drivers is their publication lag and their low frequency, for instance, when real GDP (per capita) is employed as a proxy for travellers' income.

The use of leading indicators, such as industrial production as a leading indicator for real GDP (see also Section 3.3.2), has been proposed for short-term tourism demand forecasting and nowcasting (Chatziantoniou, Degiannakis, Eeckels, & Filis, 2016). During the past couple of years, web-based leading indicators have also been employed in tourism demand forecasting and have, in general, shown improvement in terms of forecast accuracy. However, this has not happened in each and every case, thereby confirming the traded wisdom that there is no single best tourism demand forecasting approach (Li, Song, & Witt, 2005). Examples of those web-based leading indicators include Google Trends indices (Bangwayo-Skeete & Skeete, 2015), Google Analytics indicators (Gunter & Önder, 2016), as well as Facebook 'likes' (Gunter, Önder, & Gindl, 2019).

The reason why these expressions of interaction of users with the Internet have proven worthwhile as predictors in a large number of cases is that it is sensible to assume potential travellers gather information about their destination of interest prior to the actual trip, with the Internet being characterised by comparably low search costs, ergo allowing potential travellers to forage information (Pirolli & Card, 1999) with only little effort (Zipf, 2016). A forecaster should include this information in their own set of relevant information at the forecast origin Lütkepohl (2005), if taking it into account results in an improved forecast accuracy, with web-based leading indicators thus effectively Granger-causing (Granger, 1969) actual tourism demand (see Section 2.5.1).

Naturally, tourism demand forecasting is closely related to aviation forecasting (see Section 3.8.2), as well as traffic flow forecasting (see Section 3.8.3). A sub-discipline of tourism demand forecasting can be found with hotel room demand forecasting. The aforementioned perishability of tourism products and services is particularly evident for hotels as a hotel room not sold is lost revenue that cannot be regenerated. Accurate hotel room demand forecasts are crucial for successful hotel revenue management (Pereira, 2016) and are relevant for planning purposes such as adequate staffing during MICE (i.e., Meetings, Incentives, Conventions, and Exhibitions/Events) times, scheduling of renovation periods during low seasons, or balancing out overbookings and "no shows" given constrained hotel room supply (Ivanov & Zhechev, 2012).

Particularly since the onset of the COVID-19 pandemic in 2020, which has been characterised by global travel restrictions and tourism businesses being locked down to varying extents, scenario forecasting and other forms of hybrid and judgmental forecasting played an important

[148] This subsection was written by Ulrich Gunter.

role (Zhang, Song, Wen, & Liu, 2021, see Section 2.11.5), thereby highlighting an important limitation of quantitative tourism demand forecasting as currently practised. Based on the rapid development of information technology and artificial intelligence, Li and Jiao (2020), however, envisage a "super-smart tourism forecasting system" (Li & Jiao, 2020, p. 264) for the upcoming 75 years of tourism demand forecasting. According to these authors, this system will be able to automatically produce forecasts at the micro level (i.e., for the individual traveller and tourism business) in real time while drawing on a multitude of data sources and integrating multiple (self-developing) forecast models.

3.8.2. *Forecasting for aviation*[149]

Airports and airlines have long invested in forecasting arrivals and departures of aircrafts. These forecasts are important in measuring airspace and airport congestions, designing flight schedules, and planning for the assignment of stands and gates (Barnhart & Cohn, 2004). Various techniques have been applied to forecast aircrafts' arrivals and departures. For instance, Rebollo and Balakrishnan (2014) apply random forests to predict air traffic delays of the National Airspace System using both temporal and network delay states as covariates. Manna, Biswas, Kundu, Rakshit, Gupta, and Barman (2017) develop a statistical model based on a gradient boosting decision tree to predict arrival and departure delays, using the data taken from the United States Department of Transportation (Bureau of Transportation Statistics, 2020). Rodríguez-Sanz, Comendador, Valdés, Pérez-Castán, Montes, and Serrano (2019) develop a Bayesian Network model to predict flight arrivals and delays using the radar data, aircraft historical performance and local environmental data. There are also a few studies that have focused on generating probabilistic forecasts of arrivals and departures, moving beyond point estimates. For example, Tu, Ball, and Jank (2008) develop a predictive system for estimating flight departure delay distributions using flight data from Denver International Airport. The system employs the smoothing spline method to model seasonal trends and daily propagation patterns. It also uses mixture distributions to estimate the residual errors for predicting the entire distribution.

In the airline industry, accurate forecasts on demand and booking cancellations are crucial to revenue management, a concept that was mainly inspired by the airline and hotel industries (Lee, 1990; McGill & Van Ryzin, 1999, see also Section 3.8.1 for a discussion on hotel occupancy forecasting). The proposals of forecasting models for flight demand can be traced back to Beckmann and Bobkoski (1958), where these authors demonstrate that Poisson and Gamma models can be applied to fit airline data. Then, the use of similar flights' short-term booking information in forecasting potential future bookings has been discussed by airline practitioners such as Adams and Michael (1987) at Quantas as well as Smith, Leimkuhler, and Darrow (1992) at American Airlines. Regressions models (see Section 2.3.2) and time series models such as exponential smoothing (see Section 2.3.1) and ARIMA (see Section 2.3.4) have been discussed in Botimer (1997), Sa (1987), and Wickham (1995). There are also studies focusing on disaggregate airline demand forecasting. For example, Martinez and Sanchez (1970) apply empirical probability distributions to predict bookings and cancellations of individual passengers travelling with Iberia Airlines. Carson, Cenesizoglu, and Parker (2011) show that aggregating the forecasts of individual airports using airport-specific data could provide better forecasts at a national level. More recently, machine learning methods have also been introduced to generate forecasts for airlines. This can be seen in Weatherford, Gentry, and Wilamowski (2003) where they apply neural networks to forecast the time series of the number of reservations. Moreover, Hopman, Koole, and van der Mei (2021) show that an extreme gradient boosting model which forecasts itinerary-based bookings using ticket price, social media posts and airline reviews outperforms traditional time series forecasts.

Forecasting passenger arrivals and delays in the airports have received also some attention in the literature, particularly in the past decade. Wei and Hansen (2006) build an aggregate demand model for air passenger traffic in a hub-and-spoke network. The model is a log-linear regression that uses airline service variables such as aircraft size and flight distance as predictors. Barnhart, Fearing, and Vaze (2014) develop a multinomial logit regression model, designed to predict delays of US domestic passengers. Their study also uses data from the US Department of Transportation (Bureau of Transportation Statistics, 2020). Guo, Grushka-Cockayne, and De Reyck (2020) recently develop a predictive system that generates distributional forecasts of connection times for transfer passengers at an airport, as well as passenger flows at the immigration and security areas. Their approach is based on the application of regression trees combined with copula-based simulations. This predictive system has been implemented at Heathrow airport since 2017.

With an increasing amount of available data that is associated with activities in the aviation industry, predictive analyses and forecasting methods face new challenges as well as opportunities, especially in regard to updating forecasts in real time. The predictive system developed by Guo et al. (2020) is able to generate accurate forecasts using real-time flight and passenger information on a rolling basis. The parameters of their model, however, do not update over time. Therefore, a key challenge in this area is for future studies to identify an efficient way to dynamically update model parameters in real time.

3.8.3. *Traffic flow forecasting*[150]

Traffic flow forecasting is an important task for traffic management bodies to reduce traffic congestion, perform planning and allocation tasks, as well as for travelling individuals to plan their trips. Traffic flow is complex spatial and time-series data exhibiting multiple seasonalities and affected by spatial exogenous influences such

149 This subsection was written by Xiaojia Guo.

150 This subsection was written by Alexander Dokumentov.

as social and economic activities and events, various government regulations, planned road works, weather, traffic accidents, etc. (Polson & Sokolov, 2017).

Methods to solve traffic flow forecasting problems vaguely fall into three categories. The first uses parametric statistical methods such as ARIMA, seasonal ARIMA, space–time ARIMA, Kalman filters, etc. (see, for example, Kamarianakis & Prastacos, 2005; Vlahogianni, Golias, & Karlaftis, 2004; Vlahogianni, Karlaftis, & Golias, 2014; Whittaker, Garside, & Lindveld, 1997). The second set of approaches uses purely of neural networks (Mena-Oreja & Gozalvez, 2020). The third group of methods uses various machine learning, statistical non-parametric techniques or mixture of them (see, for example, Hong, 2011; Zhang, Qi, Henrickson, Tang, & Wang, 2017; Zhang, Zou, Tang, Ash, & Wang, 2016, but also Section 2.7.8 and Section 2.7.10 for an overview of NN and ML methods).

Although neural networks are probably the most promising technique for traffic flow forecasting (see, for example, Do, Vu, Vo, Liu, & Phung, 2019; Polson & Sokolov, 2017), statistical techniques, such as Seasonal-Trend decomposition based on Regression (STR, see Section 2.2.2), can outperform when little data is available or they can be used for imputation, de-noising, and other pre-processing before feeding data into neural networks which often become less powerful when working with missing or very noisy data.

Traffic flow forecasting is illustrated below using vehicle flow rate data from road camera A1.GT.24538 on A1 highway in Luxembourg (des Mobilités, 2020) from 2019-11-19 06:44:00 UTC to 2019-12-23 06:44:00 UTC. Most of the data points are separated by 5 min intervals. Discarding points which do not follow this schedule leads to a data set where all data points are separated by 5 min intervals, although values at some points are missing. The data is split into training and test sets by setting aside last 7 days of data. As Hou, Edara, and Sun (2014) and Polson and Sokolov (2017) suggest, spatial factors are less important for long term traffic flow forecasting, and therefore they are not taken into account and only temporal data is used. Application of STR (Dokumentov, 2017) as a forecasting technique to the log transformed data leads to a forecast with Mean Squared Error 102.4, Mean Absolute Error 62.8, and Mean Absolute Percentage Error (MAPE) 14.3% over the test set, outperforming Double-Seasonal Holt-Winters by 44% in terms of MAPE. The decomposition and the forecast obtained by STR are shown on Fig. 15 and the magnified forecast and the forecasting errors are on Fig. 16.

3.8.4. Call arrival forecasting[151]

Forecasting of inbound call arrivals for call centres supports a number of key decisions primarily around staffing (Aksin, Armony, & Mehrotra, 2007). This typically involves matching staffing level requirements to service demand as summarised in Fig. 17. To achieve service level objectives, an understanding of the call load is required in terms of the call arrivals (Gans, Koole, & Mandelbaum, 2003). As such, forecasting of future call volume or call arrival rates is an important part of call centre management.

There are several properties to call arrival data. Depending on the level of aggregation and the frequency with which data is collected, e.g., hourly, call arrival data may exhibit intraday (within-day), intraweek, and intrayear multiple seasonal patterns (Avramidis, Deslauriers, & L'Ecuyer, 2004; Brown et al., 2005a, and Section 2.3.5). In addition, arrival data may also exhibit interday and intraday dependencies, with different time periods within the same day, or across days within the same week, showing strong levels of autocorrelation (Brown et al., 2005a; Shen & Huang, 2005; Tanir & Booth, 1999). Call arrivals may also be heteroscedastic with variance at least proportional to arrival counts (Taylor, 2008), and overdispersed under a Poisson assumption having variance per time period typically much larger than its expected value (Avramidis et al., 2004; Jongbloed & Koole, 2001; Steckley, Henderson, & Mehrotra, 2005). These properties have implications for various approaches to modelling and forecasting call arrivals.

The first family of methods are time series methods requiring no distributional assumptions. Early studies employed auto regressive moving average (ARMA; see Section 2.3.4) models (Andrews & Cunningham, 1995; Antipov & Meade, 2002; Tandberg, Easom, & Qualls, 1995; Xu, 1999), exponential smoothing (Bianchi, Jarrett, & Hanumara, 1993, 1998, see Section 2.3.1), fast Fourier transforms (Lewis, Herbert, & Bell, 2003), and regression (Tych, Pedregal, Young, & Davies, 2002, see Section 2.3.2). The first methods capable of capturing multiple seasonality were evaluated by Taylor (2008) and included double seasonal exponential smoothing (Taylor, 2003b) and multiplicative double seasonal ARMA (SARMA). Since then several advanced time series methods have been developed and evaluated (De Livera et al., 2011; Taylor, 2010; Taylor & Snyder, 2012), including artificial neural networks (Li, Huang, & Gong, 2011; Millán-Ruiz & Hidalgo, 2013; Pacheco, Millán-Ruiz, & Vélez, 2009) and models for density forecasting (Taylor, 2012).

Another family of models relies on the assumption of a time-inhomogeneous Poisson process adopting fixed (Brown et al., 2005a; Jongbloed & Koole, 2001; Shen & Huang, 2008a; Taylor, 2012) and mixed modelling (Aldor-Noiman, Feigin, & Mandelbaum, 2009; Avramidis et al., 2004; Ibrahim & L'Ecuyer, 2013) approaches to account for the overdispersed nature of the data, and in some cases, interday and intraday dependence.

The works by Soyer and Tarimcilar (2008) and Weinberg, Brown, and Stroud (2007) model call volumes from a Bayesian point of view. Other Bayesian inspired approaches have been adopted mainly for estimating various model parameters, but also allowing for intraday updates of forecasts (Aktekin & Soyer, 2011; Landon, Ruggeri, Soyer, & Tarimcilar, 2010).

A further class of approach addresses the dimensionality challenge related to high frequency call data using Singular Value Decomposition (SVD). Shen and Huang

[151] This subsection was written by Devon K. Barrow.

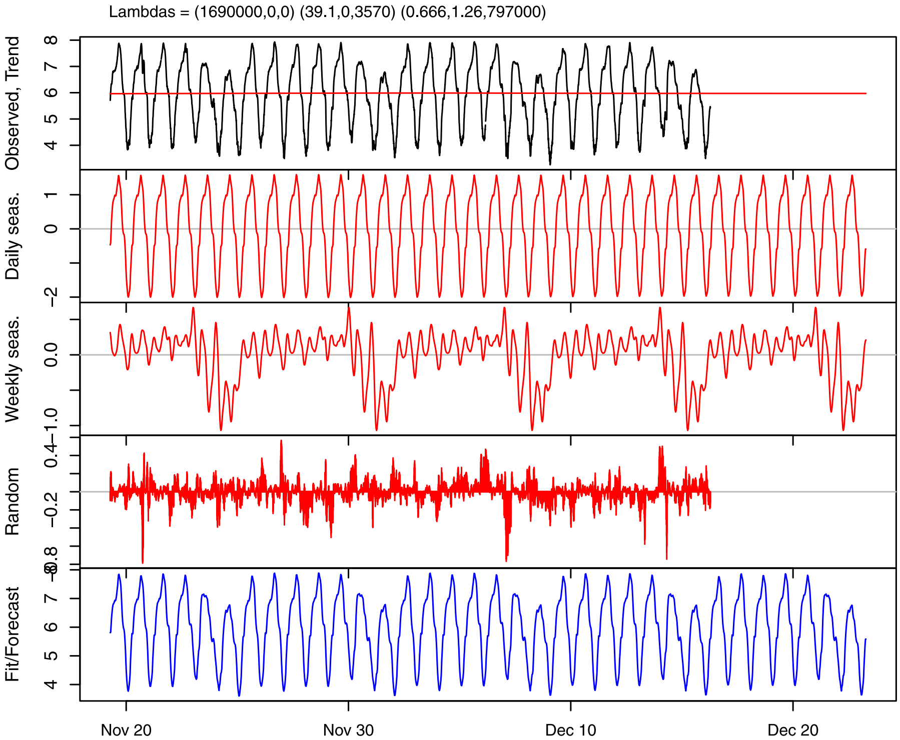

Fig. 15. STR decomposition of the log transformed training data and the forecasts for the traffic flow data.

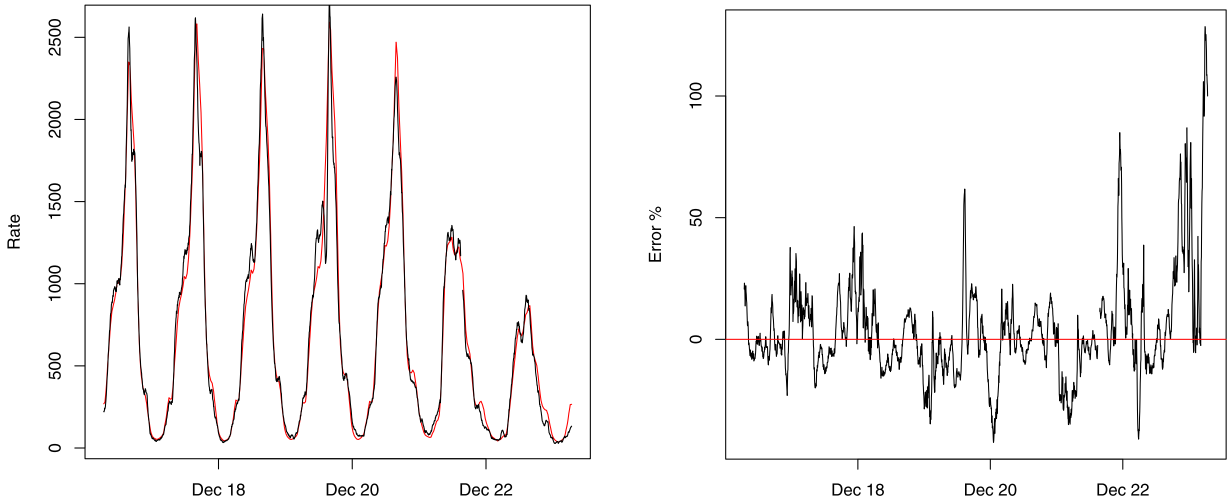

Fig. 16. Left: forecast (red) and the test data (black); Right: the prediction error over time for the traffic flow data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

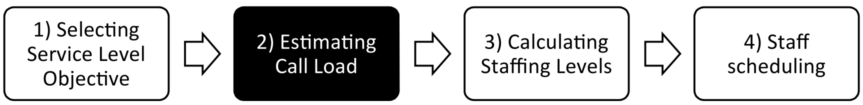

Fig. 17. The staffing decision process in call centres.

(2005) and Shen and Huang (2008a) use the same technique to achieve dimensionality reduction of arrival data, and to create a forecasting model that provides both interday forecasts of call volume, and an intraday updating mechanism. Several further studies have extended the basic SVD approach to realise further modelling innovations, for example, to forecast call arrival rate profiles and generate smooth arrival rate curves (Shen, 2009; Shen &

Huang, 2008b; Shen, Huang, & Lee, 2007). A more comprehensive coverage of different forecasting approaches for call arrival rate and volume can be found in a recent review paper by Ibrahim, Ye, L'Ecuyer, and Shen (2016).

3.8.5. Elections forecasting[152]

With the exception of weather forecasts, there are few forecasts which have as much public exposure as election forecasts. They are frequently published by mass media, with their number and disclosure reaching a frenzy as the Election Day approaches. This explains the significant amount of methods, approaches and procedures proposed and the paramount role these forecasts play in shaping people's confidence in (soft/social) methods of forecasting.

The problem escalates because, regardless whether the goal of the election forecast is an attempt to ascertain the winner in two-choice elections (e.g., a referendum or a Presidential election) or to reach estimates within the margins of error in Parliamentary systems, the knowledge of the forecasts influences electors' choices (Pavía, Gil-Carceller, Rubio-Mataix, Coll, Alvarez-Jareño, Aybar, & Carrasco-Arroyo, 2019). Election forecasts not only affect voters but also political parties, campaign organizations and (international) investors, who are also watchful of their evolution.

Scientific approaches to election forecasting include polls, information (stock) markets and statistical models. They can also be sorted by when they are performed; and new methods, such as social media surveillance (see also Section 2.9.3), are also emerging (Ceron, Curini, & Iacus, 2016; Huberty, 2015). Probabilistic (representative) polls are the most commonly used instrument to gauge public opinions. The progressive higher impact of non-sampling errors (coverage issues, non-response bias, measurement error: Biemer, 2010) is, however, severely testing this approach. Despite this, as Kennedy, Wojcik, and Lazer (2017) show in a recent study covering 86 countries and more than 500 elections, polls are still powerful and robust predictors of election outcomes after adjustments (see also Jennings, Lewis-Beck, & Wlezien, 2020). The increasing need of post-sampling adjustments of probabilistic samples has led to a resurgence of interest in non-probabilistic polls (Elliott & Valliant, 2017; Pavía & Larraz, 2012; Wang, Rothschild, Goel, & Gelman, 2015), abandoned in favour of probabilistic sampling in 1936, when Gallup forecasted Roosevelt's triumph over Landon using a small representative sample despite Literacy Digest failing to do so with a sample of near 2.5 million responses (Squire, 1988).

A person knows far more than just her/his voting intention (Rothschild, 2009) and when s/he makes a bet, the rationality of her/his prediction is reinforced because s/he wants to win. Expectation polls try to exploit the first issue (Graefe, 2014), while prediction markets, as efficient aggregators of information, exploit both these issues to yield election forecasts (see also Sections 2.6.4 and 2.11.4). Several studies have proven the performance of these approaches (Berg, Nelson, & Rietz, 2008; Erikson & Wlezien, 2012; Williams & Reade, 2016; Wolfers & Zitzewitz, 2004), even studying their links with opinion polls (Brown, Reade, & Vaughan Williams, 2019). Practice has also developed econometric models (Fair, 1978) that exploit structural information available months before the election (e.g., the evolution of the economy or the incumbent popularity). Lewis-Beck has had great success in publishing dozens of papers using this approach (see, e.g., Lewis-Beck, 2005).

Special mention also goes to Election-Day forecasting strategies, which have been systematically commissioned since the 1950s (Mitofsky, 1991). Exit (and entrance) polls (Klofstad & Bishin, 2012; Pavía, 2010), quick-counts (Pavía-Miralles & Larraz-Iribas, 2008), and statistical models (Bernardo, 1984; Moshman, 1964; Pavía-Miralles, 2005) have been used to anticipate outcomes on Election Day. Some of these strategies (mainly random quick-counts) can be also employed as auditing tools to disclose manipulation and fraud in weak democracies (Scheuren & Alvey, 2008).

3.8.6. Sports forecasting[153]

Forecasting is inherent to sport. Strategies employed by participants in sporting contests rely on forecasts, and the decision by promoters to promote, and consumers to attend such events are conditioned on forecasts: predictions of how interesting the event will be. First in this section, we look at forecast competitions in sport, and following this we consider the role forecasts play in sporting outcomes.

Forecast competitions are common; see Section 2.12.7. Sport provides a range of forecast competitions, perhaps most notably the competition between bookmakers and their customers – betting. A bet is a contingent contract, a contract whose payout is conditional on specified future events occurring. Bets occur fundamentally because two agents disagree about the likelihood of that event occurring, and hence it is a forecast.

Bookmakers have been extensively analysed as forecasters; Forrest, Goddard, and Simmons (2005) evaluated biases in the forecasts implied by bookmaker odds over a period where the betting industry became more competitive, and found that relative to expert forecasts, bookmaker forecasts improved.

With the internet age, prediction markets have emerged, financial exchanges where willing participants can buy and sell contingent contracts. In theory, such decentralised market structures ought to provide the most efficient prices and hence efficient forecasts (Nordhaus, 1987). A range of papers have tested this in the sporting context (Angelini & De Angelis, 2019; Croxson & Reade, 2014; Gil & Levitt, 2007), with conclusions tending towards a lack of efficiency.

Judgmental forecasts by experts are commonplace too (see also Section 2.11); traditionally in newspapers, but more recently on television and online. Reade, Singleton, and Brown (2020) evaluate forecasts of scorelines from two such experts against bookmaker prices, a statistical

152 This subsection was written by Jose M. Pavía.

153 This subsection was written by J. James Reade.

model, and the forecasts from users of an online forecasting competition. Singleton, Reade, and Brown (2019) find that when forecasters in the same competition revise their forecasts, their forecast performance worsens. This forecasting competition is also analysed by Butler, Butler, and Eakins (2020) and Reade et al. (2020).

Sport is a spectacle, and its commercial success is conditioned on this fact. Hundreds of millions of people globally watch events like the Olympics and the FIFA World Cup – but such interest is conditioned on anticipation, a forecast that something interesting will happen. A superstar is going to be performing, the match will be a close encounter, or it will matter a lot for a bigger outcome (the championship, say). These are the central tenets of sport economics back to Neale (1964) and Rottenberg (1956), most fundamentally the 'uncertainty of outcome hypothesis'. A multitude of sport attendance prediction studies investigate this (see, for example, Coates & Humphreys, 2010; Forrest & Simmons, 2006; Hart, Hutton, & Sharot, 1975; Sacheti, Gregory-Smith, & Paton, 2014; van Ours, 2021), and Van Reeth (2019) considers this for forecasting TV audiences for the Tour de France.

Cities and countries bid to host large events like the World Cup based on forecasts regarding the impact of hosting such events. Forecasts that are often inflated for political reasons (Baade & Matheson, 2016). Equally, franchise-based sports like many North American sports attract forecasts regarding the impact of a team locating in a city, usually resulting in public subsidies for the construction of venues for teams to play at Coates and Humphreys (1999). Governments invest in sporting development, primarily to achieve better performances at global events, most notably the Olympics (Bernard & Busse, 2004).

Many sporting events themselves rely on forecasts to function; high jumpers predict what height they will be able to jump over, and free diving contestants must state the depth they will dive to. Less formally, teams will set themselves goals: to win matches, to win competitions, to avoid the 'wooden spoon'. Here, forecast outcomes are influenced by the teams, and competitors, taking part in competitions and, as such, are perhaps less commonly thought of as genuine forecasts. Important works predicting outcomes range from Dixon and Coles (1997) in soccer, to Kovalchik and Reid (2019) for tennis, while the increasing abundance of data means that machine learning and deep learning methods are beginning to dominate the landscape. See, for example, Hubáček, Šourek, and Železný (2019) and Maymin (2019) for basketball, and Mulholland and Jensen (2019) for NFL.

3.8.7. *Forecasting for megaprojects*[154]

Megaprojects are significant activities characterised by a multi-organisation structure, which produces highly visible infrastructure or asset with very crucial social impacts (Aaltonen, 2011). Megaprojects are complex, require huge capital investment, several stakeholders are identified and, usually a vast number of communities and the public are the receivers of the project's benefits. There is a need megaprojects especially those that deliver social and economic goods and create economic growth (Flyvbjerg, Bruzelius, & Rothengatter, 2003). Typical features of megaprojects include some or all the following: (*i*) delivering a substantial piece of physical infrastructure with a life expectancy that spans across decades, (*ii*) main contractor or group of contractors are privately owned and financed, and (*iii*) the contractor could retain an ownership stake in the project and the client is often a government or public sector organisation (Sanderson, 2012).

However, megaprojects are heavily laced with extreme human and technical complexities making their delivery and implementation difficult and often unsuccessful (Merrow, McDonnell, & Arguden, 1988; The R.F.E. Working Group Report, 2015). This is largely due to the challenge of managing megaprojects including extreme complexity, increased risk, tight budget and deadlines, lofty ideals (Fiori & Kovaka, 2005). Due to the possibility and consequences of megaproject failure (Mišić & Radujković, 2015), forecasting the outcomes of megaprojects is becoming of growing importance. In particular, it is crucial to identify and assess the risks and uncertainties as well as other factors that contribute to disappointing outcomes of megaprojects in order to mitigate them (Flyvbjerg et al., 2003; Miller & Lessard, 2007).

Literature review in forecasting in megaprojects is scarce. However, there are a few themes that have emerged in the extant literature as characteristics of megaprojects that should be skilfully managed to provide a guideline for the successful planning and construction of megaprojects (Fiori & Kovaka, 2005; Flyvbjerg, 2007; Sanderson, 2012). Turner and Zolin (2012) even claim that we cannot even properly define what success is. They argue that we need to reliable scales in order to predict multiple perspectives by multiple stakeholders over multiple time frames — so definitely a very difficult long term problem. This could be done via a set of leading performance indicators that will enable managers of Megaprojects to forecast during project execution how various stakeholders will perceive success months or even years into the operation. At the very early stages of a project's lifecycle, a number of decisions must been taken and are of a great importance for the performance and successful deliverables/outcomes. Flyvbjerg (2007) stress the importance of the front-end considerations particularly for Megaprojects Failure to account for unforeseen events frequently lead to cost overruns.

Litsiou et al. (2019) suggest that forecasting the success of megaprojects is particularly a challenging and critical task due to the characteristics of such projects. Megaproject stakeholders typically implement impact assessments and/or cost benefit Analysis tools (Litsiou et al., 2019). As Makridakis, Hogarth, and Gaba (2010) suggested, judgmental forecasting is suitable where quantitative data is limited, and the level of uncertainty is very high; elements that we find in megaprojects. By comparing the performance of three judgmental methods, unaided judgment, semi-structured analogies (sSA), and interaction groups (IG), used by a group of 69 semi-experts, Litsiou et al. (2019) found that, the use of sSA outperforms unaided judgment in forecasting performance (see also Section 2.11.4). The difference is amplified further when pooling of analogies through IG is introduced.

[154] This subsection was written by Konstantia Litsiou.

3.8.8. Competing products[155]

Competition among products or technologies affects prediction due to local systematic deviations and saturating effects related to policies, and evolving interactions. The corresponding sales time series must be jointly modelled including the time varying reciprocal influence. Following the guidelines in subsection Section 2.3.20, some examples are reported below.

Based on IMS-Health quarterly number of cimetidine and ranitidine packages sold in Italy, the CRCD model (Guseo & Mortarino, 2012) was tested to evaluate a diachronic competition that produced substitution. Cimetidine is a histamine antagonist that inhibits the production of stomach acid and was introduced by Smith, Kline & French in 1976. Ranitidine is an alternative active principle introduced by Glaxo in 1981 and was found to have far-improved tolerability and a longer-lasting action. The main effect in delayed competition is that the first compound spread fast but was suddenly outperformed by the new one principle that modified its stand-alone regime. Guseo and Mortarino (2012) give some statistical and forecasting comparisons with the restricted Krishnan-Bass-Kummar Diachronic model (KBKD) by Krishnan et al. (2000). Previous results were improved with the UCRCD model in Guseo and Mortarino (2014) by considering a decomposition of word-of-mouth (WOM) effects in two parts: within-brand and cross-brand contributions. The new active compound exploited a large cross-brand WOM and a positive within-brand effect. After the start of competition, cimetidine experienced a negative WOM effect from its own adopters and benefited from the increase of the category's market potential driven by the antagonist. Forecasting is more realistic with the UCRCD approach and it avoids mistakes in long-term prediction.

Restricted and unrestricted UCRCD models were applied in Germany by Guidolin and Guseo (2016) to the competition between nuclear power technologies and renewable energy technologies (wind and solar; see also Sections 3.4.5, 3.4.6 and 3.4.8) in electricity production. Due to the 'Energiewende' policy started around 2000, the substitution effect, induced by competition, is confirmed by the electricity production data provided by BP.[156] An advance is proposed in Furlan, Mortarino, and Zahangir (2020) with three competitors (nuclear power, wind, and solar technologies) and exogenous control functions obtaining direct inferences that provide a deeper analysis and forecasting improvements in energy transition context.

Previous mentioned intersections between Lotka–Volterra approach and diffusion of innovations competition models suggested a more modulated access to the residual carrying capacity. The Lotka–Volterra with churn model (LVch) by Guidolin and Guseo (2015) represents 'churn effects' preserving within and cross-brand effects in a synchronic context.

An application of LVch model is discussed with reference to the competition/substitution between compact cassettes and compact discs for pre-recorded music in the US market. Obtained results of LVch outperform restricted and unrestricted UCRCD analyses. In this context the residual market is not perfectly accessible to both competitors and this fact, combined with WOM components, allows for better interpretation and forecasting especially in medium and long-term horizons.

A further application of the LVch model, Lotka–Volterra with asymmetric churn (LVac), is proposed in Guidolin and Guseo (2020). It is based on a statistical reduction: The late entrant behaves as a standard Bass (1969) model that modifies the dynamics and the evolution of the first entrant in a partially overlapped market. The case study is offered by a special form of competition where the iPhone produced an inverse cannibalisation of the iPad. The former suffered a local negative interaction with some benefits: A long-lasting life cycle and a larger market size induced by the iPad.

A limitation in models for diachronic competition relates to high number of rivals, implying complex parametric representations with respect to the observed information. A second limitation, but also an opportunity, is the conditional nature of forecasting if the processes partially depend upon exogenous control functions (new policy regulations, new radical innovations, regular and promotional prices, etc.). These tools may be used to simulate the effect of strategic interventions, but a lack of knowledge of such future policies may affect prediction.

3.8.9. Forecasting under data integrity attacks[157]

Data integrity attacks, where unauthorized parties access protected or confidential data and inject false information using various attack templates such as ramping, scaling, random attacks, pulse and smooth-curve, has become a major concern in data integrity control in forecasting (Giani, Bitar, Garcia, McQueen, Khargonekar, & Poolla, 2013; Singer & Friedman, 2014; Sridhar & Govindarasu, 2014; Yue, 2017).

Several previous studies have given attention in anomaly detection pre-processing step in forecasting workflow with varying degree of emphasis. However, according to Yue (2017), the detection of data integrity attacks is very challenging as such attacks are done by highly skilled adversaries in a coordinated manner without notable variations in the historical data patterns (Liang, He, & Chen, 2019). These attacks can cause over-forecasts that demand unnecessary expenses for the upgrade and maintenance, and can eventually lead to poor planning and business decisions (Luo et al., 2018a; Luo, Hong, & Fang, 2018b; Wu, Yu, Cui, & Lu, 2020).

Short-term load forecasting (see Section 3.4.3) is one major field that are vulnerable to malicious data integrity attacks as many power industry functions such as economic dispatch, unit commitment and automatic generation control heavily depend on accurate load forecasts (Liang et al., 2019). The cyberattack on U.S. power grid in 2018 is one such major incident related to the topic. According to the study conducted by Luo et al. (2018a), the widely used load forecasting models fail to

155 This subsection was written by Renato Guseo.

156 https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html (Accessed: 2020-09-01).

157 This subsection was written by Priyanga Dilini Talagala.

produce reliable load forecast in the presence of such large malicious data integrity attacks. A submission to the Global Energy Forecasting Competition 2014 (GEFCom2014) incorporated an anomaly detection pre-processing step with a fixed anomalous threshold to their load forecasting framework (Xie & Hong, 2016). The method was later improved by Luo, Hong, Yue (2018) by replacing the fixed threshold with a data driven anomalous threshold. Sridhar and Govindarasu (2014) also proposed a general framework to detect scaling and ramp attacks in power systems. Akouemo and Povinelli (2016) investigated the impact towards the gas load forecasting using hybrid approach based on Bayesian maximum likelihood classifier and a forecasting model. In contrast to the previous model based attempts, Yue, Hong, and Wang (2019) proposed a descriptive analytic-based approach to detect cyberattacks including long anomalous subsequences (see Section 2.2.3), that are difficult to detect by the conventional anomaly detection methods.

The problem of data integrity attacks is not limited to load forecasting. Forecasting fields such as elections forecasting (see Section 3.8.5), retail forecasting (see Section 3.2.4), airline flight demand forecasting (see Section 3.8.2) and stock price forecasting (see Section 3.3.13) are also vulnerable to data integrity attacks (Luo et al., 2018a; Seaman, 2018). For instant, Wu et al. (2020) explored the vulnerability of traffic modelling and forecasting in the presence of data integrity attacks with the aim of providing useful guidance for constrained network resource planning and scheduling.

However, despite of the increasing attention toward the topic, advancements in cyberattacks on critical infrastructure raise further data challenges. Fooling existing anomaly detection algorithms via novel cyberattack templates is one such major concern. In response to the above concern, Liang et al. (2019) proposed a data poisoning algorithm that can fool existing load forecasting approaches with anomaly detection component while demanding further investigation into advanced anomaly detection methods. Further, adversaries can also manipulate other related input data without damaging the target data series. Therefore, further research similar to (Sobhani et al., 2020) are required to handle such data challenges.

3.8.10. The forecastability of agricultural time series[158]

The forecasting of agricultural time series falls under the broader group of forecasting commodities, of which agricultural and related products are a critical subset. While there has been considerable work in the econometrics and forecasting literature on common factor models in general there is surprisingly little work so far on the application of such models for commodities and agricultural time series – and this is so given that there is considerable literature in the linkage between energy and commodities, including agricultural products, their prices and futures prices, their returns and volatilities. Furthermore, a significant number of papers is fairly recent which indicates that there are many open avenues of future research on these topics, and in particular for applied forecasting. The literature on the latter connection can consider many different aspects in modelling as we illustrate below. We can identify two literature strands, a much larger one on the various connections of energy with commodities and the agricultural sector (and in this strand we include forecasting agricultural series) and a smaller one that explores the issue of common factors.

An early reference of the impact of energy on the agricultural sector is Tewari (1990) and then after a decade we find Cohin and Chantret (2010) on the long-run impact of energy prices on global agricultural markets. Byrne, Fazio, and Fiess (2013) is an early reference for co-movement of commodity prices followed by Daskalaki, Kostakis, and Skiadopoulos (2014) on common factors of commodity future returns and then a very recent paper from Alquist, Bhattarai, and Coibion (2020) who link global economic activity with commodity price co-movement. The impact of energy shocks on US agricultural productivity was investigated by Wang and McPhail (2014) while Koirala, Mishra, D'Antoni, and Mehlhorn (2015) explore the non-linear correlations of energy and agricultural prices with Albulescu, Tiwari, and Ji (2020) exploring the latter issue further, the last two papers using copulas. Xiong, Li, Bao, Hu, and Zhang (2015) is an early reference of forecasting agricultural commodity prices while Kyriazi et al. (2019), Li, Li, Liu, Zhu and Wei (2020), and Wang, Wang, Li, and Zhou (2019) consider three novel and completely different approaches on forecasting agricultural prices and agricultural futures returns. López Cabrera and Schulz (2016) explore volatility linkages between energy and agricultural commodity prices and then Tian, Yang, and Chen (2017) start a mini-stream on volatility forecasting on agricultural series followed among others by the work of Luo, Klein, Ji, and Hou (2019) and of Degiannakis, Filis, Klein, and Walther (2020). de Nicola, De Pace, and Hernandez (2016) examine the co-movement of energy and agricultural returns while Kagraoka (2016) and Lübbers and Posch (2016) examine common factors in commodity prices. Pal and Mitra (2019) and Wei Su, Wang, Tao, and Oana-Ramona (2019) both investigate the linkages of crude oil and agricultural prices. Finally, Tiwari, Nasreen, Shahbaz, and Hammoudeh (2020) examine the time-frequency causality between various commodities, including agricultural and metals.

There is clearly room for a number of applications in the context of this recent research, such along the lines of further identifying and then using common factors in constructing forecasting models, exploring the impact of the COVID-19 crisis in agricultural production or that of climate changes on agricultural prices.

3.8.11. Forecasting in the food and beverage industry[159]

Reducing the ecological impact and waste, and increasing the efficiency of the food and beverage industry are currently major worldwide issues. To this direction, efficient and sustainable management of perishable food and the control of the beverage quality is of paramount importance. A particular focus on this topic is placed on supply

[158] This subsection was written by Dimitrios Thomakos.

[159] This subsection was written by Daniele Apiletti.

chain forecasting (see Section 3.2.2), with advanced monitoring technologies able to track the events impacting and affecting the food and beverage processes (La Scalia, Micale, Miglietta, & Toma, 2019). Such technologies are typically deployed inside manufacturing plants, yielding to Industry 4.0 solutions (Ojo, Shah, Coutroubis, Jiménez, & Ocana, 2018) that are enabled by state-of-the-art forecasting applications in smart factories. The transition from plain agriculture techniques to smart solutions for food processing is a trend that fosters emerging forecasting data-driven solutions in many parts of the world, with special attention to the sustainability aspects (Zailani, Jeyaraman, Vengadasan, & Premkumar, 2012).

Various forecasting approaches have been successfully applied in the context of the food and beverage industry, from Monte Carlo simulations based on a shelf-life model (La Scalia et al., 2019), to association rule mining (see Section 2.9.2) applied to sensor-based equipment monitoring measurements (Apiletti & Pastor, 2020), multi-objective mathematical models for perishable supply chain configurations, forecasting costs, delivery time, and emissions (Wang, Nhieu, Chung, & Pham, 2021), and intelligent agent technologies for network optimisation in the food and beverage logistics management (Mangina & Vlachos, 2005).

We now focus on the case of forecasting the quality of beverages, and particularly coffee. Espresso coffee is among the most popular beverages, and its quality is one of the most discussed and investigated issues. Besides human-expert panels, electronic noses, and chemical techniques, forecasting the quality of espresso by means of data-driven approaches, such as association rule mining, is an emerging research topic (Apiletti & Pastor, 2020; Apiletti, Pastor, Callà, & Baralis, 2020; Kittichotsatsawat, Jangkrajarng, & Tippayawong, 2021).

The forecasting model of the espresso quality is built from a real-world dataset of espresso brewing by professional coffee-making machines. Coffee ground size, coffee ground amount, and water pressure have been selected among the most influential external variables. The ground-truth quality evaluation has been performed for each shot of coffee based on three well-known quality variables selected by domain experts and measured by specific sensors: the extraction time, the average flow rate, and the espresso volume. An exhaustive set of more than a thousand coffees has been produced to train a model able to forecast the effect of non-optimal values on the espresso quality.

For each variable considered, different categorical values are considered: ground size can be coarse, optimal, or fine; ground amount can be high, optimal, or low; brewing water pressure can be high, optimal, or low. The experimental setting of categorical variables enables the application of association rule mining (see Section 2.9.2), a powerful data-driven exhaustive and explainable approach (Han et al., 2011; Tan et al., 2005), successfully exploited in different application contexts (Acquaviva et al., 2015; Di Corso et al., 2018).

Several interesting findings emerged. If the water pressure is low, the amount of coffee ground is too high, and the grinding is fine, then we can forecast with confidence a low-quality coffee due to excessive percolation time. If the amount of coffee ground is low, the ground is coarse, and the pressure is high, then we can forecast a low-quality coffee due to excessive flow rate. Furthermore, the coarseness of coffee ground generates an excessive flow rate forecast, despite the optimal values of dosage and pressure, with very high confidence.

3.8.12. *Dealing with logistic forecasts in practice*[160]

The forecaster faces three major difficulties when using the logistic equation (S curve); see also Section 2.3.19. A first dilemma is whether he or she should fit an S curve to the cumulative number or to the number per unit of time. Here the forecaster must exercise wise judgment. What is the "species" and what is the niche that is being filled? To the frustration of business people there is no universal answer. When forecasting the sales of a new product it is often clear that one should fit the cumulative sales because the product's market niche is expected to eventually fill up. But if we are dealing with something that is going to stay with us for a long time (for example, the Internet or a smoking habit), then one should not fit cumulative numbers. At times this distinction may not be so obvious. For example, when COVID-19 first appeared many people (often amateurs) began fitting S curves to the cumulative number of infections (for other attempts on forecasting COVID-19, see Section 3.6.2). Some of them were rewarded because indeed the diffusion of the virus in some countries behaved accordingly (Debecker & Modis, 1994). But many were frustrated and tried to "fix" the logistic equation by introducing more parameters, or simply gave up on trying to use logistics with COVID 19. And yet, many cases (e.g., the US) can be illuminated by logistic fits but on the daily number of infections, not on the cumulative number. As of August 1, 2020, leaving out the three eastern states that had gotten things under control, the rest of the US displayed two classic S curve steps followed by plateaus (see Fig. 18). The two plateaus reflect the number of infections that American society was willing to tolerate at the time, as the price to pay for not applying measures to restrict the virus diffusion.

The second difficulty in using the logistic equation has to do with its ability to predict from relatively early measurements the final ceiling. The crucial question is how early can the final ceiling be determined and with what accuracy. Some people claim that before the midpoint no determination of a final level is trustworthy (Marinakis & Walsh, 2021). Forecasters usually abstain from assigning quantitative uncertainties on the parameters of their S curve forecasts mostly because there is no theory behind it. However, there is a unique study by Debecker and Modis (2021) that quantifies the uncertainties on the parameters determined by logistic fits. The study was based on 35,000 S curve fits on simulated data, smeared by random noise and covering a variety of conditions. The fits were carried out via a χ^2 minimisation technique. The

[160] This subsection was written by Theodore Modis.

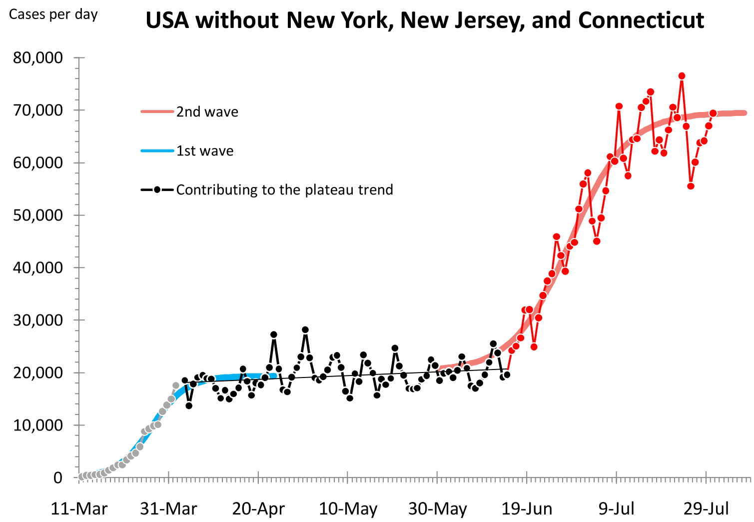

Fig. 18. Two logistic-growth steps during the early diffusion of COVID-19 in America (March to July, 2020).

study produced lookup tables and graphs for determining the uncertainties expected on the three parameters of the logistic equation as a function of the range of the S curve populated by data, the error per data point, and the confidence level required.

The third difficulty using the logistic equation comes from the fact that no matter what fitting program one uses, the fitted S curve will flatten toward a ceiling *as early and as low* as it is allowed by the constraints of the procedure. As a consequence fitting programs may yield logistic fits that are often biased toward a low ceiling. Bigger errors on the data points accentuate this bias by permitting larger margins for the determination of the S curve parameters. To compensate for this bias the user must explore several fits with different weights on the data points during the calculation of the χ^2. He or she should then favour the answer that gives the highest ceiling for the S curve (most often obtained by weighting more heavily the recent historical data points). Of course, this must be done with good justification; here again the forecaster must exercise wise judgment.

3.9. The future of forecasting practice[161]

> Plus ça change, plus c'est la même chose.
>
> Jean-Baptiste Karr (1849)

It would be a more straightforward task to make predictions about the future of forecasting practice if we had a better grasp of the present state of forecasting practice. For that matter, we lack even a common definition of forecasting practice. In a recent article, Makridakis, Bonneli et al. (2020) lamented the failure of truly notable advances in forecasting methodologies, systems, and processes during the past decades to convince many businesses to adopt systematic forecasting procedures, leaving a wide swath of commerce under the guidance of ad hoc judgment and intuition. At the other extreme, we see companies with implementations that combine state-of-the-art methodology with sophisticated accommodations of computing time and costs, as well as consideration of the requirements and capabilities of a diverse group of stakeholders (Yelland, Baz, & Serafini, 2019). So, it is not hyperbole to state that business forecasting practices are all over the place. What surely is hyperbole, however, are the ubiquitous claims of software providers about their products accurately forecasting sales, reducing costs, integrating functions, and elevating the bottom line (Makridakis, Bonneli et al., 2020; Sorensen, 2020). For this section, we grilled a dozen practitioners and thought leaders ("the Group") about developments playing out in the next decade of forecasting practice, and have categorised their responses:

- Nature of forecasting challenges;
- Changes in the forecasting toolbox;
- Evolution in forecasting processes such as integration of planning functions;
- Expectations of forecasters; and
- Scepticism about real change.

Forecasting Challenges: Focusing on operations, the Group sees demand forecasting becoming ever more difficult due to product/channel proliferation, shorter lead times, shorter product histories, and spikes in major disruptions.

- Operational forecasts will have shorter forecast horizons to increase strategic agility required by business to compete, sustain, and survive.
- New models will need to incorporate supply-chain disruption. Demand chains will need to be restarted, shortening historical data sets and making traditional models less viable due to limited history.

[161] This subsection was written by Len Tashman.

- Lead times will decrease as companies see the problems in having distant suppliers. Longer lead times make accurate forecasting more difficult.

Forecasting Tool Box: Unsurprisingly, this category received most of the Group's attention. All predict greater reliance on AI/ML for automating supply-and-demand planning tasks and for reconciling discrepancies in hierarchical forecasting. Longer-horizon causal forecasting models will be facilitated by big data, social media, and algorithmic improvements by quantum computing. Post-COVID, we will see a greater focus on risk management/mitigation. The Cloud will end the era of desktop solutions.

- Quantum computers will improve algorithms used in areas like financial forecasting (e.g., Monte Carlo simulations), and will change our thinking about forecasting and uncertainty.
- Although social media is a tool for "what's trending now", new models will be developed to use social-media data to predict longer-term behaviour. Step aside Brown (exponential smoothing) and Bass (diffusion).
- Greater automation of routine tasks (data loading, scrubbing, forecast generation and tuning, etc.) through AI/ML-powered workflow, configurable limits, and active alerts. More black box under the hood, but more clarity on the dashboard.
- Greater focus on risk management/mitigation through what-if scenarios, simulations, and probabilistic forecasting.

Forecasting Processes and Functional Integration: Systems will become more integrated, promoting greater collaboration across functional areas and coordination between forecast teams and those who rely upon them. Achieving supply-chain resilience will become as important as production efficiency, and new technology such as Alert and Root Cause Analysis systems will mitigate disruptions.

- S&OP will expand from its home in operations to more fully integrate with other functions such as finance and performance management, especially in larger multinationals.
- The pandemic has forced firms to consider upping supply-chain resilience. Firms are building in capacity, inventory, redundancy into operations—somewhat antithetical to the efficiency plays that forecasting brings to the table.
- Forecasting will be more closely tied to Alert and Root Cause Analysis systems, which identify breakdowns in processes/systems contributing to adverse events, and prevent their recurrence.

Expectations of Forecasters: Agreement was universal that the forecaster's job description will broaden and become more demanding, but that technology will allow some redirection of effort from producing forecasts to communicating forecasting insights.

- The interest around disease models increases our awareness of the strengths and weaknesses of mathematical models. Forecasters may need to become more measured in their claims, or do more to resist their models being exploited.
- We will see a transformation from demand planner to demand analyst, requiring additional skill sets including advanced decision making, data and risk analysis, communication, and negotiation.
- Professional forecasters will be rare except in companies where this expertise is valued. Fewer students are now educated or interested in statistical modelling, and time is not generally available for training.
- Forecasters will learn the same lesson as optimisation folks in the 1990s and 2000s: the importance of understanding the application area—community intelligence.

Scepticism: Many were sceptical about the current enthusiasm for AI/ML methods; disappointed about the slow adoption of promising new methods into software systems and, in turn, by companies that use these systems; and pessimistic about the respect given to and influence of forecasters in the company's decision making.

- While AI/ML are important additions to the forecaster's toolbox, they will not automatically solve forecasting issues. Problems include data hunger, capacity brittleness, dubious input data, fickle trust by users (Kolassa, 2020c), and model bias.
- Practices in the next decade will look very similar to the present. Not that much has changed in the last decade, and academic developments are slow to be translated into practice.
- Politics, gaming, and the low priority given to forecasting are the prime drivers of practice, thus limiting interest in adopting new methodologies.
- None of the topical items (AI/ML, big data, demand sensing, new forecasting applications) will have much of an impact on forecasting practice. Forecasting departments hop from one trend to the other without making much progress towards better forecasting accuracy.
- Software companies will struggle, despite good offerings. Most companies do not want to invest in excellent forecasting engines; whatever came with their ERP system is "good enough".
- Forecasting will continue to suffer from neglect by higher levels of management, particularly when forecasts are inconveniently contrary to the messages management hopes to convey.

Note finally that the COVID-19 pandemic has elevated practitioner concerns about disruptions to normal patterns as well as the fear of an increasingly volatile environment in which forecasts must be made. There are indications that companies will place more stress on judgmental scenarios, likely in conjunction with statistical/ML methods.

4. Forecasting: benefits, practices, value, and limitations[162]

> Mr. Buffett said his advice for the cash left to his wife was that 10 per cent should go to short-term government bonds and 90 per cent into a very low-cost S&P 500 index fund.

The purpose of this unique article is to provide an encyclopedic knowledge about the various aspects of forecasting. In this article, there are more than 140 sections and subsections, with more than 2,100 references, written by 80 of some of the best-known forecasting researchers and practitioners in the world, making it into a selective, encyclopedic piece covering, into a single source, a great deal of the available knowledge about the theory and practice of forecasting. We hope that this article will serve as an easy-to-use reference source. We aim to convert it into an online resource that will be regularly updated as new information becomes available.

But some people argue if there is any value in attempting to predict the future and if forecasting is any different than fortune telling, given the large numbers of mistaken forecasts made in the past, including our inability to accurately predict the progression of COVID-19 and its economic and human consequences? What is, therefore, the usefulness of a paper like the present one when crystal balling is not possible, and uncertainty reigns? It is the aim of this concluding article to set the record straight, explaining the benefits and practical value of forecasting while reporting its limitations too.

The Myriad of Forecasts: All planning and the great majority of decisions we make require forecasting. Deciding what time to get up in the morning, not to be late for work implies a correct prediction of the commuting time to go to the office. Determining what to study is another decision requiring elaborate predictions about the demand for future jobs decades away. In the business world, firms must decide/forecast how many products to manufacture, the price they should be sold, how much money to spend on advertising and promotion, how much and in what type of new technologies to invests and a plethora of other future-oriented decisions requiring both predictions and assessing their inevitable uncertainty. Whether we like it or not, we have no choice but making these forecasts to benefit as much as possible from their value, knowing perfectly well that all predictions are uncertain while some may turn out to be wrong.

The Pervasiveness of Uncertainty: Apart from some areas of hard sciences, all other forecasts are uncertain and must be accompanied with a measure of its magnitude, expressed as a prediction interval, or as a probability distribution around the most likely forecast. Although the value and usage of forecasts is clear, that of uncertainty is not. Worse, it becomes an unwelcome source of anxiety whose usefulness is misunderstood. Executives want to know the exact sales of their firm for next month to set up their production schedule. Instead, they are given prediction intervals (PIs) around such forecast and told that most of the time, sales will be within this interval, assuming the fluctuations follow some distributional assumptions. They argue that forecasting must decrease, not amplify, uncertainty and that the PIs are too wide and 'uninformative' to be used for making practical business decisions. The trouble is that these PIs are based on past fluctuations and present the best estimation of future uncertainty, even if they seem too wide. Worse, empirical research has shown that they are too narrow, underestimating uncertainty often considerably.

Assessing Uncertainty and Dealing with its Implied Risks: Uncertainty entails risks, requiring action to minimise their negative consequences. There are two kinds of uncertainty that can be illustrated by a commuting example. The first relates to fluctuations in the commuting time under normal driving conditions when there are no serious accidents, road works or major snowstorms. Such fluctuations are small and can be captured by a normal curve that allows to balance the risk of arriving earlier or later than the desired time. In the opposite case, uncertainty is fat-tailed and hard to estimate, as delays can be substantial depending upon the seriousness of the accident or that of the snowstorm while the risk of being early to work is smaller than being late. Moreover, such risk is substantially different when going to the airport to catch a flight, requiring starting much earlier than the average time it takes to go to the airport to minimise the risk of missing the flight.

More Accurate Ways of Forecasting and Assessing Uncertainty: Extensive empirical research, including forecasting competitions, has shown that systematic approaches improve the accuracy of forecasting and the correct assessment of uncertainty resulting in substantial benefits when compared to ad-hoc judgmental alternatives (Makridakis, Bonneli et al., 2020). The biggest advantage of such approaches is their ability to identify and estimate, in a mathematically optimal manner, past patterns and relationships that are subsequently extrapolated to predict their continuation, avoiding the over optimism and wishful thinking associated with judgmental approaches. At the same time, it must be clear that the accuracy of the forecasts and the correctness of uncertainty will depend on the established patterns/relationship not changing much during the forecasting period.

Using Benchmarks to Evaluate the Value of Forecasting: The accuracy of the forecasts and the correct assessment of uncertainty must be judged not on their own but in comparison to some simple, readily available benchmarks. In stock market forecasts, for instance, the accuracy of predictions is compared to that of today's price used as the forecast for future periods. Empirical comparisons have shown that such a benchmark beats the great majority of professional forecasters, hence Buffet's advice in the epigram for his wife to invest in a low-cost index fund that selects stocks randomly. In weather forecasting, meteorologists are judged by the improvement of their forecasts over the naive prediction that tomorrow's weather will be the same as today.

162 This subsection was written by Spyros Makridakis.

Concluding remark: Accepting the advantages and limitations of systematic forecasting methods and most importantly avoiding any exaggerated expectations of what it can achieve is critical. Such methods do not possess any prophetic powers, they simply extrapolate established patterns and relationships to predict the future and assess its uncertainty. Their biggest advantage is their objectivity and ability for optimal extrapolation. Their biggest disadvantages are: (*i*) the patterns and the relationships must remain fairly constant during the forecasting phase for the forecasts to be accurate, and (*ii*) uncertainty must not be fat-tailed so it can be measured quantitatively.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Fotios Petropoulos would like to thank all the co-authors of this article for their very enthusiastic response and participation in this initiave. He would also like to thank Pierre Pinson for inviting this paper to be submitted to the *International Journal of Forecasting*. The constructive comments and suggestions from this advisory board were vital in improving the paper. He also thanks Artur Tarassow for offering a list of Gretl's software functionalities.

Jakub Bijak's work received funding from the European Union's Horizon 2020 research and innovation programme, grant 870299 QuantMig: Quantifying Migration Scenarios for Better Policy.

Clara Cordeiro is partially financed by national funds through FCT – Fundação para a Ciência e a Tecnologia, Portugal under the project UIDB/00006/2020.

Fernando Luiz Cyrino Oliveira acknowledges the support of the Coordination for the Improvement of Higher Level Personnel (CAPES), Brazil – grant number 001, the Brazilian National Council for Scientific and Technological Development (CNPq) – grant number 307403/2019-0, and the Carlos Chagas Filho Research Support Foundation of the State of Rio de Janeiro (FAPERJ) – grant numbers 202.673/2018 and 211.086/2019.

Shari De Baets was funded by the FWO Research Foundation Flanders.

Joanne Ellison acknowledges the support of the ESRC FertilityTrends project (grant number ES/S009477/1) and the ESRC Centre for Population Change (grant number ES/R009139/1).

Piotr Fiszeder was supported by the National Science Centre, Poland project number 2016/21/B/HS4/00662 entitled "Multivariate volatility models - the application of low and high prices".

David T. Frazier has been supported by Australian Research Council (ARC) Discovery Grants DP170100729 and DP200101414, and ARC Early Career Researcher Award DE200101070.

Mariangela Guidolin acknowledges the support of the University of Padua, Italy, through the grant BIRD188753/18.

David F. Hendry gratefully acknowledges funding from the Robertson Foundation, USA and Nuffield College, UK.

Yanfei Kang acknowledges the support of the National Natural Science Foundation of China (number 11701022) and the National Key Research and Development Program, China (number 2019YFB1404600).

Stephan Kolassa would like to thank Tilmann Gneiting for some very helpful tips.

Gael M. Martin has been supported by Australian Research Council (ARC) Discovery Grants DP170100729 and DP200101414.

Alessia Paccagnini acknowledges the research support by COST Action "Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry" (FinAI) CA19130.

Jose M. Pavía acknowledges the support of the Spanish Ministry of Science, Innovation and Universities and the Spanish Agency of Research, co-funded with FEDER funds, grant ECO2017-87245-R, and of Consellería d'Innovació, Universitats, Ciència i Societat Digital, Generalitat Valenciana – grant number AICO/2019/053.

Diego J. Pedregal and Juan Ramon Trapero Arenas acknowledge the support of the European Regional Development Fund and Junta de Comunidades de Castilla-La Mancha (JCCM/FEDER, UE) under the project SBPLY/19/180501/000151 and by the Vicerrectorado de Investigación *y* Política Científica from UCLM, Spain through the research group fund program PREDILAB; DOCM 26/02/2020 [2020-GRIN-28770].

David E. Rapach thanks Ilias Filippou and Guofu Zhou for valuable comments.

J. James Reade and Han Lin Shang acknowledge Shixuan Wang for his constructive comments.

Michał Rubaszek is thankful for the financial support provided by the National Science Centre, Poland, grant No. 2019/33/B/HS4/01923 entitled "Predictive content of equilibrium exchange rate models".

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the views of their affiliated institutions and organisations.

Appendix A. List of acronyms

ABC	Approximate Bayesian Computation
ACC	Ant Colony Clustering
ACD	Autoregressive Conditional Duration
ADIDA	Aggregate–Disaggregate Intermittent Demand Approach
ADL	Autoregressive Distributed Lag
ADMM	Alternating Direction Method of Multipliers
AI	Artificial Intelligence

AIC	Akaike's Information Criterion
AICc	Akaike's Information Criterion corrected (for small sample sizes)
ANACONDA	Analysis of National Causes of Death for Action
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
AO	Additive Outlier
AR	AutoRegressive (model)
ARX	AutoRegressive with eXogenous variables (model)
ARCH	AutoRegressive Conditional Heteroskedasticity
ARMA	AutoRegressive-Moving Average (model)
ARIMA	AutoRegressive Integrated Moving Average (model)
ARIMAX	AutoRegressive Integrated Moving Average with eXogenous variables (model)
B&M	Brick and Mortar
BATS	Box–Cox transform, ARMA errors, Trend, and Seasonal components (model)
BEER	Behavioural Equilibrium Exchange Rate
BEKK	Baba-Engle-Kraft-Kroner GARCH
BEKK-HL	BEKK High Low
BIC	Bayesian Information Criterion
BLAST	Building Loads Analysis and System Thermodynamics
BM	Bass Model
BMC	Bootstrap Model Combination
BPNN	Back-Propagation Neural Network
CARGPR	Conditional AutoRegressive Geometric Process Range (model)
CARR	Conditional AutoRegressive Range (model)
CARRS	Conditional AutoRegressive Rogers and Satchell (model)
CBC	Choice Based Conjoint (analysis)
CBO	Congressional Budget Office
CDS	Credit Default Swap
CNN	Convolutional Neural Network
COVID-19	Coronavirus disease 2019
CVAR	Cointegrated Vector AutoRegressive (model)
CAViaR	Conditional AutoRegressive Value At Risk
CL	Cross Learning
CPFR	Collaborative Planning, Forecasting, and Replenishment
CRCD	Competition and Regime Change Diachronic (model)
CRPS	Continuous Ranked Probability Score
CSR	Complete Subset Regression
CV	Cross-Validation
DA	Deterministic Annealing
DC	Distribution Centre
DCC	Dynamic Conditional Correlation
DCC-RGARCH	Range GARCH DCC
DFM	Dynamic Factor Model
DGP	Data Generating Process
DJI	Dow Jones Industrial
DM	Diebold–Mariano (test)
DSGE	Dynamic Stochastic General Equilibrium
DSHW	Double Seasonal Holt-Winters
DSTCC-CARR	Double Smooth Transition Conditional Correlation CARR
DT	Delay Time
EEG	ElectroEncephaloGram
EGARCH	Exponential GARCH
EH	Expectations Hypothesis
EMD	Empirical Mode Decomposition
ENet	Elastic Net
ENSO	El Niño Southern Oscillation
ERCOT	Electric Reliability Council of Texas
ES	Expected Shortfall
ESP-r	Environmental Systems Performance – research
ESTAR	Exponential STAR
ETS	ExponenTial Smoothing (or Error, Trend, Seasonality)
EVT	Extreme Value Theory
EWMA	Exponentially Weighted Moving Average
FAR	Functional AutoRegressive (model)
FASSTER	Forecasting with Additive Switching of Seasonality, Trend and Exogenous Regressors
FCM	Fuzzy C-Means
FIGARCH	Fractionally Integrated GARCH
FIS	Fuzzy Inference System
FFNN	Feed-Forward Neural Network
FFORMA	Feature-based FORecast Model Averaging
FMCG	Fast Moving Consumer Goods
FPCA	Functional Principal Component Analysis
FRB/EDO	Federal Reserve Board's Estimated, Dynamic, Optimisation-based (model)
FSS	Forecasting Support System
FVA	Forecast Value Added
GARCH	General AutoRegressive Conditional Heteroscedasticity
GARCH-PARK-R	GARCH PARKinson Range
GARCH-TR	GARCH True Range

GB	Givon-Bass (model)
GBM	Generalised Bass Model
GDP	Gross Domestic Product
GGM	Guseo-Guidolin Model (GGM)
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
GM	Generalised M-estimator
GMM	Generalised Methods of Moments
GPU	Graphics Processing Unit
GRNN	Generalised Regression Neural Network
HAR	Heterogeneous AutoRegressive (model)
HDFS	Hadoop Distributed File System
HMD	Human Mortality Database
HP	Hodrick–Prescott
HPU	House Price Uncertainty
HVAC	Heating, Ventilation, and Air Conditioning (system)
HAR	Heterogeneous AutoRegressive (model)
HQ	Hannan-Quinn
IEA	International Energy Agency
IG	Interaction Groups
iid	independent and identically distributed
IIS	Impulse Indicator Saturation
IO	Innovation Outlier
IT	Information Technology
KBKD	Krishnan-Bass-Kummar Diachronic (model)
KISS	Keep It Simple, Stupid (principle)
*k*NN	*k* Nearest Neighbours
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
L-IVaR	Liquidity-adjusted Intraday Value-at-Risk
LASSO	Least Absolute Shrinkage and Selection Operator
LH	Low and High
LLN	Law of Large Numbers
LN-CASS	Logit-Normal Continuous Analogue of the Spike-and-Slab
LS (or LPS)	Logarithmic (Predictive) Score (log-score)
LSTAR	Logistic STAR
LSTM	Long Short-Term Memory Networks
LTLF	Long-Term Load Forecasting
LV	Lotka–Volterra (model)
LVac	Lotka–Volterra with asymmetric churn (model)
LVch	Lotka–Volterra with churn (model)
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error

MASE	Mean Absolute Scaled Error
MCMC	Markov Chain Monte Carlo
MFI	Marginal Forecast Interval
MICE	Meetings, Incentives, Conventions, and Exhibitions/Events
MIDAS	MIxed DAta Sampling
MJO	Madden Julian Oscillation
ML	Machine Learning
MLP	MultiLayer forward Perceptron
MLR	Multiple Linear Regression
MS	Markov Switching
MS VAR	Markov Switching VAR
MSARIMA	Multiple/Multiplicative Seasonal ARIMA
MSC	Multiple Seasonal Cycles
MSE	Mean Squared Error
MSRB	Markov-Switching Range-Based
MTA	Multiple Temporal Aggregation
MTLF	Medium-Term Load Forecasting
NAO	North Atlantic Oscillation
NLS	Nonlinear Least Squares
NLTK	Natural Language Toolkit
NMAE	Normalised Mean Absolute Error
NN	Neural Network
NNAR	Neural Network AutoRegressive
NOB	Non-Overlapping Blocks
NPF	New Product Forecasting
NWP	Numerical Weather Prediction
OB	Overlapping Blocks
OBR	Office for Budget Responsibility
ODE	Ordinary Differential Equations
OLS	Ordinary Least Squares
OWA	Overall Weighted Average
PAR	Periodic AutoRegressive (model)
PCA	Principal Components Analysis
pdf	probability density function
PdM	Predictive Maintenance
PFEM	Point Forecast Error Measure
PHANN	Physical Hybrid Artificial Neural Network
pHDR	predictive Highest Density Region
PI	Prediction Interval
PIT	Probability Integral Transform
PL	Product Level
PLS	Partial Least Squares
PM	Particulate Matter
POT	Peak Over Threshold
PPP	Purchasing Power Parity
PSO	Particle Swarm Intelligence
PV	PhotoVoltaic

Q_α	quantile score or pinball loss for a level $\alpha \in (0, 1)$
RAF	Royal Air Force (UK)
RB	Range-Based
RB-copula	Range-Based copula
RB-DCC	Range-Based DCC
RB-MS-DCC	Range-Based Markov-Switching DCC
RBF	Radial Basis Function
REGARCH	Range-Based Exponential GARCH
RET	Renewable Energy Technology
RGARCH	Range GARCH
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RR-HGADCC	Return and Range Heterogeneous General Asymmetric DCC
RTV	Real Time Vintage
SA	Structured Analogies
SARIMA	Seasonal AutoRegressive Integrated Moving Average (model)
SARIMAX	Seasonal AutoRegressive Integrated Moving Average with eXogenous variables
SARMA	Seasonal AutoRegressive Moving Average (model)
SARMAX	Seasonal AutoRegressive Moving Average with eXogenous variables
SBA	Syntetos-Boylan Approximation
SBC	Syntetos-Boylan-Croston (classification)
SEATS	Seasonal Extraction in ARIMA Time Series
SES	Simple (or Single) Exponential Smoothing
SETAR	Self-Exciting Threshold AutoRegressive (model)
SFI	Simultaneous Forecast Interval
SKU	Stock Keeping Unit
SGD	Stochastic Gradient Descent
SIS	Step Indicator Saturation
SL	Serial number Level
SMA	Simple Moving Average
sMAPE	symmetric Mean Absolute Percentage Error
SOM	Self-Organising Map
SS	State Space
sSA	semi-Structured Analogies
SSARIMA	Several Seasonalities (or State Space) ARIMA
STAR	Smooth Transition AutoRegressive (model)
STARR	Smooth Transition conditional AutoRegressive Range (model)
STL	Seasonal Trend decomposition using Loess
STLF	Short-Term Load Forecasting
STR	Seasonal-Trend decomposition based on Regression
SV	Stochastic Volatility
SVA	Stochastic Value Added
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SWAN	Simulating WAves Nearshore
S&OP	Sales and Operations Planning
S&P	Standard & Poor's
TAR	Threshold AutoRegressive (model)
TARMA	Threshold AutoRegressive Moving Average (model)
TARMASE	Threshold AutoRegressive Moving Average (model)
TARR	Range-Based Threshold conditional AutoRegressive (model)
TBATS	Exponential Smoothing state space model with Box–Cox transformation, ARMA errors, Trend and Seasonal components
TFR	Total Fertility Rate
TGARCH	Threshold GARCH
TMA	Threshold Moving Average (model)
TPU	Tensor Processing Unit
TRAMO	Time series Regression with ARIMA noise, Missing values and Outliers
TSB	Teunter-Syntetos-Babai (method)
UCRCD	Unbalanced Competition and Regime Change Diachronic (model)
UIP	Uncovered Interest Party
VaR	Value at Risk
VAR	Vector AutoRegressive (model)
VARX	VAR with eXogenous variables (model)
VARMA	Vector AutoRegressive Moving Average (model)
VAT	Value Added Tax
VARIMAX	Vector AutoRegressive Integrated Moving Average with eXogenous variables (model)
VECM (or VEC)	Vector Error Correction Model
VEqCM	Vector Equilibrium-Correction Model
VSTLF	Very Short-Term Load Forecasting
WLS	Weighted Least Squares
WNN	Wavelet Neural Network
WOM	Word-Of-Mouth
WW2	World War 1
WW2	World War 2
WW3	World War 3
XGBoost	eXtreme Gradient Boosting

Table B.1
A list of indicative free or open-source packages, libraries, and toolboxes linking to the theory sections of this article. The authors assume no liability for the software listed below; interested users are strongly advised to read the respective documentations and licences terms.

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
Section 2.2.1. Box–Cox transformations	R	*forecast*	BoxCox; InvBoxCox; BoxCox.lambda;	Functions to transform the input variable using a Box–Cox transformation, reverse the transformation and find optimal parameters. https://cran.r-project.org/package=forecast
Section 2.2.2. Box–Cox transformations Time series decomposition	R	*stats*	decompose; stl	Classical decomposition method (additive and multiplicative), and STL decomposition method. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html
	R	*forecast*	seasadj; seasonal; mstl; msts; tbats.components	Tools for extracting components, and multiple seasonal decomposition methods. https://cran.r-project.org/package=forecast
	R	*tsutils*	decomp; seasplot	Classical decomposition method, and functions for seasonal plots. https://cran.r-project.org/package=tsutils
	R	*stR*	AutoSTR; STR; heuristicSTR; plot.STR; seasadj.STR	Seasonal-Trend decomposition based on Regression. https://cran.r-project.org/package=stR
	R	*seasonal*	seas	Functions for X-11, SEATS, and X-13-ARIMA-SEATS decomposition methods. https://cran.r-project.org/package=seasonal
	Gretl	*buys_ballot*		Plot seasonal time series components. http://gretl.sourceforge.net/
	Gretl	*season_plot*	set_season_plot	Plot seasonal time-series components. http://gretl.sourceforge.net/
	Gretl	*tsfcst*	decompcl	Classical time series decomposition. http://gretl.sourceforge.net/
	Gretl	*StrucTiSM*	STSM_components	Decomposition using structural timeseries model. http://gretl.sourceforge.net/
Section 2.2.3. Anomaly detection and time series forecasting	R	*anomalize*	time_decompose; anomalize; time_recompose	A "tidy" workflow for detecting anomalies in data. https://cran.r-project.org/package=anomalize
	R	*oddstream*	find_odd_streams; extract_tsfeatures; set_outlier_threshold	Early detection of anomalous series within a large collection of streaming time series data. https://cran.r-project.org/package=oddstream
	R	*tsoutliers*	tso; locate.outliers.oloop; remove.outliers	Detection of outliers in time series such as Innovational outliers, additive outliers, level shifts, temporary changes and seasonal level shifts. https://cran.r-project.org/package=tsoutliers
	R	*stray*	find_HDoutliers; find_threshold; display_HDoutliers	Anomaly detection in high dimensional and temporal data. https://cran.r-project.org/package=stray
	R	*forecast*	tsoutliers; tsclean	Provides some simple heuristic methods for identifying and correcting outliers. https://cran.r-project.org/package=seasonal
	R	*OutliersO3*	OutliersO3; O3plotM; O3plotT; O3prep	Draws overview of outliers (O3) Plots. https://cran.r-project.org/package=OutliersO3
	R	CRAN Task View	Anomaly Detection with R	Contains a list of R packages that can be used for anomaly detection. https://github.com/pridiltal/ctv-AnomalyDetection
	Gretl	*tramolin*		Outlier detection/correction and missing data interpolation. http://gretl.sourceforge.net/
Section 2.2.4. Robust handling of outliers in time series forecasting	R	*gets*	isat	Function for running impulse and step indicator saturation. https://cran.r-project.org/package=gets

(*continued on next page*)

Table B.1 (*continued*).

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
Section 2.3.1. Exponential smoothing models	R	*forecast*	ets; forecast.ets; ses;	Functions for simple exponential smoothing and automatic exponential smoothing modelling. https://cran.r-project.org/package=forecast
	R	*smooth*	es	Function for automatic exponential smoothing modelling. https://cran.r-project.org/package=smooth
	Gretl	*tsfcst*	expsmpars	Simple exponential smoothing minimising the sum of squared errors. http://gretl.sourceforge.net/
Section 2.3.2. Time-series regression models	R	*stats*	lm	Fitting linear regression models. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html
	R	*leaps*	regsubsets	Functions for selecting linear regression models. https://cran.r-project.org/package=leaps
	R	*relaimpo*		Relative importance of regressors in linear models. https://cran.r-project.org/package=relaimpo
	R	*MASS*	stepAIC	Choose a model by AIC in a stepwise algorithm. https://cran.r-project.org/package=MASS
	Gretl		ols, lad, midasreg	Regression models with OLS, LAD, and MIDAS with functionality for forecasting. http://gretl.sourceforge.net/
Section 2.3.3. Theta method and models	R	*forecast*	thetaf	Returns forecasts and prediction intervals for a theta method forecast. https://cran.r-project.org/package=forecast
	R	*forecTheta*	stheta; stm; otm; dstm; dotm	Functions for forecasting univariate time series using Theta Models. https://cran.r-project.org/package=forecTheta
	R	*tsutils*	theta	Estimate Theta method. https://cran.r-project.org/package=tsutils
	Gretl	*tsfcst*	stheta	Theta-method for univariate forecasting. http://gretl.sourceforge.net/
Section 2.3.4. Autoregressive integrated moving average (ARIMA) models	R	*forecast*	auto.arima; Arima; arfima; arima.errors; arimaorder	Functions for fitting and forecasting with ARIMA models. https://cran.r-project.org/package=forecast
	R	*smooth*	auto.msarima; auto.ssarima; msarima; ssarima	State-space and multiple seasonalities implementations of ARIMA models. https://cran.r-project.org/package=smooth
	Gretl		arima	Functions for fitting and forecasting with SARIMAX models. http://gretl.sourceforge.net/
	Gretl	*auto_arima*		Find best fitting SARIMAX model with functions for forecasting. http://gretl.sourceforge.net/
	Gretl	*armax*		Automatically determine the best ARMAX model. http://gretl.sourceforge.net/
	R	*aTSA*	adf.test	Augmented Dickey–Fuller test. https://cran.r-project.org/package=aTSA
	R	*tseries*	kpss.test	Kwiatkowski-Phillips–Schmidt–Shin (KPSS) test. https://cran.r-project.org/package=tseries
	R	*forecast*	ndiffs; nsdiffs	Estimates the number of (seasonal) differences required to make a given time series stationary. https://cran.r-project.org/package=forecast
	Gretl		adf; adf-gls; kpss; levinlin	Various unit-root tests for time-series and panel data. http://gretl.sourceforge.net/
	Gretl	*DHF_test*		Package for Dickey-Hasza-Fuller seasonal Unit Root Test. http://gretl.sourceforge.net/
	Gretl	*DP*		Package for testing for a double unit root. http://gretl.sourceforge.net/
	Gretl	*GHegy*		Package Seasonal unit roots tests. http://gretl.sourceforge.net/
	Gretl	*Kapetanios*		Package for Kapetanios' unit root test with possible structural breaks. http://gretl.sourceforge.net/
	Gretl	*PPtest*		Package for running Phillips–Perron unit root test. http://gretl.sourceforge.net/
	Gretl	*VSG_test*		Package for test proposed by Ventosa-Santaulària and Gómez-Zaldívar. http://gretl.sourceforge.net/
Section 2.3.5. Forecasting for multiple seasonal cycles	R	*smooth*	msarima; ssarima	Functions for forecasting data with multiple seasonal cycles. https://cran.r-project.org/package=smooth
	R	*fable*	model; forecast; fasster; ETS; ARIMA; TSLM	Forecasting models for tidy time series. https://cran.r-project.org/package=fable

(*continued on next page*)

International Journal of Forecasting 38 (2022) 705–871

Table B.1 (*continued*).

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
	R, Python	*prophet*		Facebook's automatic forecasting procedure. https://cran.r-project.org/package=prophet
	R	*tidymodels*		Collection of packages for modelling and machine learning using tidyverse principles. https://cran.r-project.org/package=tidymodels
	R	*forecast*	tbats; dshw	Functions for forecasting data with multiple seasonal cycles. https://cran.r-project.org/package=forecast
	R	*fable.prophet*	prophet; forecast	A tidy R interface to the prophet forecasting procedure using fable. https://github.com/mitchelloharawild/fable.prophet
Section 2.3.6. State-space models	Matlab	*SSpace*		General modelling of linear, non-linear and non-Gaussian State Space systems. https://github.com/djpedregal/SSpace
	Matlab	*SSM*		General modelling of linear, non-linear and non-Gaussian State Space systems. https://www.mathworks.com/help/econ/ssm-class.html
	Matlab	*SSMMATLAB*		A Set of MATLAB Programs for the Statistical Analysis of State Space Models. https://github.com/vgomezenriquez/ssmmatlab
	Matlab	*E4*		A MATLAB toolbox for time series analysis in State Space form. https://www.ucm.es/e-4
	R	*UComp*		Automatic identification of Unobserved Components models in State Space form. https://cran.r-project.org/package=UComp
	R	*statespacer*		State Space modelling, mainly ARIMA and Basic Structural Models. https://cran.r-project.org/package=statespacer
	R	*smooth*		Forecasting using single error State Space models. https://cran.r-project.org/package=smooth
	R	*bssm*		Bayesian Inference of Non-Gaussian State Space Models. https://cran.r-project.org/package=bssm
	R	*mssm*		Multivariate State Space models. https://cran.r-project.org/package=mssm
	R	*KFAS*		Kalman Filter and Smoother for Exponential Family State Space Models. https://cran.r-project.org/package=KFAS
	R	*TSSS*		Time Series Analysis with State Space Model, based on the methods in Kitagawa (1993). https://cran.r-project.org/package=TSSS
	R	*dlm*		Bayesian and Likelihood Analysis of Dynamic Linear Models (Gaussian State Space models). https://cran.r-project.org/package=dlm
	Python	*statsmodels*	statespace	Time Series Analysis by State Space Methods. https://www.statsmodels.org/stable/index.html
	Gretl		kfilter; ksmooth; kdsmooth; ksimul	State Space Modeling functionality with function for forecasting. http://gretl.sourceforge.net/
	Gretl	*StrucTiSM*	STSM_fcast	Harvey-style Structural Time Series Models with function for forecasting. http://gretl.sourceforge.net/
Section 2.3.7. Models for population processes	R	*dembase*		General-purpose tools for demographic analysis. https://github.com/StatisticsNZ/dembase
	R	*demest*		Bayesian statistical methods for demography. https://github.com/StatisticsNZ/demest
	R	*demlife*		Creating and working with life tables. https://github.com/StatisticsNZ/demlife
	R	*BayesPop*		Generating population projections for all countries of the world using several probabilistic components, such as total fertility rate and life expectancy. https://cran.r-project.org/package=bayesPop

(*continued on next page*)

Table B.1 (*continued*).

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
	R	*bayesTFR*		Making probabilistic projections of total fertility rate for all countries of the world, using a Bayesian hierarchical model. https://cran.r-project.org/package=bayesTFR
	R	*bayesLife*		Making probabilistic projections of life expectancy for all countries of the world, using a Bayesian hierarchical model. https://cran.r-project.org/package=bayesLife
	Spread-sheet	*DAPPS*		Demographic Analysis and Population Projection System: Standalone spreadsheet-based software for demographic estimation and projections, prepared by the US Census Bureau. https://www.census.gov/data/software/dapps.Overview.html
Section 2.3.9. Forecasting with many variables	R	*gets*	getsm, getsv, isat, isatvar	Package that implements general to specific model selection, indicator saturation, with functionality for forecasting. https://cran.r-project.org/package=gets
	R	*vars*		Functions and routines for VAR Modelling. https://cran.r-project.org/package=vars
	Gretl		var; system	Fitting system-models with functionality for forecasting. http://gretl.sourceforge.net/
Section 2.3.10. Functional time series models	R	*ftsa*	ftsm; farforecast; T_stationarity	Functional time series analysis. https://cran.r-project.org/package=ftsa
Section 2.3.11. ARCH/GARCH models	R	*tseries*	garch	Fit GARCH models to time series. https://cran.r-project.org/package=tseries
	Python	*PyFlux*		Time series analysis and prediction tools that focus on autoregressive methods (ARIMA, ARCH, GARCH, etc.). https://pyflux.readthedocs.io/en/latest/index.html
	Gretl		arch, garch	Fit (G)ARCH models to time series. http://gretl.sourceforge.net/
	Gretl	*gig*	gig_estimate; gig_var_fcast	Estimate various types of GARCH models. http://gretl.sourceforge.net/
Section 2.3.12. Markov switching models	R	*MSwM*		Fitting Markov switching models. https://cran.r-project.org/package=MSwM
	R	*NHMSAR*		Non-homogeneous Markov switching autoregressive models. https://cran.r-project.org/package=NHMSAR
Section 2.3.13. Threshold models	R	*TAR*		Bayesian modelling of autoregressive threshold time series models. https://cran.r-project.org/package=TAR
	R	*TSA*	tar; star	Functions for threshold models (and general time series analysis). https://cran.r-project.org/package=TSA
	Gretl	*Threshold_Panel*	THRESH_SETUP	Hansen's panel threshold model. http://gretl.sourceforge.net/
	Gretl	*SETAR*		Estimation of a SETAR model. http://gretl.sourceforge.net/
Section 2.3.15. Forecasting with DSGE models	R	*BMR*	forecast	Bayesian Macroeconometrics in R (BMR) is a package for estimating and forecasting Bayesian VAR and DSGE. https://github.com/kthohr/BMR
	Mat-lab/GNU Octave	*Dynare*		Software platform for solving, estimating, and making forecasts with DSGE. https://www.dynare.org/
Section 2.3.18. Innovation diffusion models	R	*DIMORA*		Estimation of Bass Model, Generalised Bass Model, GGM, UCRCD. https://cran.r-project.org/package=DIMORA
	R	*diffusion*	diffusion	Various diffusion models to forecast new product growth. Currently the package contains Bass, Gompertz and Gamma/Shifted Gompertz curves. https://cran.r-project.org/package=diffusion
Section 2.3.19. The natural law of growth in competition	R	*LS2Wstat*	scurve	An S curve function between two constant values. https://cran.r-project.org/package=LS2Wstat
Section 2.3.21. Estimation and representation of uncertainty	R	*hdrcde*	cde	Conditional kernel density estimation to produce marginal distributions (uncertainty forecasts). https://cran.r-project.org/package=hdrcde
	R	*gamlss*	gamlss	Semi-parametric models for uncertainty forecasting. https://cran.r-project.org/package=gamlss

(*continued on next page*)

Table B.1 (*continued*).

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
	R	*gamboostLSS*	mboostLSS	Semi-parametric component-wise gradient boosting models for uncertainty forecasting. https://cran.r-project.org/package=gamboostLSS
	Python	*scikit-learn*	GradientBoostingRegressor; RandomForestQuantileRegressor	Machine learning models (gradient boosting trees and random forests) for quantile forecasting. https://scikit-learn.org/stable/
	R	*quantreg*	rq; lprq; nlqr	Estimation and inference methods for models of conditional quantiles. https://cran.r-project.org/package=quantreg
	R	*EnvStats*	FcnsByCatPredInts; pointwise	Functions for computing prediction intervals and simultaneous prediction intervals. https://cran.r-project.org/package=EnvStats
	R	*rmgarch*	dccfit-methods; dccforecast-methods	Multivariate GARCH Models (e.g., forecasting covariance matrix). https://cran.r-project.org/package=rmgarch
Section 2.3.22. Forecasting under fat tails	R	*FatTailsR*		Functions for Kiener distributions and fat tails. https://cran.r-project.org/package=FatTailsR
Section 2.4.3. Bayesian forecasting with copulas	R	*VineCopula*		Statistical analysis of vine copula models. https://cran.r-project.org/package=VineCopula
	R	*cdcopula*		Covariate-dependent copula models. https://github.com/feng-li/cdcopula
	R	*FactorCopula*		Factor Copula Models for Mixed Continuous and Discrete Data. https://cran.r-project.org/package=FactorCopula
Section 2.5.1. Leading indicators and Granger causality	R	*lmtest*	grangertest	Test for Granger causality. https://cran.r-project.org/package=lmtest
	Gretl		var; omit	Standard Granger-causality test. http://gretl.sourceforge.net/
	Gretl	*BreitungCandelonTest*		Breitung-Candelon test of frequency-wise Granger (non-)causality. http://gretl.sourceforge.net/
Section 2.5.3. Variable Selection	R	*glmnet*		Generalised linear models with Lasso or elastic net regularisation. https://cran.r-project.org/package=glmnet
	Gretl		omit	Sequential removing of variables to a model. http://gretl.sourceforge.net/
	Gretl	*addlist*		Sequential addition of variables to a model. http://gretl.sourceforge.net/
	Gretl	*regls*		Add-on for regularised least squares such as Ridge, Lasso and Elastic-Net. http://gretl.sourceforge.net/
	Gretl	*fsboost*	fsreg	Forward-stagewise boosted regression estimates with functionality for forecasting. http://gretl.sourceforge.net/
Section 2.5.4. Model Selection	R	*gets*		Functions for automatic general to specific model selection. https://cran.r-project.org/package=gets
Section 2.5.5. Cross-validation for time-series data	R	*forecast*	CVar	k-fold Cross-Validation applied to an autoregressive model. https://cran.r-project.org/package=forecast
	Gretl		fcast	Forecasting command with functionality for recursive-window forecasts. http://gretl.sourceforge.net/

(*continued on next page*)

Table B.1 (*continued*).

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
Section 2.6.1. Forecast combination: a brief review of statistical approaches	R	*forecastHybrid*		Functions for ensemble time series forecasts. https://cran.r-project.org/package=forecastHybrid
Section 2.7.2. Forecasting on distributed systems	Database	*InfluxDB*		Scalable datastore for metrics, events, and real-time analytics. https://www.influxdata.com/time-series-database/
	Database	*OpenTSDB*		A scalable, distributed Time Series Database. http://opentsdb.net/
	Database	*RRDtool*		A program for easily maintaining a database of time-series data. https://oss.oetiker.ch/rrdtool/
	Database	*Timely*		A time series database application that provides secure access to time series data. https://code.nsa.gov/timely/
	Python, R & Spark	*darima*		Implementations of distributed ARIMA models on Spark platform. https://github.com/xqnwang/darima
Section 2.7.3. Agent-based models	R	*SpaDES*		Spatially explicit discrete event simulation models. https://cran.r-project.org/package=SpaDES
Section 2.7.4. Feature-based time series forecasting	R	*tsfeatures*	tsfeatures	Methods for extracting various features from time series data. https://cran.r-project.org/package=tsfeatures
	Python	*tsfresh*		Calculates a large number of time series characteristics, the so called features. Further the package contains methods to evaluate the explaining power and importance of such characteristics for regression or classification tasks. https://tsfresh.readthedocs.io/en/latest/
	Matlab	*hctsa*		Code framework that enables the extraction of thousands of time-series features from a time series (or a time-series dataset). It also provides a range of tools for visualising and analysing the resulting time-series feature matrix. https://github.com/benfulcher/hctsa
	Python	*pyopy*		Python binding for hctsa. https://github.com/strawlab/pyopy
	R	*fforma*		Tools for forecasting using a model combination approach. It can be used for model averaging or model selection. It works by training a 'classifier' that learns to select/combine different forecast models. https://github.com/pmontman/fforma
	R	*gratis*		Efficient algorithms for generating time series with diverse and controllable characteristics, which can be used as the training data in feature-based time series forecasting. https://cran.r-project.org/package=gratis
	R	*seer*		Implementations of a novel framework for forecast model selection using time series features; FFORMS (Feature-based FORecast Model Selection). https://cran.r-project.org/package=seer
Section 2.7.5. Forecasting with bootstrap	R	*bootstrap*		Various bootstrapping functions. https://cran.r-project.org/package=bootstrap
	Gretl	*uniFCextensions*	uniFCboot	Estimate an interval forecast without assuming Gaussian innovations. http://gretl.sourceforge.net/
Section 2.7.6. Bagging for time series forecasting	R	*forecast*	baggedETS	Returns forecasts and other information for bagged ETS models. https://cran.r-project.org/package=forecast
	R	*tshacks*	baggedClusterETS; treatedETS	Returns forecasts for bagged Cluster and Treated ETS models. https://github.com/tiagomendesdantas/tshacks
Section 2.7.8. Neural Networks	Python, MXNet & PyTorch	*GluonTS*		Framework for building deep learning based models including a number of pre-built models such as feed-forward neural networks. https://github.com/awslabs/gluon-ts
	R	*nnfor*	mlp; elm	Time Series Forecasting with Neural Networks. https://cran.r-project.org/package=nnfor
	Python	*neural prophet*		Reimplementation of prophet in PyTorch, and extensions to it. https://github.com/ourownstory/neural_prophet
	Python			RNNs for forecasting in Tensorflow. https://github.com/HansikaPH/time-series-forecasting

(*continued on next page*)

Table B.1 (*continued*).

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
	R	*ANN2*	neuralnetwork	Artificial Neural Networks. https://cran.r-project.org/package=ANN2
	R	*nnet*	nnet	Feed-Forward Neural Networks and Multinomial Log-Linear Models. https://cran.r-project.org/package=nnet
	R	*forecast*	nnetar	Auto-regressive Neural Network for time series forecasting. https://cran.r-project.org/package=forecast
Section 2.7.9. Deep Probabilistic models	Python, MXNet & PyTorch	*GluonTS*		Framework for building deep learning based models including a number of pre-built models such as DeepAR, DeepState models and NBEATS. https://github.com/awslabs/gluon-ts
	Python, PyTorch	*PyTorchTS*		Clone of GluonTS in PyTorch. https://github.com/zalandoresearch/pytorch-ts
Section 2.7.10. Machine learning	R	*RSNNS*	mlp; rbf; dlvq; elman; jordan; som	Neural Networks using the Stuttgart Neural Network Simulator (SNNS). https://cran.r-project.org/package=RSNNS
	R	*rpart*	rpart; prune	Recursive partitioning and regression trees. https://cran.r-project.org/package=rpart
	R	*caret*		Classification and regression training. https://cran.r-project.org/package=caret
	R	*e1071*	svm	Misc ML functions of the Department of Statistics, Probability Theory Group. https://cran.r-project.org/package=e1071
	R	*kernlab*	gausspr	Gaussian processes for regression and classification. https://cran.r-project.org/package=kernlab
	R	*brnn*	brnn	Bayesian Regularisation for Feed-Forward Neural Networks. https://cran.r-project.org/package=brnn
	R	*grnn*	grnn	General regression neural network. https://cran.r-project.org/package=grnn
	R	*randomForest*	randomForest	Breiman and Cutler's Random Forests for Classification and Regression. https://cran.r-project.org/package=randomForest
	R	*gbm*	gbm	Generalised Boosted regression models. https://cran.r-project.org/package=gbm
	R	*neuralnet*	neuralnet	Training of simple Neural Networks. https://cran.r-project.org/package=neuralnet
	Python	*Tensorflow*		A framework, developed by Google, offering tools for designing, building, and deploying ML models. https://tensorflow.org/api_docs/python/tf
	Python	*Keras API*		A deep learning API built on top of Tensorflow. It provides high level blocks for building and training NN models. https://keras.io/
	R	*Tensorflow*		R Interface to Tensorflow. https://tensorflow.rstudio.com/
	R	*deepnet*		Deep learning toolkit. https://cran.r-project.org/package=deepnet
	R	*h2o*		R Interface for the 'H2O' Scalable Machine Learning Platform. https://cran.r-project.org/package=h2o
	R	*Apache MXNet*		A flexible library for deep learning. https://mxnet.apache.org/
	Python	*scikit-learn*		Ordinary Least Squares, Ridge regression, Lasso, Bayesian Regression, Generalized Linear Regression, Stochastic Gradient Descent and Polynomial regression, Support Vector Machines, Nearest Neighbors, Gaussian Processes, Decision Trees, Ensemble methods (Forests of randomised trees, AdaBoost and Gradient Tree Boosting), Multi-layer Perceptrons. https://scikit-learn.org/stable/
	Python	*CNTK*		A framework, developed by Microsoft, that provides tools for building ML and DL models. https://docs.microsoft.com/en-us/cognitive-toolkit/
	Python	*PyTorch*		A framework, developed by Facebook, for building ML and DL models. https://pytorch.org/

(*continued on next page*)

Table B.1 (*continued*).

Related section	Software	Package/ Library/ Toolbox	Function(s)	Comments
Section 2.7.12. Clustering-based forecasting	R	*tsfknn*	tsfknn	Time Series Forecasting Using Nearest Neighbours. https://cran.r-project.org/package=tsfknn
Section 2.7.13. Hybrid methods	Python	*ESRNN-GPU*		A GPU-enabled version of the hybrid model used by the winner of M4 competition. https://github.com/damitkwr/ESRNN-GPU
Section 2.8.1. Parametric methods for intermittent demand forecasting	R	*tsintermittent*	crost; tsb	Parametric forecasting methods for intermitternt demand. https://cran.r-project.org/package=tsintermittent
	R	*forecast*	croston	Forecasts for intermittent demand using Croston's method. https://cran.r-project.org/package=forecast
Section 2.8.2. Non-parametric intermittent demand methods	R	*tsintermittent*	imapa	MAPA for intermittent demand data. https://cran.r-project.org/package=tsintermittent
Section 2.8.3. Classification methods	R	*tsintermittent*	idclass	Time series categorisation for intermittent demand. https://cran.r-project.org/package=tsintermittent
	R	*tsutils*	abc; xyz; abcxyz	Classification functions and routines. https://cran.r-project.org/package=tsutils
Section 2.9.3. Forecasting with text information	Python	*NLTK*		The Natural Language Toolkit in Python. https://www.nltk.org/
	Python	*SpaCy*		An open source library for advanced Natural Language Processing in Python. https://spacy.io/
Section 2.10.1. Cross-sectional hierarchical forecasting	R	*hts*		Functions and routines for hierarchical and grouped time series forecasting. https://cran.r-project.org/package=hts
Section 2.10.2. Temporal aggregation	R	*MAPA*	mapa; mapasimple	Functions and wrappers for using the Multiple Aggregation Prediction Algorithm (MAPA) for time series forecasting. https://cran.r-project.org/package=MAPA
	R	*thief*	thief	Temporal Hierarchical Forecasting. https://cran.r-project.org/package=thief
	R	*tsintermittent*	imapa	MAPA for intermittent demand data with automatic model selection based on the PK classification. https://cran.r-project.org/package=tsintermittent
Section 2.10.4. Ecological inference forecasting	R	*ei*	ei	Returns local and global forecasts of inner cells in 2 × 2 tables. https://cran.r-project.org/package=ei
	R	*eiPack*	ei.MD.bayes; ei.reg; ei.reg.bayes	Returns local and global forecasts of inner cells in R × C tables under a Multinomial Dirichlet model or using ecological regression. https://cran.r-project.org/package=eiPack
	R	*lphom*	lphom; tslphom; nslphom	Returns forecasts of inner cells of a R × C table using linear programming optimisation. https://CRAN.R-project.org/package=lphom
	R	*eiCompare*	ei_est_gen; ei_good; ei_rxc	Returns forecasts of inner cells of a $R \times C$ tables using iterative versions of 2 × 2 methods and the Multinomial Dirichlet model. https://cran.r-project.org/package=eiCompare
Section 2.12.2. Point, interval, and pHDR forecast error measures	R	*forecast*	accuracy	Accuracy measures for a forecast model. https://cran.r-project.org/package=forecast
Section 2.12.4. Evaluating probabilistic forecasts	R	*scoringRules*		Scoring rules for parametric and simulated distribution forecasts. https://cran.r-project.org/package=scoringRules
	R	*verification*	crps	Continuous ranked probability score. https://cran.r-project.org/package=verification
Section 2.12.6. Statistical tests of forecast performance	R	*forecast*	dm.test	Diebold–Mariano test for predictive accuracy. https://cran.r-project.org/package=forecast
	R	*tsutils*	nemenyi	Nonparametric multiple comparisons (Nemenyi test). https://cran.r-project.org/package=tsutils
	Gretl	*FEP*	doMZtest; doHPtest; doEKTtest; doPTtest; doDLtest; doDMtest; doGWtest; doCWtest	Various statistical tests on forecast unbiasedness, efficiency, asymmetric loss and directional changes. http://gretl.sourceforge.net/
	Gretl	*DiebMar*		Diebold–Mariano test. http://gretl.sourceforge.net/

Table C.2
A list of indicative publicly available data sets.

Related section	Description	Link
Section 2.7.9. Deep Probabilistic Models	Data for wikipedia page views, Dominicks retail, electricity consumption, traffic lane occupation.	https://gluon-ts.mxnet.io/api/gluonts/gluonts.dataset.repository.datasets.html
Section 2.9.3. Forecasting with text Information	Movie reviews data provided by the Stanford NLP group.	https://nlp.stanford.edu/sentiment/code.html
Section 2.10.4. Ecological inference forecasting	Party registration in South-East North Carolina (eiPack R package).	https://www2.ncleg.net/RnR/Redistricting/BaseData2001
	ei.Datasets: Real Datasets for Assessing Ecological Inference Algorithms.	https://cran.csiro.au/web/packages/ei.Datasets/index.html
Section 2.12.7. Forecasting competitions	Data for the M, M2, M3 and M4 forecasting competitions.	https://forecasters.org/resources/time-series-data/
	Time Series Competition Data (R package)	https://github.com/robjhyndman/tscompdata
	Mcomp: Data for the M and M3 forecasting competitions (R package).	https://cran.r-project.org/package=Mcomp
	M4comp2018: Data for the M4 forecasting competition (R package).	https://github.com/carlanetto/M4comp2018
	Data for the M4 forecasting competition (csv files).	https://github.com/Mcompetitions/M4-methods/tree/master/Dataset
	Tcomp: Data from the 2010 Tourism forecasting competition (R package)	https://cran.r-project.org/package=Tcomp
	Data for the M5 forecasting competition (csv files).	https://github.com/Mcompetitions/M5-methods/tree/master/Dataset
Section 3.2.3. Forecasting for inventories	Grupo Bimbo Inventory Demand.	https://www.kaggle.com/c/grupo-bimbo-inventory-demand
Section 3.2.4. Forecasting in retail	Rossmann Store Sales.	https://www.kaggle.com/c/rossmann-store-sales
	Corporación Favorita Grocery Sales Forecasting.	https://www.kaggle.com/c/favorita-grocery-sales-forecasting
	Walmart Recruiting – Store Sales Forecasting.	https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting
	Walmart Recruiting II: Sales in Stormy Weather.	https://www.kaggle.com/c/walmart-recruiting-sales-in-stormy-weather
	Store Item Demand Forecasting Challenge.	https://www.kaggle.com/c/demand-forecasting-kernels-only
	Online Product Sales.	https://www.kaggle.com/c/online-sales
Section 3.2.8. Predictive maintenance	Robot Execution Failures.	https://archive.ics.uci.edu/ml/datasets/Robot+Execution+Failures
	Gearbox Fault Detection.	https://c3.nasa.gov/dashlink/resources/997/
	Air Pressure System Failure at Scania Trucks.	https://archive.ics.uci.edu/ml/datasets/IDA2016Challenge
	Generic, Scalable and Decentralised Fault Detection for Robot Swarms	https://zenodo.org/record/831471#.WwQIPUgvxPY
	Wind turbine data (e.g., failures).	https://opendata.edp.com/pages/homepage/
Section 3.3. Economics and finance	Two Sigma Financial Modelling Challenge.	https://www.kaggle.com/c/two-sigma-financial-modeling/overview
	Financial, economic, and alternative data sets, serving investment professionals.	https://www.quandl.com/
Section 3.3.2. Forecasting GDP and Inflation	Repository website with Dynare codes and data sets to estimate different DSGE models and use them for forecasting.	https://github.com/johannespfeifer/dsge_mod
	Data set for Macroeconomic variables for US economy.	https://fred.stlouisfed.org/
	Data set for Macroeconomic variables for OECD economy.	https://data.oecd.org/
	Data set for Macroeconomic variables for EU economy.	https://ec.europa.eu/eurostat/data/database
Section 3.3.7. House price forecasting	Zillow Prize: Zillow's Home Value Prediction (Zestimate).	https://www.kaggle.com/c/zillow-prize-1
	Sberbank Russian Housing Market.	https://www.kaggle.com/c/sberbank-russian-housing-market

(*continued on next page*)

Appendix B. Software

See Table B.1.

Appendix C. Data sets

See Table C.2.

Table C.2 (*continued*).

Related section	Description	Link
	Western Australia Rental Prices.	https://www.kaggle.com/c/deloitte-western-australia-rental-prices
Section 3.3.12. Forecasting returns to investment style	Algorithmic Trading Challenge.	https://www.kaggle.com/c/AlgorithmicTradingChallenge
Section 3.3.13. Forecasting stock returns	The Winton Stock Market Challenge.	https://www.kaggle.com/c/the-winton-stock-market-challenge
	The Big Data Combine Engineered by BattleFin.	https://www.kaggle.com/c/battlefin-s-big-data-combine-forecasting-challenge/data
Section 3.4. Energy	VSB Power Line Fault Detection.	https://www.kaggle.com/c/vsb-power-line-fault-detection
	ASHRAE – Great Energy Predictor III	https://www.kaggle.com/c/ashrae-energy-prediction
Section 3.4.3. Hybrid machine learning system for short-term load forecasting	Global Energy Forecasting Competition 2012 – Load Forecasting.	https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting
Section 3.4.6. Wind power forecasting	Global Energy Forecasting Competition 2012 – Wind Forecasting.	https://www.kaggle.com/c/GEF2012-wind-forecasting
Section 3.4.8. Solar power forecasting	Power measurements from a PV power plant and grid of numerical weather predictions.	https://doi.org/10.25747/edf8-m258
	AMS 2013–2014 Solar Energy Prediction Contest.	https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest
	SolarTechLab data set.	https://ieee-dataport.org/open-access/photovoltaic-power-and-weather-parameters
Section 3.4.9. Long-term simulation for large electrical power systems	Brazilian National Electric Systems Operator (hydro, solar, wind, nuclear and thermal generation data).	http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx
Section 3.4.10. Collaborative forecasting in the energy sector	Solar power time series from 44 small-scale PV in Évora, Portugal.	https://doi.org/10.25747/gywm-9457
	Australian Electricity Market Operator (AEMO) 5 Minute Wind Power Data.	https://doi.org/10.15129/9e1d9b96-baa7-4f05-93bd-99c5ae50b141
	Electrical energy consumption data from domestic consumers.	
	Electric vehicles charging data (arrivals, departures, current, voltage, etc.).	https://eatechnology.com/consultancy-insights/my-electric-avenue/
	Wind power plant data and numerical weather predictions from CNR (France).	https://challengedata.ens.fr/challenges/34
Section 3.5.2. Weather forecasting	How Much Did It Rain?	https://www.kaggle.com/c/how-much-did-it-rain-ii
Section 3.5.3. Air quality forecasting	EMC Data Science Global Hackathon (Air Quality Prediction).	https://www.kaggle.com/c/dsg-hackathon/overview
Section 3.6. Social good and demographic forecasting	LANL Earthquake Prediction.	https://www.kaggle.com/c/LANL-Earthquake-Prediction
Section 3.6.1. Healthcare	Flu Forecasting.	https://www.kaggle.com/c/genentech-flu-forecasting
	West Nile Virus Prediction.	https://www.kaggle.com/c/predict-west-nile-virus
Section 3.6.2. Epidemics and pandemics	COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.	https://github.com/CSSEGISandData/COVID-19
Section 3.6.3. Forecasting mortality	Human Mortality Database.	https://www.mortality.org
	EuroMOMO.	https://www.euromomo.eu/
	The Economist.	https://github.com/TheEconomist/covid-19-excess-deaths-tracker
	The New York Times.	https://github.com/Financial-Times/coronavirus-excess-mortality-data
	The Financial Times.	https://github.com/nytimes/covid-19-data/tree/master/excess-deaths
	ANACONDA- Quality assessment of mortality data.	https://crvsgateway.info/anaconda
	Australian Human Mortality Database.	https://demography.cass.anu.edu.au/research/australian-human-mortality-database

(*continued on next page*)

Table C.2 (*continued*).

Related section	Description	Link
	Canadian Human Mortality Database.	http://www.bdlc.umontreal.ca/CHMD/
	French Human Mortality Database.	https://frdata.org/fr/french-human-mortality-database
Section 3.6.4. Forecasting fertility	Human Fertility Database: fertility data for developed countries with complete birth registration based on official vital statistics.	https://www.humanfertility.org/
	World Fertility Data: UN's collection of fertility data based on additional data sources such as surveys.	https://www.un.org/development/desa/pd/data/world-fertility-data
Section 3.6.5. Forecasting migration	Integrated Modelling of European Migration (IMEM) Database, with estimates of migration flows between 31 European countries by origin, destination, age and sex, for 2002–2008.	https://www.imem.cpc.ac.uk/
	QuantMig data inventory: meta-inventory on different sources of data on migration and its drivers, with European focus.	https://quantmig.eu/data_inventory/
	Bilateral international migration flow estimates for 200 countries (Abel & Cohen, 2019).	https://doi.org/10.1038/s41597-019-0089-3
	UN World Population Prospects: UN global population estimates and projections, including probabilistic	https://population.un.org/wpp/
Section 3.8. Other applications	Forecast Eurovision Voting.	https://www.kaggle.com/c/Eurovision2010
	Reducing Commercial Aviation Fatalities.	https://www.kaggle.com/c/reducing-commercial-aviation-fatalities
	Porto Seguro's Safe Driver Prediction.	https://www.kaggle.com/c/porto-seguro-safe-driver-prediction
	Recruit Restaurant Visitor Forecasting.	https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting
	Restaurant Revenue Prediction.	https://www.kaggle.com/c/restaurant-revenue-prediction
	Coupon Purchase Prediction.	https://www.kaggle.com/c/coupon-purchase-prediction
	Bike Sharing Demand	https://www.kaggle.com/c/bike-sharing-demand
	Google Analytics Customer Revenue Prediction.	https://www.kaggle.com/c/ga-customer-revenue-prediction
	Santander Value Prediction Challenge.	https://www.kaggle.com/c/santander-value-prediction-challenge
	Santander Customer Transaction Predictio.n	https://www.kaggle.com/c/santander-customer-transaction-prediction
	Acquire Valued Shoppers Challenge	https://www.kaggle.com/c/acquire-valued-shoppers-challenge
	Risky Business	https://www.kaggle.com/c/risky-business
	Web Traffic Time Series Forecasting.	https://www.kaggle.com/c/web-traffic-time-series-forecasting
	A repository of data sets, including time series ones, that can be used for benchmarking forecasting methods in various applications of interest.	https://github.com/awesomedata/awesome-public-datasets
	WSDM – KKBox's Churn Prediction Challenge.	https://www.kaggle.com/c/kkbox-churn-prediction-challenge/overview
	Homesite Quote Conversion	https://www.kaggle.com/c/homesite-quote-conversion
	Liberty Mutual Group: Property Inspection Prediction.	https://www.kaggle.com/c/liberty-mutual-group-property-inspection-prediction
	Liberty Mutual Group – Fire Peril Loss Cost.	https://www.kaggle.com/c/liberty-mutual-fire-peril
	A set of more than 490,000 time series (micro, macro, demographic, finance, other) to download.	http://fsudataset.com/
	A repository of data sets, including time series ones, that can be used for benchmarking forecasting methods in various applications of interest.	https://github.com/awesomedata/awesome-public-datasets

(*continued on next page*)

Table C.2 (*continued*).

Related section	Description	Link
Section 3.8.1. Tourism demand forecasting	TourMIS: Database with annual and monthly tourism time series (e.g., arrivals and bednights) covering European countries, regions, and cities (free registration required).	https://www.tourmis.info/index_e.html
	Tourism Forecasting.	https://www.kaggle.com/c/tourism1; https://www.kaggle.com/c/tourism2
Section 3.8.2. Forecasting for aviation	Airline and Airport performance data provided by the U.S. Department of Transportation	https://www.transtats.bts.gov/
Section 3.8.3. Traffic flow forecasting	New York City Taxi Fare Prediction.	https://www.kaggle.com/c/demand-forecasting-kernels-only
	RTA Freeway Travel Time Prediction	https://www.kaggle.com/c/RTA
	ECML/PKDD 15: Taxi Trip Time Prediction.	https://www.kaggle.com/c/pkdd-15-taxi-trip-time-prediction-ii
	BigQuery-Geotab Intersection Congestion.	https://www.kaggle.com/c/bigquery-geotab-intersection-congestion/overview
	Traffic volume counts collected by DOT for New York Metropolitan Transportation Council.	https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts-2014--2019-/ertz-hr4r
Section 3.8.5. Elections forecasting	New Zealand General Elections – Official results and statistics.	https://www.electionresults.govt.nz
	Spanish Elections – Official results and statistics.	https://dataverse.harvard.edu/dataverse/SEA
Section 3.8.6. Sports forecasting	NFL Big Data Bowl.	https://www.kaggle.com/c/nfl-big-data-bowl-2020
Section 3.8.9. Forecasting under data integrity attacks	Microsoft Malware Prediction.	https://www.kaggle.com/c/microsoft-malware-prediction

References

Aaltonen, K. (2011). Project stakeholder analysis as an environmental interpretation process. *International Journal of Project Management*, *29*(2), 165–183.

Aastveit, K. A., Mitchell, J., Ravazzolo, F., & van Dijk, H. K. (2019). *The evolution of forecast density combinations in economics*. Oxford University Press.

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., et al. (2016). Deep learning with differential privacy. In *Proc. of the 2016 ACM SIGSAC conference on computer and communications security (ACM CCS)* (pp. 308–318).

Abel, G. J. (2018). Non-zero trajectories for long-run net migration assumptions in global population projection models. *Demographic Research*, *38*(54), 1635–1662.

Abel, G. J., & Cohen, J. E. (2019). Bilateral international migration flow estimates for 200 countries. *Scientific Data*, *6*(1), 82.

Abouarghoub, W., Nomikos, N. K., & Petropoulos, F. (2018). On reconciling macro and micro energy transport forecasts for strategic decision making in the tanker industry. *Transportation Research Part E: Logistics and Transportation Review*, *113*, 225–238.

AbouZahr, C., de Savigny, D., Mikkelsen, L., Setel, P. W., Lozano, R., & Lopez, A. D. (2015). Towards universal civil registration and vital statistics systems: the time is now. *The Lancet*, *386*(10001), 1407–1418.

Abraham, B., & Box, G. E. P. (1979). Bayeslan analysis of some outlier problems in time series. *Biometrika*, *66*(2), 229–236.

Abraham, B., & Chuang, A. (1989). Outlier detection and time series modeling. *Technometrics*, *31*(2), 241–248.

Abramson, B., & Finizza, A. (1991). Using belief networks to forecast oil prices. *International Journal of Forecasting*, *7*(3), 299–315.

Abramson, B., & Finizza, A. (1995). Probabilistic forecasts from probabilistic models: a case study in the oil market. *International Journal of Forecasting*, *11*(1), 63–72.

Abramson, G., & Zanette, D. (1998). Statistics of extinction and survival in Lotka–Volterra systems. *Physical Review E*, *57*, 4572–4577.

Achen, C. H., & Phillips Shively, W. (1995). *Cross-level inference*. University of Chicago Press.

Acquaviva, A., Apiletti, D., Attanasio, A., Baralis, E., Castagnetti, F. B., Cerquitelli, T., et al. (2015). Enhancing energy awareness through the analysis of thermal energy consumption. In *EDBT/ICDT workshops* (pp. 64–71).

Adams, W., & Michael, V. (1987). Short-term forecasting of passenger demand and some application in quantas. In *AGIFORS symposium proc, vol. 27*.

Afanasyev, D. O., & Fedorova, E. A. (2019). On the impact of outlier filtering on the electricity price forecasting accuracy. *Applied Energy*, *236*, 196–210.

Agarwal, A., Dahleh, M., & Sarkar, T. (2019). A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM conference on economics and computation* (pp. 701–726).

Aggarwal, C., & Zhai, C. (2012). *Mining text data*. Springer Science & Business Media.

Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, *31*(2), 316–321.

Ahlburg, D. A., & Vaupel, J. W. (1990). Alternative projections of the U.S. population. *Demography*, *27*(4), 639–652.

Ahmad, A., Hassan, M., Abdullah, M., Rahman, H., Hussin, F., Abdullah, H., et al. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, *33*, 102–109.

Ahmad, A., Javaid, N., Mateen, A., Awais, M., & Khan, Z. A. (2019). Short-term load forecasting in smart grids: An intelligent modular approach. *Energies*, *12*(1), 164.

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, *147*, 77–89.

Aït-Sahalia, Y., Cacho-Diaz, J., & Laeven, R. J. A. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, *117*(3), 585–606.

Aizenman, J., & Jinjarak, Y. (2013). Real estate valuation, current account and credit growth patterns, before and after the 2008-9 crisis. *Tech. Rep. 19190*, National Bureau of Economic Research.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csáki (Eds.), *Proceedings of the second international symposium on information theory* (pp. 267–281). Budapest: Csáki.

Akouemo, H. N., & Povinelli, R. J. (2016). Probabilistic anomaly detection in natural gas time series data. *International Journal of Forecasting*, *32*(3), 948–956.

Akram, F., Binning, A., & Maih, J. (2015). Joint prediction bands for macroeconomic risk management. *Tech. Rep. No. 5/2016*, Centre for Applied Macro- and Petroleum economics (CAMP) Working Paper Series.

Aksin, Z., Armony, M., & Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, *16*(6), 665–688.

Aktekin, T., & Soyer, R. (2011). Call center arrival modeling: A Bayesian state-space approach. *Naval Research Logistics*, *58*(1), 28–42.

Al-Azzani, M. A., Davari, S., & England, T. J. (2020). An empirical investigation of forecasting methods for ambulance calls-a case study. *Health Systems*, 1–18.

Al-Homoud, M. S. (2001). Computer-aided building energy analysis techniques. *Building and Environment*, *36*(4), 421–433.

Albon, C. (2018). *Python machine learning cookbook*. O'Reilly UK Ltd..

Albulescu, C. T., Tiwari, A. K., & Ji, Q. (2020). Copula-based local dependence among energy, agriculture and metal commodities markets. *Energy*, *202*, Article 117762.

Aldor-Noiman, S., Feigin, P. D., & Mandelbaum, A. (2009). Workload forecasting for a call center: Methodology and a case study. *The Annals of Applied Statistics*, *3*(4), 1403–1447.

Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., et al. (2019). GluonTS: Probabilistic time series models in python. *Journal of Machine Learning Research*.

Alho, J. M., Hougaard Jensen, S. E., & Lassila, J. (Eds.), (2008). *Uncertain demographics and fiscal sustainability*. Cambridge University Press.

Alho, J. M., & Spencer, B. D. (1985). Uncertain population forecasting. *Journal of the American Statistical Association*, *80*(390), 306–314.

Alho, J. M., & Spencer, B. D. (2005). *Statistical demography and forecasting*. New York: Springer.

Ali, M. M., & Boylan, J. E. (2011). Feasibility principles for downstream demand inference in supply chains. *Journal of the Operational Research Society*, *62*(3), 474–482.

Ali, M. M., Boylan, J. E., & Syntetos, A. A. (2012). Forecast errors and inventory performance under forecast information sharing. *International Journal of Forecasting*, *28*(4), 830–841.

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*(3), 1047–1091.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T., et al. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography*, *48*(3), 815–839.

Almeida, C., & Czado, C. (2012). Efficient Bayesian inference for stochastic time-varying copula models. *Computational Statistics & Data Analysis*, *56*(6), 1511–1527.

Aloui, R., Hammoudeh, S., & Nguyen, D. K. (2013). A time-varying copula approach to oil and stock market dependence: The case of transition economies. *Energy Economics*, *39*, 208–221.

Alquist, R., Bhattarai, S., & Coibion, O. (2020). Commodity-price comovement and global economic activity. *Journal of Monetary Economics*, *112*, 41–56.

Alquist, R., Kilian, L., & Vigfusson, R. J. (2013). Forecasting the price of oil. *vol. 2*, In *Handbook of economic forecasting* (pp. 427–507). Elsevier.

Alvarado-Valencia, J. A., & Barrero, L. H. (2014). Reliance, trust and heuristics in judgmental forecasting. *Computers in Human Behavior*, *36*, 102–113.

Alvarado-Valencia, J., Barrero, L. H., Önkal, D., & Dennerlein, J. T. (2017). Expertise, credibility of system forecasts and integration methods in judgmental demand forecasting. *International Journal of Forecasting*, *33*(1), 298–313.

Alvarez-Ramirez, J., Soriano, A., Cisneros, M., & Suarez, R. (2003). Symmetry/anti-symmetry phase transitions in crude oil markets. *Physica A: Statistical Mechanics and Its Applications*, *322*, 583–596.

Amarasinghe, A., Wichmann, O., Margolis, H. S., & Mahoney, R. T. (2010). Forecasting dengue vaccine demand in disease endemic and non-endemic countries. *Human Vaccines*, *6*(9), 745–753.

Amendola, A., Niglio, M., & Vitale, C. (2006). The moments of SETARMA models. *Statistics & Probability Letters*, *76*(6), 625–633.

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, *25*(2), 177–190.

An, S., & Schorfheide, F. (2007). Bayesian analysis of DSGE models. *Econometric Reviews*, *26*(2–4), 113–172.

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, *9*, 1518–1530.

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Englewood Cliffs, NJ: Prentice-Hall.

Anderson, V. O., & Nochmals, U. (1914). The elimination of spurious correlation due to position in time or space. *Biometrika*, *10*(2/3), 269–279.

Andersson, E., Kühlmann-Berenzon, S., Linde, A., Schiöler, L., Rubinova, S., & Frisén, M. (2008). Predictions by early indicators of the time and height of the peaks of yearly influenza outbreaks in Sweden. *Scandinavian Journal of Public Health*, *36*(5), 475–482.

Andrade, J., Filipe, J., Reis, M., & Bessa, R. (2017). Probabilistic price forecasting for day-ahead and intraday markets: Beyond the statistical model. *Sustainability*, *9*(1990), 1–29.

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, *27*(3), 870–886.

Andrés, M. A., Peña, D., & Romo, J. (2002). Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference*, *100*(1), 1–11.

Andrews, B. H., & Cunningham, S. M. (1995). LL Bean improves call-center forecasting. *Interfaces*, *25*(6), 1–13.

Andrews, R. L., Currim, I. S., Leeflang, P., & Lim, J. (2008). Estimating the SCAN* PRO model of store sales: HB, FM or just OLS? *International Journal of Research in Marketing*, *25*(1), 22–33.

Andrieu, C., Doucet, A., & Holenstein, R. (2011). Particle Markov chain Monte Carlo. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *72*(2), 269–342.

Andrieu, C., & Roberts, G. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, *37*(2), 697–725.

Aneiros-Pérez, G., & Vieu, P. (2008). Nonparametric time series prediction: A semi-functional partial linear modeling. *Journal of Multivariate Analysis*, *99*(5), 834–857.

Ang, A., & Bekaert, G. (2002). Short rate nonlinearities and regime switches. *Journal of Economic Dynamics & Control*, *26*(7), 1243–1274.

Ang, A., Bekaert, G., & Wei, M. (2008). The term structure of real rates and expected inflation. *The Journal of Finance*, *63*(2), 797–849.

Angelini, G., & De Angelis, L. (2019). Efficiency of online football betting markets. *International Journal of Forecasting*, *35*(2), 712–721.

Angus, J. E. (1992). Asymptotic theory for bootstrapping the extremes. *Communications in Statistics. Theory and Methods*, *22*(1), 15–30.

Anselin, L., & Tam Cho, W. K. (2002). Spatial effects and ecological inference. *Political Analysis*, *10*(3), 276–297.

Antipov, A., & Meade, N. (2002). Forecasting call frequency at a financial services call centre. *Journal of the Operational Research Society*, *53*(9), 953–960.

Antonakakis, N., Chatziantoniou, I., Floros, C., & Gabauer, D. (2018). The dynamic connectedness of U.K. regional property returns. *Urban Studies*, *55*(14), 3110–3134.

Apiletti, D., Baralis, E., Cerquitelli, T., Garza, P., Michiardi, P., & Pulvirenti, F. (2015). Pampa-HD: A parallel MapReduce-based frequent pattern miner for high-dimensional data. In *2015 IEEE international conference on data mining workshop (ICDMW)* (pp. 839–846). IEEE.

Apiletti, D., Baralis, E., Cerquitelli, T., Garza, P., Pulvirenti, F., & Michiardi, P. (2017). A parallel mapreduce algorithm to efficiently support itemset mining on high dimensional data. *Big Data Research*, *10*, 53–69.

Apiletti, D., & Pastor, E. (2020). Correlating espresso quality with coffee-machine parameters by means of association rule mining. *Electronics*, *9*(1), 100.

Apiletti, D., Pastor, E., Callà, R., & Baralis, E. (2020). Evaluating espresso coffee quality by means of time-series feature engineering. In *EDBT/ICDT workshops*.

Archak, N., Ghose, A., & Ipeirotis, P. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509.

Arinze, B. (1994). Selecting appropriate forecasting models using rule induction. *Omega*, *22*(6), 647–658.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79.

Armstrong, J. S. (2001a). *Principles of forecasting: A handbook for researchers and practitioners*. Springer Science & Business Media.

Armstrong, J. S. (2001b). Combining forecasts. In *International series in operations research & management science*, *Principles of forecasting* (pp. 417–439). Springer, Boston, MA.

Armstrong, J. S. (2007). Significance tests harm progress in forecasting. *International Journal of Forecasting*, *23*(2), 321–327.

Armstrong, C. (2017). Omnichannel retailing and demand planning. *The Journal of Business Forecasting*, *35*(4), 10–15.

Armstrong, J. S., & Collopy, F. (1998). Integration of statistical methods and judgment for time series forecasting: Principles from empirical research. In G. Wright, & P. Goodwin (Eds.), *Forecasting with judgment* (pp. 269–293). New York: John Wiley & Sons Ltd.

Armstrong, J. S., & Green, K. C. (2018). Forecasting methods and principles: Evidence-based checklists. *Journal of Global Scholars of Marketing Science*, *28*(2), 103–159.

Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, *68*(8), 1717–1731.

Arnott, R. D., Beck, N., Kalesnik, V., & West, J. (2016). How can 'smart beta' go horribly wrong? SSRN:3040949.

Aron, J., & Muellbauer, J. (2020). Measuring excess mortality: the case of England during the Covid-19 pandemic. https://www.oxfordmartin.ox.ac.uk/publications/measuring-excess-mortality-the-case-of-england-during-the-covid-19-pandemic/, Accessed on 2020-08-20.

Arora, S., Taylor, J. W., & Mak, H.-Y. (2020). Probabilistic forecasting of patient waiting times in an emergency department. arXiv:2006.00335.

Arrhenius, S. A. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science (Fifth Series)*, *41*, 237–275.

Artis, M., & Marcellino, M. (2001). Fiscal forecasting: The track record of the IMF, OECD and EC. *The Econometrics Journal*, *4*(1), S20–S36.

Arvan, M., Fahimnia, B., Reisi, M., & Siemsen, E. (2019). Integrating human judgement into quantitative forecasting methods: A review. *Omega*, *86*, 237–252.

Asai, M. (2013). Heterogeneous asymmetric dynamic conditional correlation model with stock return and range. *Journal of Forecasting*, *32*(5), 469–480.

Asai, M., & Brugal, I. (2013). Forecasting volatility via stock return, range, trading volume and spillover effects: The case of Brazil. *North American Journal of Economics and Finance*, *25*, 202–213.

Asimakopoulos, S., & Dix, A. (2013). Forecasting support systems technologies-in-practice: A model of adoption and use for product forecasting. *International Journal of Forecasting*, *29*(2), 322–336.

Asimakopoulos, S., Paredes, J., & Warmedinger, T. (2020). Real-time fiscal forecasting using mixed-frequency data. *Scandinavian Journal of Economics*, *122*, 369–390.

Askanazi, R., Diebold, F. X., Schorfheide, F., & Shin, M. (2018). On the comparison of interval forecasts. *Journal of Time Series Analysis*, *39*(6), 953–965.

Asness, C. S. (2016). Invited editorial comment: the siren song of factor timing aka "smart beta timing" aka "style timing". *Journal of Portfolio Management*, *42*(5), 1–6.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, *16*(4), 521–530.

Assmus, G. (1984). New product forecasting. *Journal of Forecasting*, *3*(2), 121–138.

Athanasopoulos, G., Ahmed, R. A., & Hyndman, R. J. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, *25*(1), 146–166.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, *262*(1), 60–74.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*(3), 822–844.

Athanasopoulos, G., Song, H., & Sun, J. A. (2018). Bagging in tourism demand modeling and forecasting. *Journal of Travel Research*, *57*(1), 52–68.

Athey, S. (2018). The impact of machine learning on economics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An Agenda* (pp. 507–547). University of Chicago Press.

Atiya, A. F. (2020). Why does forecast combination work so well? *International Journal of Forecasting*, *36*(1), 197–200, URL https://www.sciencedirect.com/science/article/pii/S0169207019300779, M4 Competition.

Atiya, A. F., El-shoura, S. M., Shaheen, S. I., & El-sherif, M. S. (1999). A comparison between neural-network forecasting techniques–case study: river flow forecasting. *IEEE Transactions on Neural Networks*, *10*(2), 402–409.

Atkinson, A. C., Riani, M., & Corbellini, A. (2021). The Box–Cox transformation: Review and extensions. *Statistical Science*, *36*(2), 239–255.

Aue, A., Norinho, D. D., & Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, *110*(509), 378–392.

Austin, C., & Kusumoto, F. (2016). The application of big data in medicine: current implications and future directions. *Journal of Interventional Cardiac Electrophysiology*, *47*(1), 51–59.

Avramidis, A. N., Deslauriers, A., & L'Ecuyer, P. (2004). Modeling daily arrivals to a telephone call center. *Management Science*, *50*(7), 896–908.

Axelrod, R. (1997). Advancing the art of simulation in the social sciences. In *Simulating social phenomena* (pp. 21–40). Springer.

Ayton, P., Önkal, D., & McReynolds, L. (2011). Effects of ignorance and information on judgments and decisions. *Judgment and Decision Making*, *6*(5), 381–391.

Azose, J. J., & Raftery, A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, *52*(5), 1627–1650.

Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6460–6465.

Baade, R. A., & Matheson, V. A. (2016). Going for the gold: The economics of the olympics. *Journal of Economic Perspectives*, *30*(2), 201–218.

Baardman, L., Levin, I., Perakis, G., & Singhvi, D. (2018). Leveraging comparables for new product sales forecasting. *Production and Operations Management*, *27*(12), 2340–2343.

Babai, M. Z., Ali, M. M., & Nikolopoulos, K. (2012). Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *Omega*, *40*(6), 713–721.

Babai, M. Z., Dallery, Y., Boubaker, S., & Kalai, R. (2019). A new method to forecast intermittent demand in the presence of inventory obsolescence. *International Journal of Production Economics*, *209*, 30–41.

Babai, M. Z., Syntetos, A., & Teunter, R. (2014). Intermittent demand forecasting: An empirical study on accuracy and the risk of obsolescence. *International Journal of Production Economics*, *157*, 212–219.

Babai, M. Z., Tsadiras, A., & Papadopoulos, C. (2020). On the empirical performance of some new neural network methods for forecasting intermittent demand. *IMA Journal of Management Mathematics*, *31*(3), 281–305.

Babu, A., Levine, A., Ooi, Y. H., Pedersen, L. H., & Stamelos, E. (2020). Trends everywhere. *Journal of Investment Management*, *18*(1), 52–68.

Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, *40*(6), 722–737.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the seventh international conference on language resources and evaluation (LREC'10), vol. 10* (pp. 2200–2204).

Bacha, H., & Meyer, W. (1992). A neural network architecture for load forecasting. In *IJCNN international joint conference on neural networks, vol. 2* (pp. 442–447).

Baecke, P., De Baets, S., & Vanderheyden, K. (2017). Investigating the added value of integrating human judgement into statistical demand forecasting systems. *International Journal of Production Economics*, *191*, 85–96.

Baicker, K., Chandra, A., & Skinner, J. S. (2012). Saving money or just saving lives? Improving the productivity of US health care spending. *Annual Review of Economics*, *4*(1), 33–56.

Baillie, R. T., & Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, *1–2*(52), 91–113.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30.
Baker, J. (2021). Maximizing forecast value added through machine learning and nudges. *Foresight: The International Journal of Applied Forecasting*, *60*.
Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, *131*(4), 1593–1636.
Balbo, N., Billari, F. C., & Mills, M. (2013). Fertility in advanced societies: A review of research. *European Journal of Population*, *29*(1), 1–38.
Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics*, *11*(1), 81–92.
Balke, N. S., & Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, *38*(3), 627–645.
Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92.
Bańbura, M., Giannone, D., & Reichlin, L. (2011). Nowcasting (chapter 7). In M. P. Clements, & D. F. Hendry (Eds.), *The oxford handbook of economic forecasting*. Oxford University Press.
Bandara, K., Bergmeir, C., & Hewamalage, H. (2020a). LSTM-msnet: leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Transactions on Neural Networks and Learning Systems*.
Bandara, K., Bergmeir, C., & Smyl, S. (2020b). Forecasting across time series databases using recurrent neural networks on groups of similar series: a clustering approach. *Expert Systems with Applications*, *140*, Article 112896.
Bandyopadhyay, S. (2009). A dynamic model of cross-category competition: theory, tests and applications. *Journal of Retailing*, *85*(4), 468–479.
Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, *46*, 454–464.
Banks, J., Blundell, R., Oldfield, Z., & Smith, J. P. (2015). House price volatility and the housing ladder. *Tech. Rep. 21255*, National Bureau of Economic Research.
Bannister, R. N., Chipilski, H. G., & Martinez-Alvarado, O. (2020). Techniques and challenges in the assimilation of atmospheric water observations for numerical weather prediction towards convective scales. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 1–48.
Bansal, R., Tauchen, G., & Zhou, H. (2004). Regime shifts, risk premiums in the term structure, and the business cycle. *Journal of Business & Economic Statistics*, *22*(4), 396–409.
Bansal, R., & Zhou, H. (2002). Term structure of interest rates with regime shifts. *The Journal of Finance*, *57*(5), 1997–2043.
Banu, S., Hu, W., Hurst, C., & Tong, S. (2011). Dengue transmission in the Asia-Pacific region: impact of climate change and socio-environmental factors. *Tropical Medicine & International Health*, *16*(5), 598–607.
Bao, Y., Lee, T.-H., & Saltoglu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, *26*(3), 203–225.
Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento, C., Prendinger, H., & Henriques, E. M. P. (2018). Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers & Industrial Engineering*, *115*, 41–53.
Barbetta, S., Coccia, G., Moramarco, T., Brocca, L., & Todini, E. (2017). The multi temporal/multi-model approach to predictive uncertainty assessment in real-time flood forecasting. *Journal of Hydrology*, *551*, 555–576.
BarclayHedge (2018). Survey: Majority of hedge fund pros use AI/Machine learning in investment strategies. In *BarclayHedge*. https://www.barclayhedge.com/insider/barclayhedge-survey-majority-of-hedge-fund-pros-use-ai-machine-learning-in-investment-strategies, Accessed on 2020-09-01.
Barker, J. (2020). Machine learning in M4: What makes a good unstructured model?. *International Journal of Forecasting*, *36*(1), 150–155.
Barnhart, C., & Cohn, A. (2004). Airline schedule planning: Accomplishments and opportunities. *Manufacturing & Service Operations Management*, *6*(1), 3–22.
Barnhart, C., Fearing, D., & Vaze, V. (2014). Modeling passenger travel and delays in the national air transportation system. *Operations Research*, *62*(3), 580–601.
Barnichon, R., & Garda, P. (2016). Forecasting unemployment across countries: The ins and outs. *European Economic Review*, *84*, 165–183.
Barr, J. (2018). New – predictive scaling for EC2, powered by machine learning. In *AWS news blog*. https://aws.amazon.com/blogs/aws/new-predictive-scaling-for-ec2-powered-by-machine-learning, Accessed on 2020-09-01.
Barron, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Machine Learning*, *14*(1), 115–133.
Barroso, P. (2015). Momentum has its moments. *Journal of Financial Economics*, *116*(1), 111–121.
Barrow, D. K., & Crone, S. F. (2016a). A comparison of AdaBoost algorithms for time series forecast combination. *International Journal of Forecasting*, *32*(4), 1103–1119.
Barrow, D. K., & Crone, S. F. (2016b). Cross-validation aggregation for combining autoregressive neural network forecasts. *International Journal of Forecasting*, *32*(4), 1120–1137.
Barrow, D. K., & Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: implications for inventory management. *International Journal of Production Economics*, *177*, 24–33.
Barrow, D., Kourentzes, N., Sandberg, R., & Niklewski, J. (2020). Automatic robust estimation for exponential smoothing: Perspectives from statistics and machine learning. *Expert Systems with Applications*, *160*, Article 113637.
Bartelsman, E. J., Kurz, C. J., & Wolf, Z. (2011). Using census microdata to forecast US aggregate productivity. In *Working paper*.
Bartelsman, E. J., & Wolf, Z. (2014). Forecasting aggregate productivity using information from firm-level data. *The Review of Economics and Statistics*, *96*(4), 745–755.
Bartezzaghi, E., Verganti, R., & Zotteri, G. (1999). A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, *59*(1), 499–510.
Bass, F. (1969). A new product growth model for consumer durables. *Management Science*, *15*, 215–227.
Bass, F. M., Gordon, K., Ferguson, T. L., & Githens, M. L. (2001). DIRECTV: Forecasting diffusion of a new technology prior to product launch. *INFORMS Journal on Applied Analytics*, *31*(3S), S82–S93.
Bass, F., Krishnan, T., & Jain, D. (1994). Why the bass model fits without decision variables. *Marketing Science*, *13*, 203–223.
Bassetti, F., Casarin, R., & Ravazzolo, F. (2018). Bayesian nonparametric calibration and combination of predictive distributions. *Journal of the American Statistical Association*, *113*(522), 675–685.
Basturk, N., Borowska, A., Grassi, S., Hoogerheide, L., & van Dijk, H. (2019). Forecast density combinations of dynamic models and data driven portfolio strategies. *Journal of Econometrics*, *210*(1), 170–186.
Basu, S., Fernald, J. G., Oulton, N., & Srinivasan, S. (2003). The case of the missing productivity growth, or does information technology explain why productivity accelerated in the United States but not in the United Kingdom? *NBER Macroeconomics Annual*, *18*, 9–63.
Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468.
Baumeister, C., Guérin, P., & Kilian, L. (2015). Do high-frequency financial data help forecast oil prices? The midas touch at work. *International Journal of Forecasting*, *31*(2), 238–252.
Baumeister, C., & Kilian, L. (2015). Forecasting the real price of oil in a changing world: a forecast combination approach. *Journal of Business & Economic Statistics*, *33*(3), 338–351.
Beare, B. K., Seo, J., & Seo, W. (2017). Cointegrated linear processes in Hilbert space. *Journal of Time Series Analysis*, *38*(6), 1010–1027.
Becker, R., Hurn, S., & Pavlov, V. (2008). Modelling spikes in electricity prices. *The Economic Record*, *83*(263), 371–382.
Beckmann, M., & Bobkoski, F. (1958). Airline demand: An analysis of some frequency distributions. *Naval Research Logistics Quarterly*, *5*(1), 43–51.
Beckmann, J., & Schussler, R. (2016). Forecasting exchange rates under parameter and model uncertainty. *Journal of International Money and Finance*, *60*, 267–288.
Behera, M. K., Majumder, I., & Nayak, N. (2018). Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Engineering Science and Technology An International Journal*, *21*(3).
Bekaert, G., Hodrick, R. J., & Marshall, D. A. (2001). Peso problem explanations for term structure anomalies. *Journal of Monetary Economics*, *48*(2), 241–270.

Bekiros, S., Cardani, R., Paccagnini, A., & Villa, S. (2016). Dealing with financial instability under a DSGE modeling approach with banking intermediation: A predictability analysis versus TVP-vars. *Journal of Financial Stability*, *26*(C), 216–227.
Bekiros, S. D., & Paccagnini, A. (2014). Bayesian forecasting with small and medium scale factor-augmented vector autoregressive DSGE models. *Computational Statistics & Data Analysis*, *71*(C), 298–323.
Bekiros, S., & Paccagnini, A. (2015a). Estimating point and density forecasts for the US economy with a factor-augmented vector autoregressive DSGE model. *Studies in Nonlinear Dynamics & Econometrics*, *19*(2), 107–136.
Bekiros, S. D., & Paccagnini, A. (2015b). Macroprudential policy and forecasting using hybrid DSGE models with financial frictions and state space Markov-switching tvp-vars. *Macroeconomic Dynamics*, *19*(7), 1565–1592.
Bekiros, S. D., & Paccagnini, A. (2016). Policy oriented macroeconomic forecasting with hybrid DGSE and time varying parameter VAR models. *Journal of Forecasting*, *35*(7), 613–632.
Bélanger, A., & Sabourin, P. (2017). *Microsimulation and population dynamics: an introduction to modgen 12*. Springer, Cham.
Beliën, J., & Forcé, H. (2012). Supply chain management of blood products: A literature review. *European Journal of Operational Research*, *217*(1), 1–16.
Bell, F., & Smyl, S. (2018). Forecasting at uber: An introduction. URL https://eng.uber.com/forecasting-introduction/, Accessed on 2020-09-02.
Ben Taieb, S. (2014). Machine learning strategies for multi-step-ahead time series forecasting. (Ph.D. thesis), Free University of Brussels (ULB).
Ben Taieb, S., & Atiya, A. F. (2015). A bias and variance analysis for multistep-ahead time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, *PP*(99), 1.
Ben Taieb, S., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, *39*(8), 7067–7083.
Ben Taieb, S., & Hyndman, R. (2014). Boosting multi-step autoregressive forecasts. In *Proceedings of the 31st international conference on machine learning* (pp. 109–117).
Ben Taieb, S., Sorjamaa, A., & Bontempi, G. (2010). Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, *73*(10–12), 1950–1957.
Ben Taieb, S., Taylor, J. W., & Hyndman, R. J. (2020). Hierarchical probabilistic forecasting of electricity demand with smart meter data. *Journal of the American Statistical Association*.
Benati, L. (2007). Drift and breaks in labor productivity. *Journal of Economic Dynamics and Control*, *31*(8), 2847–2877.
Bender, J., Sun, X., Thomas, R., & Zdorovtsov, V. (2018). The promises and pitfalls of factor timing. *Journal of Portfolio Management*, *44*(4), 79–92.
Bendre, M., & Manthalkar, R. (2019). Time series decomposition and predictive analytics using MapReduce framework. *Expert Systems with Applications*, *116*, 108–120.
Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, B., Maddix, D., Turkmen, C., et al. (2020). Neural forecasting: Introduction and literature overview. arXiv:2004.10240.
Bennell, J., & Sutcliffe, C. (2004). Black-scholes versus artificial neural networks in pricing FTSE 100 options. *Intelligent Systems in Accounting, Finance & Management*, *12*(4), 243–260.
Berdugo, V., Chaussin, C., Dubus, L., Hebrail, G., & Leboucher, V. (2011). Analog method for collaborative very-short-term forecasting of powergeneration from photovoltaic systems. In *Next generation data mining summit* (pp. 1–5). Athens, Greece.
Berg, J. E., Nelson, F. D., & Rietz, T. A. (2008). Prediction market accuracy in the long run. *International Journal of Forecasting*, *24*(2), 285–300.
Berger, J. O. (1985). *Statistical decision theory and bayesian analysis*. Springer.
Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213.
Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, *32*(2), 303–312.
Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70–83.
Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474.
Berlinski, D. (2009). *The devil's delusion: atheism and its scientific pretensions*. Basic Books.
Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, *120*(1), 387–422.
Bernard, A., & Busse, M. (2004). Who wins the olympic games: Economic resources and medal totals. *The Review of Economics and Statistics*, *86*(1), 413–417.
Bernardini Papalia, R., & Fernandez Vazquez, E. (2020). Entropy-based solutions for ecological inference problems: A composite estimator. *Entropy*, *22*(7), 781.
Bernardo, J. M. (1984). Monitoring the 1982 spanish socialist victory: A Bayesian analysis. *Journal of the American Statistical Association*, *79*(387), 510–515.
Bernardo, J. M. (1994). *Bayesian theory*. Wiley.
Bernstein, R. (1995). *Style investing*. New York: John Wiley & Sons.
Berry, L. R., & West, M. (2020). BayesIan forecasting of many count-valued time series. *Journal of Business & Economic Statistics*, *38*(4), 872–887.
Bertsimas, D., & Pachamanova, D. (2008). Robust multiperiod portfolio management in the presence of transaction costs. *Computers & Operations Research*, *35*(1), 3–17.
Bessa, R. J., Miranda, V., Botterud, A., Zhou, Z., & Wang, J. (2012). Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy*, *40*(1), 29–39.
Bessa, R., Möhrlen, C., Fundel, V., Siefert, M., Browell, J., Haglund El Gaidi, S., et al. (2017). Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. *Energies*, *10*(9).
Besse, P., Cardot, H., & Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, *27*(4), 673–687.
Beyaztas, U., & Shang, H. L. (2019). Forecasting functional time series using weighted likelihood methodology. *Journal of Statistical Computation and Simulation*, *89*(16), 3046–3060.
Bhansali, R. J., & Kokoszka, P. S. (2002). Computation of the forecast coefficients for multistep prediction of long-range dependent time series. *International Journal of Forecasting*, *18*(2), 181–206.
Bianchi, L., Jarrett, J. E., & Hanumara, R. C. (1993). Forecasting incoming calls to telemarketing centers. *The Journal of Business Forecasting*, *12*(2), 3.
Bianchi, L., Jarrett, J., & Hanumara, R. C. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. *International Journal of Forecasting*, *14*(4), 497–504.
Bianco, A. M., García Ben, M., Martínez, E. J., & Yohai, V. J. (2001). Outlier detection in regression models with ARIMA errors using robust estimates. *Journal of Forecasting*, *20*(8), 565–579.
Bickel, J. E. (2007). Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, *4*(2), 49–65.
Bickel, P. J., & Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, *76*(374), 296–311.
Bickel, P. J., & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 1196–1217.
Bielecki, T. R., & Rutkowski, M. (2013). *Credit risk: modeling, valuation and hedging*. Springer Science & Business Media.
Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, *74*(5), 817–848.
Bijak, J. (2010). *Forecasting international migration in Europe: A Bayesian view*. Springer, Dordrecht.
Bijak, J., & Czaika, M. (2020). Black swans and grey rhinos: Migration policy under uncertainty. *Migration Policy Practice*, *X*(4), 14–20.
Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W. F., & Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, *38*(6), 470–487.
Bijak, J., & Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society. Series A*, *173*(4), 775–796.
Billio, M., Casarin, R., Ravazzolo, F., & van Dijk, H. K. (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics*, *177*(2), 213–232.

Binder, C. C. (2017). Measuring uncertainty based on rounding: New method and application to inflation expectations. *Journal of Monetary Economics*, *90*(C), 1–12.

Bisaglia, L., & Canale, A. (2016). Bayesian nonparametric forecasting for INAR models. *Computational Statistics & Data Analysis*, *100*, 70–78.

Bisaglia, L., & Gerolimetto, M. (2019). Model-based INAR bootstrap for forecasting INAR(p) models. *Computational Statistics*, *34*, 1815–1848.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, N.Y.: Springer.

Bjerknes, V. (1904). Das problem der wettervorhersage, betrachtet vom standpunkte der mechanik und der physik. *Meteorologische Zeitschrift*, *21*, 1–7.

Blanchard, O. J., & Kahn, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica*, *48*(5), 1305–1311.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877.

Bo, R., & Li, F. (2012). Probabilistic LMP forecasting under AC optimal power flow framework: Theory and applications. *Electric Power Systems Research*, *88*, 16–24.

Boccara, N. (2004). *Modeling complex systems*. New York: Springer-Verlag.

Bohk-Ewald, C., Li, P., & Myrskylä, M. (2018). Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(37), 9187–9192.

Boje, D. M., & Murnighan, J. K. (1982). Group confidence pressures in iterative decisions. *Management Science*, *28*(10), 1187–1196.

Bojer, C. S., & Meldgaard, J. P. (2020). Kaggle's forecasting competitions: An overlooked learning opportunity. *International Journal of Forecasting*.

Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *46*(4), 779–811.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, *69*(3), 542–547.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics*, *72*(3), 498–505.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151.

Bonaldo, D. (1991). *Competizione tra prodotti farmaceutici: strumenti di previsione*. (Master's thesis), University of Padua.

Boneva, L., Fawcett, N., Masolo, R. M., & Waldron, M. (2019). Forecasting the UK economy: Alternative forecasting methodologies and the role of off-model information. *International Journal of Forecasting*, *35*(1), 100–120.

Bonham, C., & Cohen, R. (2001). To aggregate, pool, or neither: Testing the rational expectations hypothesis using survey data. *Journal of Business & Economic Statistics*, *190*, 278–291.

Bontempi, G., & Ben Taieb, S. (2011). Conditionally dependent strategies for multiple-step-ahead prediction in local learning. *International Journal of Forecasting*, *27*(3), 689–699.

Bontempi, G., Birattari, M., & Bersini, H. (1999). Local learning for iterated time series prediction. In *International conference on machine learning*. In (pp. 32–38).

Booij, N., Ris, R. C., & Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions 1. Model description and validation. *Journal of Geophysical Research: Oceans*, *104*(C4), 7649–7666.

Boone, T., & Ganeshan, R. (2008). The value of information sharing in the retail supply chain: Two case studies. *Foresight: The International Journal of Applied Forecasting*, *9*, 12–17.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, *22*(3), 547–581.

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*(1–2), 3–43.

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2018). Over-reaction in macroeconomic expectations. *NBER Working Papers 24932*, National Bureau of Economic Research, Inc.

Bordignon, S., Bunn, D. W., Lisi, F., & Nan, F. (2013). Combining day-ahead forecasts for british electricity prices. *Energy Economics*, *35*, 88–103.

Bordley, R. F. (1982). The combination of forecasts: a Bayesian approach. *Journal of the Operational Research Society*, *33*(2), 171–174.

Bork, L., & Møller, S. V. (2015). Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection. *International Journal of Forecasting*, *31*(1), 63–78.

Bosq, D. (2000). *Linear processes in function spaces*. New York: Lecture Notes in Statistics.

Bosq, D., & Blanke, D. (2007). *Inference and prediction in large dimensions*. West Sussex, England: John Wiley & Sons.

Botimer, T. (1997). Select ideas on forecasting with sales relative to bucketing and 'seasonality'. *Company Report*, Continental Airlines, Inc.

Bourdeau, M., qiang Zhai, X., Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, *48*, Article 101533.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *26*(2), 211–243.

Box, George, E. P., Jenkins, & Gwilym (1976). *Time series analysis forecasting and control*. San Francisco, CA: Holden-Day.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: forecasting and control* (4th ed.). New Jersey: Wiley.

Boylan, J. E., & Babai, M. Z. (2016). On the performance of overlapping and non-overlapping temporal demand aggregation approaches. *International Journal of Production Economics*, *181*, 136–144.

Boylan, J. E., & Syntetos, A. A. (2003). Intermittent demand forecasting: size-interval methods based on averaging and smoothing. In C. C. Frangos (Ed.), *Proceedings of the international conference on quantitative methods in industry and commerce* (pp. 87–96). Athens: Technological Educational Institute.

Boylan, J., & Syntetos, A. (2006). Accuracy and accuracy implication metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, *4*, 39–42.

Boylan, J. E., & Syntetos, A. A. (2021). *Intermittent demand forecasting - context, methods and applications*. Wiley.

Boylan, J. E., Syntetos, A. A., & Karakostas, G. C. (2008). Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society*, *59*(4), 473–481.

Bozkurt, O. O., Biricik, G., & Tayşi, Z. C. (2017). Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market. *PLoS One*, *12*(4), Article e0175915.

Brandt, M. W., & Jones, C. S. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, *24*(4), 470–486.

Brass, W. (1974). Perspectives in population prediction: Illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society. Series A*, *137*(4), 532–583.

Braumoeller, B. F. (2019). *Only the dead: the persistence of war in the modern age*. Oxford University Press.

Brehmer, J., & Gneiting, T. (2020). Scoring interval forecasts: Equal-tailed, shortest, and modal interval. arXiv:2007.05709.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Brennan, J. (2020). Can novices trust themselves to choose trustworthy experts? Reasons for (reserved) optimism. *Social Epistemology*, *34*(3), 227–240.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3.

Brighton, H., & Gigerenzer, G. (2015). The bias bias. *Journal of Business Research*, *68*(8), 1772–1784.

Broer, T., & Kohlhas, A. (2018). Forecaster (mis-)behavior. *CEPR Discussion Papers 12898*, C.E.P.R. Discussion Papers.

Brooks, S., Gelman, A., Jones, G., & Meng, X. (2011). *Handbook of Markov Chain Monte Carlo*. Taylor & Francis.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., et al. (2005a). Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American Statistical Association*, *100*(469), 36–50.

Brown, A., Reade, J. J., & Vaughan Williams, L. (2019). When are prediction market prices most informative? *International Journal of Forecasting*, *35*(1), 420–428.

Brown, G., Wyatt, J., Harris, R., & Yao, X. (2005b). Diversity creation methods: a survey and categorisation. *Information Fusion*, *6*(1), 5–20.

Brücker, H., & Siliverstovs, B. (2006). On the estimation and forecasting of international migration: how relevant is heterogeneity across countries? *Empirical Economics*, *31*(3), 735–754.

Brunetti, C., & Lildholdt, P. M. (2002). Return-based and range-based (co)variance estimation - with an application to foreign exchange markets. SSRN:296875.

Bryant, J., & Zhang, J. L. (2018). *Bayesian demographic estimation and forecasting*. CRC Press.

Bu, R., & McCabe, B. P. (2008). Model selection, estimation and forecasting in INAR(p) models: A likelihood-based Markov chain approach. *International Journal of Forecasting*, *24*(1), 151–162.

Buchanan, B. G. (2019). *Artificial intelligence in finance*. London: The Alan Turing Institute.

Buckle, H. T. (1858). *vol. 1*, *History of civilization in England*. John W. Parker and Son.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, *36*(3), 391–405.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, *3*(2), 123–148.

Buizza, R. (2018). Ensemble forecasting and the need for calibration. In *Statistical postprocessing of ensemble forecasts* (pp. 15–48). Elsevier.

Bunea, A., Della Posta, P., Guidolin, M., & Manfredi, P. (2020). What do adoption patterns of solar panels observed so far tell about governments' incentive? Insights from diffusion models. *Technological Forecasting and Social Change*, *160*, Article 120240.

Bunn, D. W. (1975). A Bayesian approach to the linear combination of forecasts. *Journal of the Operational Research Society*, *26*(2), 325–329.

Bunn, D. W., & Salo, A. A. (1993). Forecasting with scenarios. *European Journal of Operational Research*, *68*(3), 291–303.

Burch, T. K. (2018). *Model-based demography: essays on integrating data, technique and theory*. Springer, Cham.

Bureau of Transportation Statistics (2020). Reporting carrier on-time performance (1987 - present). Accessed on 2020-09-09.

Burgman, M. A. (2016). *Trusting judgements: how to get the best out of experts*. Cambridge University Press.

Burman, P., Chow, E., & Nolan, D. (1994). A cross-validatory method for dependent data. *Biometrika*, *81*(2), 351–358.

Burridge, P., & Robert Taylor, A. (2006). Additive outlier detection via extreme-value theory. *Journal of Time Series Analysis*, *27*(5), 685–701.

Burton, J. W., Stein, M.-k., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239.

Busetti, F., & Marcucci, J. (2013). Comparing forecast accuracy: a Monte Carlo investigation. *International Journal of Forecasting*, *29*(1), 13–27.

Butler, D., Butler, R., & Eakins, J. (2020). Expert performance and crowd wisdom: Evidence from english premier league predictions. *European Journal of Operational Research*, *288*, 170–182.

Buys-Ballot, C. H. D. (1847). *Les changements périodiques de temperature*. Utrecht: Kemink Et Fils.

Byrne, J. P., Fazio, G., & Fiess, N. (2013). Primary commodity prices: Co-movements, common factors and fundamentals. *Journal of Development Economics*, *101*, 16–26.

Byrne, J. P., Korobilis, D., & Ribeiro, P. J. (2016). Exchange rate predictability in a changing world. *Journal of International Money and Finance*, *62*, 1–24.

Ca' Zorzi, M., Cap, A., Mijakovic, A., & Rubaszek, M. (2020). The predictive power of equilibrium exchange rate models. *Working Paper Series 2358*, European Central Bank.

Ca' Zorzi, M., Kolasa, M., & Rubaszek, M. (2017). Exchange rate forecasting with DSGE models. *Journal of International Economics*, *107*(C), 127–146.

Ca' Zorzi, M., Muck, J., & Rubaszek, M. (2016). Real exchange rate forecasting and PPP: This time the random walk loses. *Open Economies Review*, *27*(3), 585–609.

Ca' Zorzi, M., & Rubaszek, M. (2020). Exchange rate forecasting on a napkin. *Journal of International Money and Finance*, *104*, Article 102168.

Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, *12*(3), 309–316.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., et al. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, *13*(1), 1–35.

Calvo, E., & Escolar, M. (2003). The local voter: A geographically weighted approach to ecological inference. *American Journal of Political Science*, *47*(1), 189–204.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509–1531.

Canale, A., & Ruggiero, M. (2016). Bayesian nonparametric forecasting of monotonic functional time series. *Electronic Journal of Statistics*, *10*(2), 3265–3286.

Cappelen, Å., Skjerpen, T., & Tønnessen, M. (2015). Forecasting immigration in official population projections using an econometric model. *International Migration Review*, *49*(4), 945–980.

Cardani, R., Paccagnini, A., & Villa, S. (2015). Forecasting in a DSGE model with banking intermediation: Evidence from the US. *Working Paper 292*, University of Milano-Bicocca, Department of Economics.

Cardani, R., Paccagnini, A., & Villa, S. (2019). Forecasting with instabilities: An application to DSGE models with financial frictions. *Journal of Macroeconomics*, *61*(C), Article 103133.

Carlstein, E. (1990). Resampling techniques for stationary time-series: some recent developments. *Tech. rep.*, North Carolina State University, Department of Statistics.

Carmo, J. L., & Rodrigues, A. J. (2004). Adaptive forecasting of irregular demand processes. *Engineering Applications of Artificial Intelligence*, *17*(2), 137–143.

Carnevale, C., Angelis, E. D., Finzi, G., Turrini, E., & Volta, M. (2020). Application of data fusion techniques to improve air quality forecast: A case study in the northern Italy. *Atmosphere*, *11*(3).

Carnevale, C., Finzi, G., Pederzoli, A., Turrini, E., & Volta, M. (2018). An integrated data-driven/data assimilation approach for the forecast of PM10 levels in northern Italy. In C. Mensink, & G. Kallos (Eds.), *Air pollution modeling and its application XXV* (pp. 225–229). Springer International Publishing.

Carnevale, C., Finzi, G., Pisoni, E., & Volta, M. (2016). Lazy learning based surrogate models for air quality planning. *Environmental Modelling and Software*, *83*, 47–57.

Carriero, A., Clements, M. P., & Galvão, A. B. (2015). Forecasting with Bayesian multivariate vintage-based VARs. *International Journal of Forecasting*, *31*(3), 757–768.

Carroll, R. (2003). *The skeptic's dictionary: A collection of strange beliefs, amusing deceptions, and dangerous delusions*. Wiley.

Carson, R., Cenesizoglu, T., & Parker, R. (2011). Forecasting (aggregate) demand for US commercial air travel. *International Journal of Forecasting*, *27*(3), 923–941.

Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*(1), 41–75.

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480.

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. d. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, *137*, Article 106024.

Casals, J., Garcia-Hiernaux, A., Jerez, M., Sotoca, S., & Trindade, A. (2016). *State-space methods for time series analysis: theory, applications and software*. Chapman-Hall / CRC Press.

Casarin, R., Leisen, F., Molina, G., & ter Horst, E. (2015). A Bayesian beta Markov random field calibration of the term structure of implied risk neutral densities. *Bayesian Analysis*, *10*(4), 791–819.

Castle, J. L., Clements, M. P., & Hendry, D. F. (2015a). Robust approaches to forecasting. *International Journal of Forecasting*, *31*(1), 99–112.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2018a). Selecting a model for forecasting. In *Working paper*. Economics Department, Oxford University.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2020a). Modelling non-stationary 'big data'. *International Journal of Forecasting*.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2020b). Robust discovery of regression models. *Working paper 2020-W04*, Nuffield College, Oxford University.

Castle, J., Doornik, J., & Hendry, D. (2021). *The value of robust statistical forecasts in the COVID-19 pandemic, Vol. 256* (pp. 19–43). National Institute Economic Review.

Castle, J. L., Doornik, J. A., Hendry, D. F., & Pretis, F. (2015b). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, *3*(2), 240–264.

Castle, J., Doornik, J., Hendry, D., & Pretis, F. (2015c). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, *3*(2), 240–264.

Castle, J. L., Fawcett, N. W., & Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, *158*(1), 25–36.

Castle, J. L., & Hendry, D. F. (2010). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting*, *29*, 200–214.

Castle, J. L., & Hendry, D. F. (2020a). Climate econometrics: An overview. *Foundations and Trends in Econometrics* , *10*, 145–322.

Castle, J. L., & Hendry, D. F. (2020b). Identifying the causal role of CO2 during the ice ages. *Discussion paper 898*, Economics Department, Oxford University.

Castle, J. L., Hendry, D. F., & Kitov, O. I. (2018). Forecasting and nowcasting macroeconomic variables: A methodological overview. In EuroStat (Ed.), *Handbook on rapid estimates* (pp. 53–107). Brussels: UN/EuroStat.

Castle, J. L., Hendry, D. F., & Martinez, A. B. (2020c). The paradox of stagnant real wages yet rising 'living standards' in the UK. *Tech. rep.*, VoxEU.

Caswell, H. (2019a). *Sensitivity analysis: matrix methods in demography and ecology*. Springer, Cham.

Caswell, H. (2019b). The formal demography of kinship: A matrix formulation. *Demographic Research*, *41*(24), 679–712.

Caswell, H. (2020). The formal demography of kinship II: Multistate models, parity, and sibship. *Demographic Research*, *42*(38), 1097–1146.

Catalán, B., & Trívez, F. J. (2007). Forecasting volatility in GARCH models with additive outliers. *Quantitative Finance*, *7*(6), 591–596.

Cavalcante, L., Bessa, R. J., Reis, M., & Browell, J. (2016). LASSO vector autoregression structures for very short-term wind power forecasting. *Wind Energy*, *20*, 657–675.

Cazelles, B., Chavez, M., McMichael, A. J., & Hales, S. (2005). Nonstationary influence of el nino on the synchronous dengue epidemics in thailand. *PLoS Medicine*, *2*(4), Article e106.

Cederman, L.-E. (2003). Modeling the size of wars: From billiard balls to sandpiles. *The American Political Science Review*, *97*(1), 135–150.

Ceron, A., Curini, L., & Iacus, S. M. (2016). *Politics and big data: nowcasting and forecasting elections with social media*. Routledge.

Chae, Y. T., Horesh, R., Hwang, Y., & Lee, Y. M. (2016). Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy and Buildings*, *111*, 184–194.

Chakraborty, T., Chattopadhyay, S., & Ghosh, I. (2019). Forecasting dengue epidemics using a hybrid methodology. *Physica A: Statistical Mechanics and its Applications*, *527*, Article 121266.

Chan, N. H., & Genovese, C. R. (2001). A comparison of linear and nonlinear statistical techniques in performance attribution. *IEEE Transactions on Neural Networks*, *12*(4), 922–928.

Chan, C. K., Kingsman, B. G., & Wong, H. (1999). The value of combining forecasts in inventory management–a case study in banking. *European Journal of Operational Research*, *117*(2), 199–210.

Chan, J. S., Lam, C. P., Yu, P. L., Choy, S. T., & Chen, C. W. (2012). A Bayesian conditional autoregressive geometric process model for range data. *Computational Statistics & Data Analysis*, *56*(11), 3006–3019.

Chan, F., & Pauwels, L. L. (2018). Some theoretical results on forecast combinations. *International Journal of Forecasting*, *34*(1), 64–74.

Chan, K. S., & Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, *7*(3), 179–190.

Chan, J. C., & Yu, X. (2020). Fast and accurate variational inference for large Bayesian VARs with stochastic volatility. In *CAMA working paper*.

Chandola, V., Banerjee, A., & Kumar, V. (2007). Outlier detection: A survey. *ACM Computing Surveys*, *14*, 15.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*(3), 1–58.

Chang, Y., Kim, C. S., & Park, J. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, *192*(1), 152–167.

Chaouch, M. (2014). Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves. *IEEE Transactions on Smart Grid*, *5*(1), 411–419.

Chase, C. (2021). Assisted demand planning using machine learning. In M. Gilliland, L. Tashman, & U. Sglavo (Eds.), *Business forecasting: the emerging role of artificial intelligence and machine learning* (pp. 110–114). Wiley.

Chatfield, C. (1986). Simple is best? *International Journal of Forecasting*, *2*(4), 401–402.

Chatziantoniou, I., Degiannakis, S., Eeckels, B., & Filis, G. (2016). Forecasting tourist arrivals using origin country macroeconomics. *Applied Economics*, *48*(27), 2571–2585.

Chavez-Demoulin, V., Davison, A. C., & McNeil, A. J. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance*, *5*(2), 227–234.

Checchi, F., & Roberts, L. (2005). Interpreting and using mortality data in humanitarian emergencies. *Humanitarian Practice Network*, *52*.

Chen, R. (1995). Threshold variable selection in open-loop threshold autoregressive models. *Journal of Time Series Analysis*, *16*(5), 461–481.

Chen, C. W. S., Chiang, T. C., & So, M. K. P. (2003). Asymmetrical reaction to US stock-return news: evidence from major stock markets based on a double-threshold model. *Journal of Economics and Business*, *55*(5), 487–502.

Chen, C. W., Gerlach, R., Hwang, B. B., & McAleer, M. (2012). Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range. *International Journal of Forecasting*, *28*(3), 557–574.

Chen, C. W., Gerlach, R., & Lin, E. M. (2008). Volatility forecasting using threshold heteroskedastic models of the intra-day range. *Computational Statistics & Data Analysis*, *52*(6), 2990–3010.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). ACM.

Chen, Y., Kang, Y., Chen, Y., & Wang, Z. (2020). Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, *399*, 491–501.

Chen, J., Li, K., Rong, H., Bilal, K., Li, K., & Philip, S. Y. (2019a). A periodicity-based parallel time series prediction algorithm in cloud computing environments. *Information Sciences*, *496*, 506–537.

Chen, C., & Liu, L.-M. (1993a). Forecasting time series with outliers. *Journal of Forecasting*, *12*(1), 13–35.

Chen, C., & Liu, L.-M. (1993b). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, *88*(421), 284–297.

Chen, Y., Marron, J. S., & Zhang, J. (2019b). Modeling seasonality and serial dependence of electricity price curves with warping functional autoregressive dynamics. *The Annals of Applied Statistics*, *13*(3), 1590–1616.

Chen, C. W. S., & So, M. K. P. (2006). On a threshold heteroscedastic model. *International Journal of Forecasting*, *22*(1), 73–89.

Chen, M.-F., Wang, R.-H., & Hung, S.-L. (2015). Predicting health-promoting self-care behaviors in people with pre-diabetes by applying bandura social learning theory. *Applied Nursing Research*, *28*(4), 299–304.

Chen, R., Yang, L., & Hafner, C. (2004). Nonparametric multistep-ahead prediction in time series analysis. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *66*(3), 669–686.

Cheng, G., & Yang, Y. (2015). Forecast combination with outlier protection. *International Journal of Forecasting*, *31*(2), 223–237.

Cheng, C., Yu, L., & Chen, L. J. (2012). Structural nonlinear damage detection based on ARMA-GARCH model. *Applied Mechanics and Materials*, *204–208*, 2891–2896.

Cheung, Y.-W., Chinn, M. D., & Pascual, A. G. (2005). Empirical exchange rate models of the nineties: Are any fit to survive? *Journal of International Money and Finance*, *24*(7), 1150–1175.

Cheung, Y.-W., Chinn, M. D., Pascual, A. G., & Zhang, Y. (2019). Exchange rate prediction redux: New models, new data, new currencies. *Journal of International Money and Finance*, *95*, 332–362.

Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, *21*(4), 746–785.

Chew, V. (1968). Simultaneous prediction intervals. *Technometrics*, *10*(2), 323–330.

Chiang, M. H., & Wang, L. M. (2011). Volatility contagion: A range-based volatility approach. *Journal of Econometrics*, *165*(2), 175–189.

Chicco, G., Cocina, V., Di Leo, P., Spertino, F., & Massi Pavan, A. (2015). Error assessment of solar irradiance forecasts and AC power from energy conversion model in grid-connected photovoltaic systems. *Energies*, *9*(1), 8.

Chinco, A., Clark-Joseph, A. D., & Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, *74*(1), 449–492.

Chiroma, H., Abdulkareem, S., & Herawan, T. (2015). Evolutionary neural network model for west texas intermediate crude oil price prediction. *Applied Energy*, *142*, 266–273.

Choi, E., Özer, O., & Zheng, Y. (2020). Network trust and trust behaviors among executives in supply chain interactions. *Management Science*.

Chong, Y. Y., & Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, *53*(4), 671–690.

Chou, R. Y.-T. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit, and Banking*, *37*(3), 561–582.

Chou, R. Y., & Cai, Y. (2009). Range-based multivariate volatility model with double smooth transition in conditional correlation. *Global Finance Journal*, *20*(2), 137–152.

Chou, R. Y., Chou, H., & Liu, N. (2015). Range volatility: A review of models and empirical studies. In C. F. Lee, & J. C. Lee (Eds.), *Handbook of financial econometrics and statistics* (pp. 2029–2050). Springer New York.

Chou, R. Y., & Liu, N. (2010). The economic value of volatility timing using a range-based volatility model. *Journal of Economic Dynamics and Control*, *34*(11), 2288–2301.

Chou, R. Y., Wu, C. C., & Liu, N. (2009). Forecasting time-varying covariance with a range-based dynamic conditional correlation model. *Review of Quantitative Finance and Accounting*, *33*(4), 327–345.

Choudhury, A., & Urena, E. (2020). Forecasting hourly emergency department arrival using time series analysis. *British Journal of Healthcare Management*, *26*(1), 34–43.

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing*, *307*, 72–77.

Christensen, T., Hurn, S., & Lindsay, K. (2009). It never rains but it pours: Modeling the persistence of spikes in electricity prices. *Energy Journal*, *30*(1), 25–48.

Christensen, T. M., Hurn, A. S., & Lindsay, K. A. (2012). Forecasting spikes in electricity prices. *International Journal of Forecasting*, *28*(2), 400–411.

Christiano, L. J., Eichenbaum, M. S., & Trabandt, M. (2018). On DSGE models. *Journal of Economic Perspectives*, *32*(3), 113–140.

Christoffersen, P., & Langlois, H. (2013). The joint dynamics of equity market factors. *Journal of Financial and Quantitative Analysis*, *48*(5), 1371–1404.

Chung, H., Kiley, M. T., & Laforte, J.-P. (2010). Documentation of the estimated, dynamic, optimization-based (EDO) model of the U.S. economy: 2010 version. *Finance and Economics Discussion Series 2010–29*, Board of Governors of the Federal Reserve System (U.S.).

Chung, C., Niu, S.-C., & Sriskandarajah, C. (2012). A sales forecast model for short-life-cycle products: New releases at blockbuster. *Production and Operations Management*, *21*(5), 851–873.

Cirillo, P., & Taleb, N. N. (2016a). Expected shortfall estimation for apparently infinite-mean models of operational risk. *Quantitative Finance*, *16*(10), 1485–1494.

Cirillo, P., & Taleb, N. N. (2016b). On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications*, *452*, 29–45.

Cirillo, P., & Taleb, N. N. (2019). The decline of violent conflicts: What do the data really say? In A. Toje, & B. N. V. Steen (Eds.), *The causes of peace: what we know* (pp. 57–86). The Causes of Peace: What We Know.

Claeskens, G., Magnus, J. R., Vasnev, A. L., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, *32*(3), 754–762.

Clark, D. A. (1990). Verbal uncertainty expressions: A critical review of two decades of research. *Current Psychology*, *9*(3), 203–235.

Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*(1), 85–110.

Clark, T. E., & McCracken, M. W. (2009). Tests of equal predictive ability with real-time data. *Journal of Business & Economic Statistics*, *27*(4), 441–454.

Clark, T., & McCracken, M. (2013). Advances in forecast evaluation. *vol. 2*, In *Handbook of economic forecasting* (pp. 1107–1201). Elsevier.

Clark, T., & West, K. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of Econometrics*, *135*(1–2), 155–186.

Clauset, A. (2018). Trends and fluctuations in the severity of interstate wars. *Science Advances*, *4*(2), eaao3580.

Clauset, A., & Gleditsch, K. S. (2018). Trends in conflicts: What do we know and what can we know? In A. Gheciu, & W. C. Wohlforth (Eds.), *The oxford handbook of international security*. Oxford University Press.

Cleave, N., Brown, P. J., & Payne, C. D. (1995). Evaluation of methods for ecological inference. *Journal of the Royal Statistical Society, Series A*, *158*(1), 55–72.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.

Clemen, R. T. (2008). Comment on cooke's classical method. *Reliability Engineering & System Safety*, *93*(5), 760–765.

Clemen, R. T., & Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business & Economic Statistics*, *4*(1), 39–46.

Clements, M. P. (2009). Internal consistency of survey respondents' forecasts: Evidence based on the survey of professional forecasters. In J. L. Castle, & N. Shephard (Eds.), *The methodology and practice of econometrics. a festschrift in honour of David F. Hendry. Chapter 8* (pp. 206–226). Oxford: Oxford University Press.

Clements, M. P. (2010). Explanations of the inconsistencies in survey respondents forecasts. *European Economic Review*, *54*(4), 536–549.

Clements, M. P. (2011). An empirical investigation of the effects of rounding on the SPF probabilities of decline and output growth histograms. *Journal of Money, Credit and Banking*, *43*(1), 207–220.

Clements, M. P. (2014a). Forecast uncertainty - ex ante and ex post: US inflation and output growth. *Journal of Business & Economic Statistics*, *32*(2), 206–216.

Clements, M. P. (2014b). US inflation expectations and heterogeneous loss functions, 1968–2010. *Journal of Forecasting*, *33*(1), 1–14.

Clements, M. P. (2017). Assessing macro uncertainty in real-time when data are subject to revision. *Journal of Business & Economic Statistics*, *35*(3), 420–433.

Clements, M. P. (2018). Are macroeconomic density forecasts informative? *International Journal of Forecasting*, *34*, 181–198.

Clements, M. P. (2019). *Macroeconomic survey expectations*. Palgrave Texts in Econometrics. Palgrave Macmillan.

Clements, M. P., & Galvão, A. B. (2012). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. *Journal of Business & Economic Statistics*, *30*(4), 554–562.

Clements, M. P., & Galvão, A. B. (2013a). Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *International Journal of Forecasting*, *29*(4), 698–714.

Clements, M. P., & Galvão, A. B. (2013b). Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics*, *28*(3), 458–477.

Clements, M. P., & Galvão, A. B. (2017). Data revisions and real-time probabilistic forecasting of macroeconomic variables. *Discussion Paper ICM-2017-01*, ICMA, Henley Business School, Reading.

Clements, M. P., & Galvão, A. B. (2019). Data revisions and real-time forecasting. *The Oxford Research Encyclopedia of Economics and Finance*.

Clements, M. P., & Harvey, D. I. (2011). Combining probability forecasts. *International Journal of Forecasting*, *27*(2), 208–223.

Clements, M. P., & Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge University Press.

Clements, M. P., & Hendry, D. F. (1999). *Zeuthen lecture book series*, *Forecasting non-stationary economic time series*. Cambridge, MA: MIT Press.

Clements, M. P., & Hendry, D. F. (2005). Evaluating a model by forecast performance. *Oxford Bulletin of Economics and Statistics*, *67*, 931–956.

Clements, A. E., Herrera, R., & Hurn, A. S. (2015). Modelling interregional links in electricity price spikes. *Energy Economics*, *51*, 383–393.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.

Clottey, T., Benton Jr., W. C., & Srivastava, R. (2012). Forecasting product returns for remanufacturing operations. *Decision Sciences*, *43*(4), 589–614.

Cludius, J., Hermann, H., Matthes, F. C., & Graichen, V. (2014). The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016: Estimation and distributional implications. *Energy Economics*, *44*, 302–313.

Coates, D., & Humphreys, B. R. (1999). The growth effects of sport franchises, stadia, and arenas. *Journal of Policy Analysis and Management*, *18*(4), 601–624.

Coates, D., & Humphreys, B. (2010). Week-to-week attendance and competitive balance in the national football league. *International Journal of Sport Finance*, *5*(4), 239.

Coccia, G. (2011). Analysis and developments of uncertainty processors for real time flood forecasting. (Ph.D. thesis), Alma Mater Studiorum University of Bologna.

Coccia, G., & Todini, E. (2011). Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrology and Earth System Sciences*, *15*(10), 3253–3274.

Cohin, A., & Chantret, F. (2010). The long-run impact of energy prices on world agricultural markets: The role of macro-economic linkages. *Energy Policy*, *38*(1), 333–339.

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, *120*(1), 116–159.

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644–2678.

Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, *38*(10), 1394–1414.

Commandeur, J. J. F., Koopman, S. J., & Ooms, M. (2011). Statistical software for state space methods. *Journal of Statistical Software*, *41*(1), 1–18.

Congdon, P. (1990). Graduation of fertility schedules: an analysis of fertility patterns in London in the 1980s and an application to fertility forecasts. *Regional Studies*, *24*(4), 311–326.

Consolo, A., Favero, C., & Paccagnini, A. (2009). On the statistical identification of DSGE models. *Journal of Econometrics*, *150*(1), 99–115.

Continuous Mortality Investigation (2020). The CMI mortality projections model, cmi_2019. *Working paper*, London: Institute of Actuaries and Faculty of Actuaries.

Cook, S., & Thomas, C. (2003). An alternative approach to examining the ripple effect in U.K. house prices. *Applied Economics Letters*, *10*(13), 849–851.

Cooke, R. M. (1991). *Experts in uncertainty: opinion and subjective probability in science*. Oxford University Press.

Copeland, M. T. (1915). Statistical indices of business conditions. *Quarterly Journal of Economics*, *29*(3), 522–562.

Corani, G. (2005). Air quality prediction in milan: Feed-forward neural networks, pruned neural networks and lazy learning. *Ecological Modelling*, *185*, 513–529.

Cordeiro, C., & Neves, M. (2006). The bootstrap methodology in time series forecasting. In J. Black, & A. White (Eds.), *Proceedings of CompStat2006* (pp. 1067–1073). Springer Verlag.

Cordeiro, C., & Neves, M. (2009). Forecasting time series with BOOT.EXPOS procedure. *REVSTAT-Statistical Journal*, *7*(2), 135–149.

Cordeiro, C., & Neves, M. M. (2010). Boot.EXPOS In NNGC competition. In *The 2010 international joint conference on neural networks (IJCNN)* (pp. 1–7). IEEE.

Cordeiro, C., & Neves, M. M. (2013). Predicting and treating missing data with boot.expos. In *Advances in regression, survival analysis, extreme values, markov processes and other statistical applications* (pp. 131–138). Springer.

Cordeiro, C., & Neves, M. M. (2014). Forecast intervals with boot.expos. In *New advances in statistical modeling and applications* (pp. 249–256). Springer.

Corominas, A., Lusa, A., & Dolors Calvet, M. (2015). Computing voter transitions: The elections for the catalan parliament, from 2010 to 2012. *Journal of Industrial Engineering and Management*, *8*(1), 122–136.

Corradi, V., Swanson, N. R., & Olivetti, C. (2001). Predictive ability with cointegrated variables. *Journal of Econometrics*, *104*(2), 315–358.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Couharde, C., Delatte, A.-L., Grekou, C., Mignon, V., & Morvillier, F. (2018). EQCHANGE: A world database on actual and equilibrium effective exchange rates. *International Economics*, *156*, 206–230.

Courgeau, D. (2012). Probability and social science: methodologial relationships between the two approaches?. *Tech. Rep. 43102*, Germany: University Library of Munich.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*, 777–795.

Creal, D. D., & Tsay, R. S. (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics*, *189*(2), 335–345.

Croll, J. (1875). *Climate and time in their geological relations, a theory of secular changes of the earth's climate*. New York: D. Appleton.

Crone, S. F., Hibon, M., & Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, *27*(3), 635–660.

Cross, J. L. (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, *36*(3), 899–916.

Cross, R., & Sproull, L. (2004). More than an answer: Information relationships for actionable knowledge. *Organization Science*, *15*(4), 446–462.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, *23*(3), 289–303.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In G. Elliott, C. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting, volume 1. Handbook of economics 24* (pp. 961–982). Elsevier, Horth-Holland.

Croushore, D. (2011a). Forecasting with real-time data vintages (chapter 9). In M. P. Clements, & D. F. Hendry (Eds.), *The oxford handbook of economic forecasting* (pp. 247–267). Oxford University Press.

Croushore, D. (2011b). Frontiers of real-time data analysis. *Journal of Economic Literature*, *49*, 72–100.

Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, *105*(1), 111–130.

Croxson, K., & Reade, J. J. (2014). Information and efficiency: goal arrival in soccer. *The Economic Journal*, *124*(575), 62–91.

Cunado, J., & De Gracia, F. P. (2005). Oil prices, economic activity and inflation: Evidence for some Asian countries. *The Quarterly Review of Economics and Finance*, *45*(1), 65–83.

Cunningham, C. R. (2006). House price uncertainty, timing of development, and vacant land prices: Evidence for real options in seattle. *Journal of Urban Economics*, *59*(1), 1–31.

Cunningham, A., Eklund, J., Jeffery, C., Kapetanios, G., & Labhard, V. (2009). A state space approach to extracting the signal from uncertain data. *Journal of Business & Economic Statistics*, *30*, 173–180.

Curran, M., & Velic, A. (2019). Real exchange rate persistence and country characteristics: A global analysis. *Journal of International Money and Finance*, *97*, 35–56.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, *2*(4), 303–314.

Czado, C., Gneiting, T., & Held, L. (2009). Predictive model assessment for count data. *Biometrics*, *65*(4), 1254–1261.

Dagum, E. B. (1988). *The X11ARIMA/88 seasonal adjustment method: foundations and user's manual*. Statistics Canada, Time Series Research and Analysis Division.

Dai, Q., & Singleton, K. (2003). Term structure dynamics in theory and reality. *Review of Financial Studies*, *16*(3), 631–678.

Dai, Q., Singleton, K. J., & Yang, W. (2007). Regime shifts in a dynamic term structure model of U.S. Treasury bond yields. *Review of Financial Studies*, *20*(5), 1669–1706.

Dalkey, N. C. (1969). The delphi method: An experimental study of group opinion. *Research Memoranda*, *RM-5888-PR*.

Dalla Valle, A., & Furlan, C. (2011). Forecasting accuracy of wind power technology diffusion models across countries. *International Journal of Forecasting*, *27*(2), 592–601.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, *122*(2), 221–247.

Dantas, T. M., & Cyrino Oliveira, F. L. (2018). Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing. *International Journal of Forecasting*, *34*(4), 748–761.

Dantas, T. M., Cyrino Oliveira, F. L., & Varela Repolho, H. M. (2017). Air transportation demand forecast through bagging holt winters methods. *Journal of Air Transport Management*, *59*, 116–123.
Danti, P., & Magnani, S. (2017). Effects of the load forecasts mismatch on the optimized schedule of a real small-size smart prosumer. *Energy Procedia*, *126*, 406–413.
Dantzig, G. B., & Infanger, G. (1993). Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research*, *45*, 59–76.
Das, S., & Chen, M. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, *53*(9), 1375–1388.
Daskalaki, C., Kostakis, A., & Skiadopoulos, G. (2014). Are there common factors in individual commodity futures returns? *Journal of Banking & Finance*, *40*(C), 346–363.
Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003.
Dawid, A. P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association*, *77*(379), 605–610.
Dawid, A. P. (1984). Statistical theory: The prequential approach (with discussion and rejoinder). *Journal of the Royal Statistical Society, Series A*, *147*, 278–292.
Dawid, A. P. (1985). Calibration-based empirical probability. *The Annals of Statistics*, *13*(4), 1251–1274.
Dawid, A. P., DeGroot, M. H., Mortera, J., Cooke, R., French, S., Genest, C., et al. (1995). Coherent combination of experts' opinions. *Test*, *4*(2), 263–313.
de Albuquerquemello, V. P., de Medeiros, R. K., da Nóbrega Besarria, C., & Maia, S. F. (2018). Forecasting crude oil price: Does exist an optimal econometric model? *Energy*, *155*, 578–591.
de Almeida Marques-Toledo, C., Degener, C. M., Vinhal, L., Coelho, G., Meira, W., Codeço, C. T., et al. (2017). Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLoS Neglected Tropical Diseases*, *11*(7), Article e0005729.
de Almeida Pereira, G. A., & Veiga, A. (2019). Periodic copula autoregressive model designed to multivariate streamflow time series modelling. *Water Resources Management*, *33*(10), 3417–3431.
De Baets, S. (2019). Surveying forecasting: a review and directions for future research. *International Journal of Information and Decision Sciences*.
De Baets, S., & Harvey, N. (2020). Using judgment to select and adjust forecasts from statistical models. *European Journal of Operational Research*, *284*(3), 882–895.
De Beer, J. (1985). A time series model for cohort data. *Journal of the American Statistical Association*, *80*(391), 525–530.
De Beer, J. (1990). Projecting age-specific fertility rates by using time-series methods. *European Journal of Population*, *5*(4), 315–346.
De Beer, J. (2008). Forecasting international migration: Time series projections vs argument-based forecasts. In *International migration in Europe* (pp. 283–306). Chichester, UK: John Wiley & Sons, Ltd.
de Brito, M. P., & van der Laan, E. A. (2009). Inventory control with product returns: The impact of imperfect information. *European Journal of Operational Research*, *194*(1), 85–101.
De Gooijer, J. (1998). On threshold moving-average models. *Journal of Time Series Analysis*, *19*(1), 1–18.
De Gooijer, J. G., & Hyndman, R. J. (2006). 25 Years of time series forecasting. *International Journal of Forecasting*, *22*, 443–473.
De Iaco, S., & Maggio, S. (2016). A dynamic model for age-specific fertility rates in Italy. *Spatial Statistics*, *17*, 105–120.
de Kok, S. (2017). The quest for a better forecast error metric: Measuring more than the average error. *Foresight: The International Journal of Applied Forecasting*, *46*, 36–45.
De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527.
De Mare, J. (1980). Optimal prediction of catastrophes with applications to Gaussian processes. *The Annals of Probability*, *8*(4), 841–850.
De Menezes, L. M., Bunn, D. W., & Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, *120*(1), 190–204.
de Nicola, F., De Pace, P., & Hernandez, M. A. (2016). Co-movement of major energy, agricultural, and food commodity price returns: A time-series assessment. *Energy Economics*, *57*(C), 28–41.
de Oliveira, E. M., & Cyrino Oliveira, F. L. (2018). Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. *Energy*, *144*, 776–788.
de Queiroz, A. R. (2016). Stochastic hydro-thermal scheduling optimization: An overview. *Renewable and Sustainable Energy Reviews*, *62*, 382–395.
Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107–113.
Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, *74*, 902–924.
Debecker, A., & Modis, T. (1994). Determination of the uncertainties in S-curve logistic fits. *Technological Forecasting and Social Change*, *46*(2), 153–173.
Debecker, A., & Modis, T. (2021). Poorly known aspects of flattening the curve of COVID 19. *Technological Forecasting and Social Change*, *163*(120432).
Dees, S., Mauro, F. d., Pesaran, M. H., & Smith, L. V. (2007). Exploring the international linkages of the euro area: a global VAR analysis. *Journal of Applied Economics*, *22*(1), 1–38.
Degiannakis, S. A., Filis, G., Klein, T., & Walther, T. (2020). Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting*.
DeGroot, M. H. (2004). *Optimal statistical decisions*. Hoboken, N.J: Wiley-Interscience.
Dekker, M., van Donselaar, K., & Ouwehand, P. (2004). How to use aggregation and combined forecasting to improve seasonal demand forecasts. *International Journal of Production Economics*, *90*(2), 151–167.
Del Negro, M., & Schorfheide, F. (2004). Priors from general equilibrium models for VARS. *International Economic Review*, *45*(2), 643–673.
Del Negro, M., & Schorfheide, F. (2006). How good is what you've got? DGSE-VAR as a toolkit for evaluating DSGE models. *Economic Review-Federal Reserve Bank of Atlanta*, *91*(2), 21.
Del Negro, M., & Schorfheide, F. (2013). DSGE model-based forecasting. In G. Elliott, & A. Timmermann (Eds.), *vol. 2*, *Handbook of economic forecasting* (pp. 57–140). Amsterdam, Horth-Holland.
Dellaportas, P., Denison, D. G. T., & Holmes, C. (2007). Flexible threshold models for modelling interest rate volatility. *Econometric Reviews*, *26*(2–4), 419–437.
Delle Monache, L., Hacker, J. P., Zhou, Y., Deng, X., & Stull, R. B. (2006). Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *Journal of Geophysical Research: Atmospheres*, *111*(D24).
Demirovic, E., Stuckey, P. J., Bailey, J., Chan, J., Leckie, C., Ramamohanarao, K., et al. (2019). Predict+optimise with ranking objectives: Exhaustively learning linear functions. In *IJCAI* (pp. 1078–1085).
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *39*, 1–38.
Dempster, M., Payne, T., Romahi, Y., & Thompson, G. (2001). Computational learning techniques for intraday FX trading using popular technical indicators. *IEEE Transactions on Neural Networks*, *12*, 744–754.
Di Corso, E., Cerquitelli, T., & Apiletti, D. (2018). Metatech: Meteorological data analysis for thermal energy characterization by means of self-learning transparent models. *Energies*, *11*(6), 1336.
Diab, D. L., Pui, S.-Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non-US samples. *International Journal of Selection and Assessment*, *19*(2), 209–216.
Dichtl, H., Drobetz, W., Lohre, H., Rother, C., & Vosskamp, P. (2019). Optimal timing and tilting of equity factors. *Financial Analysts Journal*, *75*(4), 84–102.
Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama*, *263*(10), 1385–1389.
Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431.
Dickey, D. A., & Pantula, S. G. (1987). Determining the order of differencing in autoregressive processes. *Journal of Business & Economic Statistics*, *5*(4), 455–461.
Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics*, *33*(1), 1.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.
Diebold, F. X., & Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, *6*(1), 21–40.
Diebold, F. X., & Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, *6*(4), 503–508.
Diebold, F. X., & Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting*, *35*(4), 1679–1691.
Dieckmann, N. F., Gregory, R., Peters, E., & Hartman, R. (2017). Seeing what you want to see: How imprecise uncertainty ranges enhance motivated reasoning. *Risk Analysis*, *37*(3), 471–486.
Dietrich, J. K., & Joines, D. H. (1983). Rational expectations, informational efficiency, and tests using survey data: A comment. *The Review of Economics and Statistics*, *65*(3), 525–529.
Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126.
Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can even slightly modify them. *Management Science*, *64*(3), 1155–1170.
Dietzel, M., Baltzer, P. A., Vag, T., Gröschel, T., Gajda, M., Camara, O., et al. (2010). Application of breast MRI for prediction of lymph node metastases–systematic approach using 17 individual descriptors and a dedicated decision tree. *Acta Radiologica*, *51*(8), 885–894.
Ding, R., Wang, Q., Dang, Y., Fu, Q., Zhang, H., & Zhang, D. (2015). Yading: Fast clustering of large-scale time series data. *Proceedings of the VLDB Endowment*, *8*(5), 473–484.
Dion, P., Galbraith, N., & Sirag, E. (2020). Using expert elicitation to build long-term projection assumptions. In S. Mazzuco, & N. Keilman (Eds.), *Developments in demographic forecasting* (pp. 43–62). Cham: Springer International Publishing.
Dissanayake, G. S., Peiris, M. S., & Proietti, T. (2018). Fractionally differenced gegenbauer processes with long memory: A review. *Statistical Science*, *33*, 413–426.
Divakar, S., Ratchford, B. T., & Shankar, V. (2005). CHAN4CAST: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. *Marketing Science*, *24*(3), 334–350.
Dixon, M. J., & Coles, S. C. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics*, *47*(3), 265—280.
Do, L., Vu, H., Vo, B., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation Research Part C (Emerging Technologies)*, *108*, 12–28.
Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, *3*(1), 1–100.
Dokumentov, A. (2017). Smoothing, decomposition and forecasting of multidimensional and functional time series using regularisation. *Monash University*.
Dokumentov, A., & Hyndman, R. J. (2018). stR: STR Decomposition. R package version 0.4.
Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., & Ogliari, E. (2015). A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies*, *8*(2), 1–16.
Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., & Ogliari, E. (2018). Comparison of training approaches for photovoltaic forecasts by means of machine learning. *Applied Sciences*, *8*(2), 228.
Dolgin, E. (2010). *Better forecasting urged to avoid drug waste*. Nature Publishing Group.
Dong, X., Li, Y., Rapach, D. E., & Zhou, G. (2021). Anomalies and the expected market return. *The Journal of Finance*, in press.
Doornik, J. A. (2018). Autometrics. In J. L. Castle, & N. Shephard (Eds.), *The methodology and practice of econometrics* (pp. 88–121). Oxford: Oxford University Press.
Doornik, J. A., Castle, J. L., & Hendry, D. F. (2020a). Card forecasts for M4. *International Journal of Forecasting*, *36*, 129–134.
Doornik, J. A., Castle, J. L., & Hendry, D. F. (2020b). Short-term forecasting of the coronavirus pandemic. *International Journal of Forecasting*.
Doornik, J. A., & Hendry, D. F. (2015). Statistical model selection with "big data". *Cogent Economics & Finance*, *3*(1).
Doucet, A. N. d. F., & Gordon, N. J. (2001). *Sequential Monte Carlo methods in practice*. New York: Springer Verlag.
Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2010). Evaluating the goodness of fit of stochastic mortality model. *Insurance: Mathematics & Economics*, *47*(3), 255–265.
Draper, D., & Krnjajić, M. (2013). Calibration results for Bayesian model specification. *Tech. rep.*, Department of Applied Mathematics and Statistics, University of California.
Dudek, G. (2013). Forecasting time series with multiple seasonal cycles using neural networks with local learning. In *International conference on artificial intelligence and soft computing* (pp. 52–63). Springer.
Dudek, G. (2015). Generalized regression neural network for forecasting time series with multiple seasonal cycles. In *Intelligent systems'2014* (pp. 839–846). Springer.
Dudek, G. (2016). Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, *32*(3), 1057–1060.
Duncan, O. D., & Davis, B. (1953). An alternative to ecological correlation. *American Sociological Review*, *18*, 665–666.
Dungey, M., Martin, V. L., Tang, C., & Tremayne, A. (2020). A threshold mixed count time series model: estimation and application. *Studies in Nonlinear Dynamics & Econometrics*, *24*(2).
Dunis, C. L., Laws, J., & Sermpinis, G. (2010). Modelling and trading the EUR/USd exchange rate at the ECB fixing. *The European Journal of Finance*, *16*(6), 541–560.
Dunn, D. M., Williams, W. H., & Dechaine, T. L. (1976). Aggregate versus subaggregate models in local area forecasting. *Journal of the American Statistical Association*, *71*(353), 68–71.
Durante, F., & Sempi, C. (2015). *Principles of copula theory*. CRC Press.
Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford: Oxford University Press.
e Silva, E. G. d. S., Legey, L. F., & e Silva, E. A. d. S. (2010). Forecasting oil price trends using wavelets and hidden Markov models. *Energy Economics*, *32*(6), 1507–1519.
Easingwood, C. J., Mahajan, V., & Muller, E. (1983). A nonuniform influence innovation diffusion model of new product acceptance. *Marketing Science*, *2*(3), 273–295.
Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: Attitudes toward decision-making strategies. *Journal of Behavioral Decision Making*, *25*(5), 458–468.
Eaves, A. H. C., & Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, *55*(4), 431–437.
Eberhardt, M. (2012). Estimating panel time-series models with heterogeneous slopes. *The Stata Journal*, *12*(1), 61–71.
Economou, T., Stephenson, D. B., Rougier, J. C., Neal, R. A., & Mylne, K. R. (2016). On the use of Bayesian decision theory for issuing natural hazard warnings. *Proceedings of the Royal Society: Mathematical, Physical, and Engineering Sciences*, *472*(2194), Article 20160295.
Edge, R. M., & Gürkaynak, R. (2010). How useful are estimated DSGE model forecasts for central bankers? *Brookings Papers on Economic Activity*, *41*(2 (Fall)), 209–259.
Edwards, D. G., & Hsu, J. C. (1983). Multiple comparisons with the best treatment. *Journal of the American Statistical Association*, *78*(384), 965–971.
Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26.
Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 54–75.
Eggleton, I. R. C. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research*, *20*(1), 68–102.
Ehsani, S., & Linnainmaa, J. T. (2020). Factor momentum and the momentum factor. SSRN:3014521.
Eichenbaum, M., Johannsen, B. K., & Rebelo, S. (2017). Monetary policy and the predictability of nominal exchange rates. *NBER Working Papers 23158*, National Bureau of Economic Research, Inc.
Eksoz, C., Mansouri, S. A., Bourlakis, M., & Önkal, D. (2019). Judgmental adjustments through supply integration for strategic partnerships in food chains. *Omega*, *87*, 20–33.

El Balghiti, O., Elmachtoub, A. N., Grigas, P., & Tewari, A. (2019). Generalization bounds in the predict-then-optimize framework. In *Advances in neural information processing systems* (pp. 14412–14421).
El-Hendawi, M., & Wang, Z. (2020). An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting. *Electric Power Systems Research*, *182*, Article 106265.
Elangasinghe, M. A., Singhal, N., Dirks, K. N., Salmond, J. A., & Samarasinghe, S. (2014). Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modelling and k-means clustering. *Atmospheric Enviroment*, *94*, 106–116.
Elliott, G. (2015). Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics & Control*, *54*, 86–111.
Elliott, G., Timmermann, A., & Komunjer, I. (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies*, *72*(4), 1107–1125.
Elliott, M. R., & Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, *32*(2), 249–264.
Ellison, J., Dodd, E., & Forster, J. J. (2020). Forecasting of cohort fertility under a hierarchical Bayesian approach. *Journal of the Royal Statistical Society. Series A*, *183*(3), 829–856.
Elmachtoub, A. N., & Grigas, P. (2017). Smart "predict, then optimize". arXiv:1710.08005.
Elsbach, K. D., & Elofson, G. (2000). How the packaging of decision explanations affects perceptions of trustworthiness. *Academy of Management Journal*, *43*(1), 80–89.
Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*. Springer Science & Business Media.
Engel, C., Lee, D., Liu, C., Liu, C., & Wu, S. P. Y. (2019). The uncovered interest parity puzzle, exchange rate forecasting, and taylor rules. *Journal of International Money and Finance*, *95*, 317–331.
Engel, C., Mark, N. C., & West, K. D. (2008). Exchange rate models are not as bad as you think. In D. Acemoglu, K. Rogoff, & M. Woodford (Eds.), *NBER chapters*: *vol. 22*, *NBER macroeconomics annual 2007* (pp. 381–441). National Bureau of Economic Research, Inc.
Engelberg, J., Manski, C. F., & Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, *27*(1), 30–41.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987.
Engle, R. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, *20*(3), 339–350.
Engle, R. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review*, *94*(3), 405–420.
Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics*, *95*(3), 776–797.
Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Economic Theory*, *11*(1), 122–150.
Engle, R. F., & Russell, J. R. (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Empirical Finance*, *4*(2), 187–212.
Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, *66*(5), 1127–1162.
Erikson, R. S., & Wlezien, C. (2012). Markets vs. polls as election predictors: An historical assessment. *Electoral Studies*, *31*(3), 532–539.
European Banking Federation (2019). EBF position paper on AI in the banking industry. EBF_037419.
Evans, M. D. R. (1986). American fertility patterns: A comparison of white and nonwhite cohorts Born 1903-56. *Population and Development Review*, *12*(2), 267–293.
Fahimnia, B., Sanders, N., & Siemsen, E. (2020). Human judgment in supply chain forecasting. *Omega*, *94*, Article 102249.
Fair, R. C. (1978). The effect of economic events on votes for president. *The Review of Economics and Statistics*, *60*(2), 159–173.
Fan, S., Chen, L., & Lee, W.-J. (2008). Machine learning based switching model for electricity load forecasting. *Energy Conversion & Management*, *49*(6), 1331–1344.
Fan, S., Mao, C., & Chen, L. (2006). Electricity peak load forecasting with self-organizing map and support vector regression. *IEEJ Transactions on Electrical and Electronic Engineering*, *1*(3), xxxi.
Fan, Y., Nowaczyk, S., & Röognvaldsson, T. (2020). Transfer learning for remaining useful life prediction based on consensus self-organizing models. *Reliability Engineering & System Safety*, *203*, Article 107098.
Fan, Y., & Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *75*(3), 531–552.
Fan, J., & Yao, Q. (2005). *Springer series in statistics*, *Nonlinear time series: nonparametric and parametric methods* (p. 576). New York: Springer.
Faraji, J., Ketabi, A., Hashemi-Dezaki, H., Shafie-Khah, M., & Catalão, J. P. S. (2020). Optimal day-ahead self-scheduling and operation of prosumer microgrids using hybrid machine learning-based weather and load forecasting. *IEEE Access*, *8*, 157284–157305.
Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, *460*(7256), 685–686.
Faust, J., & Wright, J. H. (2013). Forecasting inflation. In G. Elliott, & A. Timmermann (Eds.), *Handbook of economic forecasting*: *vol. 2*, (pp. 2–56). Elsevier.
Favero, C. A., & Marcellino, M. (2005). Modelling and forecasting fiscal variables for the euro area. *Oxford Bulletin of Economics and Statistics*, *67*, 755–783.
Fernandes, M., de Sá Mota, B., & Rocha, G. (2005). A multivariate conditional autoregressive range model. *Economics Letters*, *86*(3), 435–440.
Fernández-Villaverde, J., & Guerrón-Quintana, P. A. (2020). Estimating DSGE models: Recent advances and future challenges. *Working Paper 27715*, National Bureau of Economic Research.
Fezzi, C., & Mosetti, L. (2020). Size matters: Estimation sample length and electricity price forecasting accuracy. *The Energy Journal*, *41*(4).
Fifić, M., & Gigerenzer, G. (2014). Are two interviewers better than one? *Journal of Business Research*, *67*(8), 1771–1779.
Figlewski, S., & Wachtel, P. (1981). The formation of inflationary expectations. *The Review of Economics and Statistics*, *63*(1), 1–10.
Figlewski, S., & Wachtel, P. (1983). Rational expectations, informational efficiency, and tests using survey data: A reply. *The Review of Economics and Statistics*, *65*(3), 529–531.
Fildes, R. (2017). Research into forecasting practice. *Foresight: The International Journal of Applied Forecasting*, *44*, 39–46.
Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, *37*(6), 570–576.
Fildes, R., & Goodwin, P. (2013). Forecasting support systems: What we know, what we need to know. *International Journal of Forecasting*, *29*(2), 290–294.
Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, *42*(1), 351–361.
Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3–23.
Fildes, R., Goodwin, P., & Önkal, D. (2019a). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, *35*(1), 144–156.
Fildes, R., Ma, S., & Kolassa, S. (2019b). Retail forecasting: research and practice. *International Journal of Forecasting*.
Fildes, R., & Petropoulos, F. (2015). Improving forecast quality in practice. *Foresight: The International Journal of Applied Forecasting*, *36*(Winter), 5–12.
Filippou, I., Rapach, D. E., Taylor, M. P., & Zhou, G. (2020). Exchange rate prediction with machine learning and a smart carry trade portfolio. SSRN:3455713.
Findley, D. F. (2005). Some recent developments and directions in seasonal adjustment. *Journal of Official Statistics*, *21*(2), 343.
Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, *16*(2), 127–152.
Fiori, F., Graham, E., & Feng, Z. (2014). Geographical variations in fertility and transition to second and third birth in britain. *Advances in Life Course Research*, *21*, 149–167.

Fiori, C., & Kovaka, M. (2005). Defining megaprojects: Learning from construction at the edge of experience. *Construction Research Congress 2005*, 1–10.
Fiorucci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F., & Koehler, A. B. (2016). Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting*, *32*(4), 1151–1161.
Fioruci, J. A., Pellegrini, T. R., Louzada, F., & Petropoulos, F. (2015). The optimised theta method. arXiv:1503.03529.
Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*, *43*(4), 557–572.
Fischhoff, B. (2007). An early history of hindsight research. *Social Cognition*, *25*(1), 10–13.
Fischhoff, B. (2012). Communicating uncertainty fulfilling the duty to inform. *Issues in Science and Technology*, *28*(4), 63–70.
Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences*, *111*(Supplement 4), 13664–13671.
Fisher, J. C., & Pry, R. H. (1971). A simple substitution model of technological change. *Technological Forecasting and Social Change*, *3*, 75–88.
Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, *111*(Supplement 4), 13593–13597.
Fissler, T., Frongillo, R., Hlavinová, J., & Rudloff, B. (2020). Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals. arXiv:1910.07912.
Fiszeder, P. (2005). Forecasting the volatility of the polish stock index – WIG20. In W. Milo, & P. Wdowiński (Eds.), *Forecasting financial markets. theory and applications* (pp. 29–42). Wydawnictwo Uniwersytetu Łódzkiego.
Fiszeder, P. (2018). Low and high prices can improve covariance forecasts: The evidence based on currency rates. *Journal of Forecasting*, *37*(6), 641–649.
Fiszeder, P., & Fałdziński, M. (2019). Improving forecasts with the co-range dynamic conditional correlation model. *Journal of Economic Dynamics and Control*, *108*, Article 103736.
Fiszeder, P., Fałdziński, M., & Molnár, P. (2019). Range-based DCC models for covariance and value-at-risk forecasting. *Journal of Empirical Finance*, *54*, 58–76.
Fiszeder, P., & Perczak, G. (2013). A new look at variance estimation based on low, high and closing prices taking into account the drift. *Statistica Neerlandica*, *67*(4), 456–481.
Fiszeder, P., & Perczak, G. (2016). Low and high prices can improve volatility forecasts during periods of turmoil. *International Journal of Forecasting*, *32*(2), 398–410.
Fixler, D. J., & Grimm, B. T. (2005). Reliability of the NIPA estimates of U.S. economic activity. *Survey of Current Business*, *85*, 9–19.
Fixler, D. J., & Grimm, B. T. (2008). The reliability of the GDP and GDI estimates. *Survey of Current Business*, *88*, 16–32.
Fliedner, G. (2003). CPFR: an emerging supply chain tool. *Industrial Management & Data Systems*, *103*(1), 14–21.
Flyvbjerg, B. (2007). Policy and planning for large-infrastructure projects: Problems, causes, cures. *Environment and Planning: B, Planning & Design*, *34*(4), 578–597.
Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. (2003). *Megaprojects and risk: an anatomy of ambition*. Cambridge University Press.
Forcina, A., & Pellegrino, D. (2019). Estimation of voter transitions and the ecological fallacy. *Quality & Quantity*, *53*(4), 1859–1874.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*, *50*(6), 1243–1255.
Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of english football. *International Journal of Forecasting*, *21*, 551–564.
Forrest, D., & Simmons, R. (2006). New issues in attendance demand: The case of the english football league. *Journal of Sports Economics*, *7*(3), 247–263.
Fortsch, S. M., & Khapalova, E. A. (2016). Reducing uncertainty in demand for blood. *Operations Research for Health Care*, *9*, 16–28.
Fortuin, L. (1984). Initial supply and re-order level of new service parts. *European Journal of Operational Research*, *15*(3), 310–319.
Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, *23*, 272–288.
Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *34*(3), 350–363.
Frankel, J., & Schreger, J. (2013). Over-optimistic official forecasts and fiscal rules in the eurozone. *Review of World Economics*, *149*, 247–272.
Franses, P. H. (1991). Seasonality, non-stationarity and the forecasting of monthly time series. *International Journal of Forecasting*, *7*(2), 199–208.
Franses, P. H., van Dijk, D., & Opschoor, A. (2014). *Time series models for business and economic forecasting*. Cambridge University Press.
Franses, P. H., & Ghijsels, H. (1999). Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting*, *15*(1), 1–9.
Franses, P. H., & Legerstee, R. (2009a). Do experts' adjustments on model-based SKU-level forecasts improve forecast quality? *Journal of Forecasting*, *36*.
Franses, P. H., & Legerstee, R. (2009b). Properties of expert adjustments on model-based SKU-level forecasts. *International Journal of Forecasting*, *25*(1), 35–47.
Franses, P. H., & Legerstee, R. (2009c). A unifying view on multi-step forecasting using an autoregression. *Journal of Economic Surveys*, *24*(3), 389–401.
Frazier, D. T., Loaiza-Maya, R., Martin, G. M., & Koo, B. (2021). Loss-based variational Bayes prediction. arXiv:2104.14054.
Frazier, D. T., Maneesoonthorn, W., Martin, G. M., & McCabe, B. P. (2019). Approximate Bayesian forecasting. *International Journal of Forecasting*, *35*(2), 521–539.
Frechtling, D. C. (2001). *Forecasting tourism demand: methods and strategies*. Routledge.
Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, *9*(6), 1218–1228.
Freedman, D. A., Klein, S. P., Ostland, M., & Roberts, M. (1998). Review of 'a solution to the ecological inference problem'. *Journal of the American Statistical Association*, *93*(444), 1518–1522.
Freeland, K., & McCabe, B. P. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting*, *20*(3), 427–434.
Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, *33*, 2326–2377.
Friedman, J. A. (2015). Using power laws to estimate conflict size. *The Journal of Conflict Resolution*, *59*(7), 1216–1241.
Fry, C., & Brundage, M. (2020). The M4 forecasting competition – a practitioner's view. *International Journal of Forecasting*, *36*(1), 156–160.
Fuhrer, J. C. (2018). *Intrinsic expectations persistence: Evidence from professional and household survey expectations*: *Working Papers 18–9*, Federal Reserve Bank of Boston.
Fulcher, B. D., & Jones, N. S. (2014). Highly comparative feature-based time-series classification. *IEEE Transactions on Knowledge and Data Engineering*, *26*(12), 3026–3037.
Fulcher, B. D., Little, M. A., & Jones, N. S. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society Interface*, *10*(83), Article 20130048.
Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, *2*(3), 183–192.
Furlan, C., & Mortarino, C. (2018). Forecasting the impact of renewable energies in competition with non-renewable sources. *Renewable and Sustainable Energy Reviews*, *81*, 1879–1886.
Furlan, C., Mortarino, C., & Zahangir, M. S. (2020). Interaction among three substitute products: An extended innovation diffusion model. *Statistical Methods & Applications*, *in press*.
Gaddis, J. L. (1989). *The long peace: Inquiries into the history of the cold war*. The Long Peace: Inquiries Into the History of the Cold War.
Gaillard, P., Goude, Y., & Nedellec, R. (2016). Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting*, *32*(3), 1038–1050.
Galbreth, M. R., Kurtuluş, M., & Shor, M. (2015). How collaborative forecasting can reduce forecast accuracy. *Operations Research Letters*, *43*(4), 349–353.

Gali, J. (2008). *Monetary policy, inflation, and the business cycle: An introduction to the new keynesian framework*. Princeton and Oxford: Princeton University Press.
Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., & Martínez-Álvarez, F. (2019). Multi-step forecasting for big data time series based on ensemble learning. *Knowledge-Based Systems*, *163*, 830–841.
Galicia, A., Torres, J. F., Martínez-Álvarez, F., & Troncoso, A. (2018). A novel Spark-based multi-step forecasting algorithm for big data time series. *Information Sciences*, *467*, 800–818.
Galvão, A. B. (2017). Data revisions and DSGE models. *Journal of Econometrics*, *196*(1), 215–232.
Galvão, A. B., Giraitis, L., Kapetanios, G., & Petrova, K. (2016). A time varying DSGE model with financial frictions. *Journal of Empirical Finance*, *38*, 690–716.
Gamble, C., & Gao, J. (2018). Safety-first AI for autonomous data centre cooling and industrial control. In *DeepMind*. Accessed on 2020-09-01, https://deepmind.com/blog/article/safety-first-ai-autonomous-data-centre-cooling-and-industrial-control.
Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, *5*(2), 79–141.
García, F. P., Pedregal, D. J., & Roberts, C. (2010). Time series methods applied to failure prediction and detection. *Reliability Engineering & System Safety*, *95*(6), 698–703.
Garcia, R., & Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, *78*(1), 111–125.
Gardner, E. S. (1985). Exponential smoothing: The state of the art. In *ICT Monograph*: *Journal of Forecasting*, In *ICT Monograph*: *4*(1).1–28,
Gardner, E. S. (2006). Exponential smoothing: The state of the art - part II. *International Journal of Forecasting*, *22*(4), 637–666.
Gardner, E., & Koehler, A. B. (2005). Comments on a patented bootstrapping method for forecasting intermittent demand. *International Journal of Forecasting*, *21*(3), 617–618.
Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, *53*(1), 67–78.
Garratt, A., Lee, K., Mise, E., & Shields, K. (2008). Real time representations of the output gap. *The Review of Economics and Statistics*, *90*, 792–804.
Gartner, W. B., & Thomas, R. J. (1993). Factors affecting new product forecasting accuracy in new firms. *Journal of Product Innovation Management*, *10*(1), 35–52.
Gasthaus, J., Benidis, K., Wang, Y., Rangapuram, S. S., Salinas, D., Flunkert, V., et al. (2019).
Gebicki, M., Mooney, E., Chen, S.-J. G., & Mazur, L. M. (2014). Evaluation of hospital medication inventory policies. *Health Care Management Science*, *17*(3), 215–229.
Gelman, A., Park, D. K., Ansolabehere, S., Price, P. N., & Minnite, L. C. (2001). Models, assumptions and model checking in ecological regressions. *Journal of the Royal Statistical Society, Series A*, *164*(1), 101–118.
Gelper, S., Fried, R., & Croux, C. (2009). Robust forecasting with exponential and holt-winters smoothing. *Journal of Forecasting*, *11*.
Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751.
George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*(423), 881–890.
Gerlach, R., Chen, C. W. S., Lin, D. S. Y., & Huang, M.-H. (2006). Asymmetric responses of international stock markets to trading volume. *Physica A: Statistical Mechanics and its Applications*, *360*(2), 422–444.
Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., et al. (2014). World population stabilization unlikely this century. *Science*, *346*(6206), 234–237.
Geweke, J. (1977). The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models*. North-Holland.
Geweke, J. (2001). BayesIan econometrics and forecasting. *Journal of Econometrics*, *100*(1), 11–15.
Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.
Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, *164*(1), 130–141.
Geweke, J., Koop, G., & van Dijk, H. (2011). *The oxford handbook of bayesian econometrics*. OUP.
Geweke, J., & Whiteman, C. (2006). BayesIan forecasting. *The Handbook of Economic Forecasting*, *1*, 3–98.
Gharbi, M., Quenel, P., Gustave, J., Cassadou, S., La Ruche, G., Girdary, L., et al. (2011). Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC Infectious Diseases*, *11*(1), 1–13.
Ghassemi, M., Pimentel, M. A., Naumann, T., Brennan, T., Clifton, D. A., Szolovits, P., et al. (2015). A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data. *2015*, In *Proceedings of the AAAI conference on artificial intelligence. aaai conference on artificial intelligence* (p. 446). NIH Public Access.
Ghysels, E., Lee, H. S., & Noh, J. (1994). Testing for unit roots in seasonal time series: Some theoretical extensions and a Monte Carlo investigation. *Journal of Econometrics*, *62*(2), 415–442.
Ghysels, E., Plazzi, A., Valkanov, R., & Torous, W. (2013). Forecasting real estate prices. In G. Elliott, & A. Timmermann (Eds.), *vol. 2*, *Handbook of economic forecasting* (pp. 509–580). Elsevier.
Giacomini, R., & Rossi, B. (2016). Model comparisons in unstable environments. *International Economic Review*, *57*(2), 369–392.
Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.
Giani, A., Bitar, E., Garcia, M., McQueen, M., Khargonekar, P., & Poolla, K. (2013). Smart grid data integrity attacks. *IEEE Transactions on Smart Grid*, *4*(3), 1244–1253.
Giannone, D. L., & Primiceri, G. M. (2017). *Macroeconomic prediction with big data: The illusion of sparsity*. The Fedral Reserve Bank of New York.
Gias, A. U., & Casale, G. (2020). COCOA: COld start aware capacity planning for function-as-a-service platforms. ArXiv:2007.01222.
Giebel, G., & Kariniotakis, G. (2017). Wind power forecasting—a review of the state of the art. In *Renewable energy forecasting* (pp. 59–109). Elsevier.
Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592–596.
Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. Viking.
Gil, R. G. R., & Levitt, S. D. (2007). Testing the efficiency of markets in the 2002 World Cup. *The Journal of Prediction Markets*, *1*(3), 255–270.
Gil-Alana, L. (2001). A fractionally integrated exponential model for UK unemployment. *Journal of Forecasting*, *20*(5), 329–340.
Gilbert, K. (2005). An ARIMA supply chain model. *Management Science*, *51*(2), 305–310.
Gilbert, C., Browell, J., & McMillan, D. (2020a). Leveraging turbine-level data for improved probabilistic wind power forecasting. *IEEE Transactions on Sustainable Energy*, *11*(3), 1152–1160.
Gilbert, C., Browell, J., & McMillan, D. (2020b). Probabilistic access forecasting for improved offshore operations. *International Journal of Forecasting*.
Gilliland, M. (2002). Is forecasting a waste of time? *Supply Chain Management Review*, *6*(4), 16–23.
Gilliland, M. (2010). *The business forecasting deal: Exposing myths, eliminating bad practices, providing practical solutions*. John Wiley & Sons.
Giraitis, L., Kapetanios, G., & Price, S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics*, *177*(2), 153–170.
Givon, M., Mahajan, W., & Müller, E. (1995). Software piracy: Estimation of the lost sales and the impact on software diffusion. *Journal of Marketing*, *59*, 29–37.
Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, *11*(8), 1203–1211.
Gleditsch, R. F., & Syse, A. (2020). *Ways to project fertility in Europe. Perceptions of current practices and outcomes*: *Tech. Rep. 929*, Statistics Norway, Research Department.
Glocker, C., & Wegmüller, P. (2018). International evidence of time-variation in trend labor productivity growth. *Economics Letters*, *167*, 115–119.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.
Glynn, A., & Wakefield, J. (2010). Ecological inference in the social sciences. *Statistical Methodology*, *7*(3), 307–322.
Gneiting, T. (2011a). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.
Gneiting, T. (2011b). Quantiles as optimal point forecasts. *International Journal of Forecasting*, *27*(2), 197–207.
Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *69*, 243–268.
Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*, 125–151.
Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.
Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, *133*(5), 1098–1118.
Gneiting, T., & Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, *7*, 1747–1782.
Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., & Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds (with discussion and rejoinder). *Test*, *17*, 211–264.
Godahewa, R., Deng, C., Prouzeau, A., & Bergmeir, C. (2020). Simulation and optimisation of air conditioning systems using machine learning. ArXiv:2006.15296.
Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-scale sentiment analysis for news and blogs. *ICWSM*, *7*(21), 219–222.
Godet, M. (1982). From forecasting to 'la prospective' a new way of looking at futures. *Journal of Forecasting*, *1*(3), 293–301.
Goh, T. N., & Varaprasad, N. (1986). A statistical methodology for the analysis of the life-cycle of reusable containers. *IIE Transactions*, *18*(1), 42–47.
Goia, A., May, C., & Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*, *26*(4), 700–711.
Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, *10*(1), 1–309.
Goldstein, J. S. (2011). *Winning the war on war: The decline of armed conflict worldwide*. Penguin.
Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review*, *109*(1), 75–90.
Golestaneh, F., Pinson, P., & Gooi, H. B. (2019). Polyhedral predictive regions for power system applications. *IEEE Transactions on Power Systems*, *34*(1), 693–704.
Goltsos, T., & Syntetos, A. (2020). Forecasting for remanufacturing. *Foresight: The International Journal of Applied Forecasting*, *56*, 10–17.
Goltsos, T. E., Syntetos, A. A., & van der Laan, E. (2019). Forecasting for remanufacturing: The effects of serialization. *Journal of Operations Management*, *65*(5), 447–467.
Gomez Munoz, C. Q., De la Hermosa Gonzalez-Carrato, R. R., Trapero Arenas, J. R., & Garcia Marquez, F. P. (2014). A novel approach to fault detection and diagnosis on wind turbines. *GlobalNEST International Journal*, *16*(6), 1029–1037.
Gonçalves, R. (2015). Minimizing symmetric mean absolute percentage error (SMAPE). Cross Validated URL: https://stats.stackexchange.com/q/145490 (version: 2016-04-15).
Gonçalves, C., Bessa, R. J., & Pinson, P. (2021a). A critical overview of privacy-preserving approaches for collaborative forecasting. *International Journal of Forecasting*, *37*(1), 322–342.
Gonçalves, C., Pinson, P., & Bessa, R. J. (2021b). Towards data markets in renewable energy forecasting. *IEEE Transactions on Sustainable Energy*, *12*(1), 533–542.
Gönül, M. S., Önkal, D., & Goodwin, P. (2009). Expectations, use and judgmental adjustment of external financial and economic forecasts: an empirical investigation. *Journal of Forecasting*, *28*(1), 19–37.
Gönül, M. S., Önkal, D., & Goodwin, P. (2012). Why should I trust your forecasts? *Foresight: The International Journal of Applied Forecasting*, *27*, 5–9.
Gönül, M. S., Önkal, D., & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems*, *42*(3), 1481–1493.
Goodman, L. A. (1953). Ecological regressions and behavior of individuals. *American Sociological Review*, *18*, 663–664.
Goodman, L. A. (1959). Some alternatives to ecological correlation. *The American Journal of Sociology*, *64*(6), 610–625.
Goodwin, P. (2000a). Correct or combine? Mechanically integrating judgmental forecasts with statistical methods. *International Journal of Forecasting*, *16*(2), 261–275.
Goodwin, P. (2000b). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, *16*(1), 85–99.
Goodwin, P. (2002). Integrating management judgment and statistical methods to improve short-term forecasts. *Omega*, *30*(2), 127–135.
Goodwin, P. (2014). Getting real about uncertainty. *Foresight: The International Journal of Applied Forecasting*, *33*, 4–7.
Goodwin, P., Dyussekeneva, K., & Meeran, S. (2013). The use of analogies in forecasting the annual sales of new electronics products. *IMA Journal of Management Mathematics*, *24*(4), 407–422.
Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy? *Journal of Behavioural Decision Making*, *12*(1), 37–53.
Goodwin, P., Fildes, R., Lawrence, M., & Nikolopoulos, K. (2007). The process of using a forecasting support system. *International Journal of Forecasting*, *23*(3), 391–404.
Goodwin, P., Fildes, R., Lawrence, M., & Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega*, *39*(3), 242–253.
Goodwin, P., Gönül, M. S., & Önkal, D. (2013). Antecedents and effects of trust in forecasting advice. *International Journal of Forecasting*, *29*(2), 354–366.
Goodwin, P., Gönül, M. S., & Önkal, D. (2019a). When providing optimistic and pessimistic scenarios can be detrimental to judgmental demand forecasts and production decisions. *European Journal of Operational Research*, *273*(3), 992–1004.
Goodwin, P., Gönül, M. S., Önkal, D., Kocabıyıkoğlu, A., & Göğüş, I. (2019b). Contrast effects in judgmental forecasting when assessing the implications of worst- and best-case scenarios. *Journal of Behavioral Decision Making*, *32*(5), 536–549.
Goodwin, P., Petropoulos, F., & Hyndman, R. J. (2017). A note on upper bounds for forecast-value-added relative to naïve forecasts. *Journal of the Operational Research Society*, *68*(9), 1082–1084.
Goodwin, P., & Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological Forecasting and Social Change*, *77*(3), 355–368.
Google code (2013). The Word2Vec project. Accessed on 2020-09-05, https://code.google.com/archive/p/word2vec.
Gorbey, S., James, D., & Poot, J. (1999). Population forecasting with endogenous migration: An application to trans-tasman migration. *International Regional Science Review*, *22*(1), 69–101.
Gordon, R. J. (2003). Exploding productivity growth: context, causes, and implications. *Brookings Papers on Economic Activity*, *2003*(2), 207–298.
Gospodinov, N. (2005). Testing for threshold nonlinearity in short-term interest rates. *Journal of Financial Econometrics*, *3*(3), 344–371.
Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., & Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, *191*(1), 207–222.
Goyal, A., & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.
Graefe, A. (2014). Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly*, *78*(1), 204–232.
Graefe, A., & Armstrong, J. S. (2011). Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting*, *27*(1), 183–195.
Graefe, A., Armstrong, J. S., Jones Jr, R. J., & Cuzán, A. G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, *30*(1), 43–54.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Granger, C. W. J., & Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *38*(2), 189–203.

Granger, C. W. J., & Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, *19*, 537–560.

Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*(2), 197–204.

Granger, C. W. J., & Swanson, N. (1996). Future developments in the study of cointegrated variables. *Oxford Bulletin of Economics and Statistics*, *58*(3), 537–553.

Graves, S. C. (1999). A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management*, *1*(1), 50–61.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, *42*(1), 27–62.

Gray, J. (2015a). *Heresies: Against Progress and Other Illusions*. Granta Books.

Gray, J. (2015b). Steven pinker is wrong about violence and war. In *the Guardian*. Accessed on 2018-05-02, http://www.theguardian.com/books/2015/mar/13/john-gray-steven-pinker-wrong-violence-war-declining.

Gray, C. W., Barnes, C. B., & Wilkinson, E. F. (1965). The process of prediction as a function of the correlation between two scaled variables. *Psychonomic Science*, *3*(1), 231.

Green, K. C., & Armstrong, J. S. (2007). Structured analogies for forecasting. *International Journal of Forecasting*, *23*(3), 365–376.

Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, *68*(8), 1678–1685.

Green, J., Hand, J. R. M., & Zhang, X. F. (2017). The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies*, *30*(12), 4389–4436.

Greenberg, E. (2008). *Introduction to bayesian econometrics*. CUP.

Greiner, D. J. (2007). Ecological inference in voting rights act disputes: Where are we now, and where do we want to be? *Jurimetrics*, *47*(2), 115–167.

Greiner, D. J., & Quinn, K. M. (2010). Exit polling and racial bloc voting: combining individual-level and rxc ecological data. *The Annals of Applied Statistics*, *4*(4), 1774–1796.

Gresnigt, F., Kole, E., & Franses, P. H. (2015). Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes. *Journal of Banking & Finance*, *56*, 123–139.

Gresnigt, F., Kole, E., & Franses, P. H. (2017a). Exploiting spillovers to forecast crashes. *Journal of Forecasting*, *36*(8), 936–955.

Gresnigt, F., Kole, E., & Franses, P. H. (2017b). Specification testing in hawkes models. *Journal of Financial Econometrics*, *15*(1), 139–171.

Gromenko, O., Kokoszka, P., & Reimherr, M. (2017). Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *79*(1), 29–50.

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, *9*(3), 233–254.

Grossi, L., & Nan, F. (2019). Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. *Technological Forecasting and Social Change*, *141*, 305–318.

Grushka-Cockayne, Y., & Jose, V. R. R. (2020). Combining prediction intervals in the M4 competition. *International Journal of Forecasting*, *36*(1), 178–185.

Grushka-Cockayne, Y., Jose, V. R. R., & Lichtendahl, K. C. (2017). Ensembles of overfit and overconfident forecasts. *Management Science*, *63*(4), 1110–1130.

Grushka-Cockayne, Y., Lichtendahl, K. C., Jose, V. R. R., & Winkler, R. L. (2017). Quantile evaluation, sensitivity to bracketing, and sharing business payoffs. *Operations Research*, *65*(3), 712–728.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, *33*(5), 2223–2273.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, *12*(1), 37–48.

Guidolin, M., & Alpcan, T. (2019). Transition to sustainable energy generation in Australia: Interplay between coal, gas and renewables. *Renewable Energy*, *139*, 359–367.

Guidolin, M., & Guseo, R. (2015). Technological change in the U.S. music industry: Within-product, cross–product and churn effects between competing blockbusters. *Technological Forecasting and Social Change*, *99*, 35–46.

Guidolin, M., & Guseo, R. (2016). The german energy transition: Modeling competition and substitution between nuclear power and renewable energy technologies. *Renewable and Sustainable Energy Reviews*, *60*, 1498–1504.

Guidolin, M., & Guseo, R. (2020). Has the iPhone cannibalized the iPad? an asymmetric competition model. *Applied Stochastic Models in Business and Industry*, *36*, 465–476.

Guidolin, M., & Mortarino, C. (2010). Cross-country diffusion of photovoltaic systems: modelling choices and forecasts for national adoption patterns. *Technological Forecasting and Social Change*, *77*(2), 279–296.

Guidolin, M., & Pedio, M. (2018). *Essentials of time series for financial applications*. Academic Press.

Guidolin, M., & Pedio, M. (2019). Forecasting and trading monetary policy effects on the riskless yield curve with regime switching Nelson–Siegel models. *Journal of Economic Dynamics & Control*, *107*, Article 103723.

Guidolin, M., & Thornton, D. L. (2018). Predictions of short-term rates and the expectations hypothesis. *International Journal of Forecasting*, *34*(4), 636–664.

Guidolin, M., & Timmermann, A. (2006). Term structure of risk under alternative econometric specifications. *Journal of Econometrics*, *131*(1), 285–308.

Guidolin, M., & Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, *150*(2), 297–311.

Gumus, M., & Kiran, M. S. (2017). Crude oil price forecasting using XGBoost. In *2017 International conference on computer science and engineering (UBMK)* (pp. 1100–1103). IEEE.

Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of Tourism Research*, *61*, 199–212.

Gunter, U., Önder, I., & Gindl, S. (2019). Exploring the predictive ability of LIKES of posts on the facebook pages of four major city DMOs in Austria. *Tourism Economics*, *25*(3), 375–401.

Guo, X., Grushka-Cockayne, Y., & De Reyck, B. (2020). Forecasting airport transfer passenger flow using real-time data and machine learning. *Manufacturing & Service Operations Management*.

Gupta, S. (1994). Managerial judgment and forecast combination: An experimental study. *Marketing Letters*, *5*(1), 5–17.

Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2013). Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *26*(9), 2250–2267.

Gürkaynak, R. S., Kısacıkoğlu, B., & Rossi, B. (2013). Do DSGE models forecast more accurately out-of-sample than VAR models? *Advances in Econometrics,VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims*, *32*, 27–79.

Guseo, R. (2010). Partial and ecological correlation: a common three-term covariance decomposition. *Statistical Methods & Applications*, *19*(1), 31–46.

Guseo, R., & Guidolin, M. (2009). Modelling a dynamic market potential: A class of automata networks for diffusion of innovations. *Technological Forecasting and Social Change*, *76*, 806–820.

Guseo, R., & Guidolin, M. (2011). Market potential dynamics in innovation diffusion: Modelling the synergy between two driving forces. *Technological Forecasting and Social Change*, *78*, 13–24.

Guseo, R., & Mortarino, C. (2010). Correction to the paper "optimal product launch times in a Duopoly: Balancing life-cycle revenues with product cost". *Operations Research*, *58*, 1522–1523.

Guseo, R., & Mortarino, C. (2012). Sequential market entries and competition modelling in multi-innovation diffusions. *European Journal of Operational Research*, *216*, 658–667.

Guseo, R., & Mortarino, C. (2014). Within-brand and cross-brand word-of-mouth for sequential multi-innovation diffusions. *IMA Journal of Management Mathematics*, *25*, 287–311.

Guseo, R., & Mortarino, C. (2015). Modeling competition between two pharmaceutical drugs using innovation diffusion models. *The Annals of Applied Statistics*, *9*, 2073–2089.

Gutierrez, R. S., Solis, A. O., & Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, *111*(2), 409–420.

Gutterman, S., & Vanderhoof, I. T. (1998). Forecasting changes in mortality: A search for a law of causes and effects. *North American Actuarial Journal*, *2*(4), 135–138.
Haas, M., Mittnik, S., & Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, *2*(4), 493–530.
Hahn, M., Frühwirth-Schnatter, S., & Sass, J. (2010). Markov chain Monte Carlo methods for parameter estimation in multidimensional continuous time Markov switching models. *Journal of Financial Econometrics*, *8*(1), 88–121.
Hajnal, J. (1955). The prospects for population forecasts. *Journal of the American Statistical Association*, *50*(270), 309–322.
Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, *32*(2), 177–203.
Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, *23*(1), 1–13.
Hall, S. G., & Mitchell, J. (2009). Recent developments in density forecasting. In T. C. Mills, & K. Patterson (Eds.), *Palgrave handbook of econometrics, Volume 2: Applied econometrics* (pp. 199–239). Palgrave MacMillan.
Hamill, T. M., & Colucci, S. J. (1997). Verification of eta-RSM short-range ensemble forecasts. *Monthly Weather Review*, *125*(6), 1312–1327.
Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics & Control*, *12*(2), 385–423.
Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, *45*(1), 39–70.
Hamilton, J. D. (2016). Macroeconomic regimes and regime shifts. In J. B. Taylor, & H. Uhlig (Eds.), *vol. 2*, *Handbook of macroeconomics* (pp. 163–201). Elsevier.
Han, Y., He, A., Rapach, D. E., & Zhou, G. (2021). Expected stock returns and firm characteristics: E-LASSO, assessment, and implications. SSRN:3185335.
Han, P. K., Klein, W. M., Lehman, T. C., Massett, H., Lee, S. C., & Freedman, A. N. (2009). Laypersons' responses to the communication of uncertainty regarding cancer risk estimates. *Medical Decision Making*, *29*(3), 391–403.
Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Elsevier.
Han, W., Wang, X., Petropoulos, F., & Wang, J. (2019). Brain imaging and forecasting: Insights from judgmental model selection. *Omega*, *87*, 1–9.
Hand, D. J. (2009). Mining the past to determine the future - problems and possibilities. *International Journal of Forecasting*, *25*(3), 441–451.
Hanley, J. A., Joseph, L., Platt, R. W., Chung, M. K., & Belisle, P. (2001). Visualizing the median as the minimum-deviation location. *The American Statistician*, *55*(2), 150–152.
Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, *B, 41*, 190–195.
Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in US labour productivity. *Journal of Economic Perspectives*, *15*(4), 117–128.
Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380.
Harford, T. (2014). Big data: A big mistake? *Significance*, *11*, 14–19.
Harrell, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis* (2nd Ed). New York, USA: Springer.
Harris, D., Martin, G. M., Perera, I., & Poskitt, D. S. (2019). Construction and visualization of confidence sets for frequentist distributional forecasts. *Journal of Computational and Graphical Statistics*, *28*(1), 92–104.
Harris, R. D., & Yilmaz, F. (2010). Estimation of the conditional variance-covariance matrix of returns using the intraday range. *International Journal of Forecasting*, *26*(1), 180–194.
Hart, R., Hutton, J., & Sharot, T. (1975). A statistical analysis of association football attendances. *Applied Statistics*, *24*(1), 17–27.
Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.
Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, *63*(3), 247–263.
Harvey, N. (2007). Use of heuristics: Insights from forecasting research. *Thinking & Reasoning*, *13*(1), 5–24.
Harvey, N. (2011). Anchoring and adjustment: A Bayesian heuristic? In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on Thinking, Judging, and Decision Making* (pp. 98–108). Oslo: Universitetsforlaget.
Harvey, A. C. (2013). *Econometric Society Monographs*, *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Cambridge University Press.
Harvey, N. (2019). Commentary: Algorithmic aversion and judgmental wisdom. *Foresight: The International Journal of Applied Forecasting*, *54*, 13–14.
Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, *12*(1), 119–137.
Harvey, N., Bolger, F., & McClelland, A. (1994). On the nature of expectations. *British Journal of Psychology*, *85*(2), 203–229.
Harvey, D. I., Leybourne, S. J., & Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business & Economic Statistics*, *16*(2), 254–259.
Harvey, C. R., Liu, Y., & Zhu, H. (2016). … And the cross-section of expected returns. *Review of Financial Studies*, *29*(1), 5–68.
Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(2), 589–607.
Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, *50*(4), 1175–1199.
Hasni, M., Aguir, M. S., Babai, M. Z., & Jemai, Z. (2019a). On the performance of adjusted bootstrapping methods for intermittent demand forecasting. *International Journal of Production Economics*, *216*, 145–153.
Hasni, M., Aguir, M. S., Babai, M. Z., & Jemai, Z. (2019b). Spare parts demand forecasting: a review on bootstrapping methods. *International Journal of Productions Research*, *57*(15–16), 4791–4804.
Hassan, S., Arroyo, J., Galán Ordax, J. M., Antunes, L., & Pavón Mestras, J. (2013). Asking the oracle: Introducing forecasting principles into agent-based modelling. *Journal of Artificial Societies and Social Simulation*, *16*(3).
Hassani, H., & Silva, E. S. (2015). Forecasting with big data: A review. *Annals of Data Science*, *2*, 5–19.
Hastie, T. J., & Tibshirani, R. J. (1990). *vol. 43*, *Generalized additive models*. CRC Press.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer-Verlag GmbH.
Haugen, R. A. (2010). *The new finance, overreaction, complexity, and their consequences* (4th ed.). Pearson Education.
Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, *41*(3), 401–439.
Hawkes, A. G. (1969). An approach to the analysis of electoral swing. *Journal of the Royal Statistical Society, Series A*, *132*(1), 68–79.
Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *33*(3), 438–443.
Hawkes, A. G. (2018). Hawkes processes and their applications to finance: a review. *Quantitative Finance*, *18*(2), 193–198.
Hawkes, A. G., & Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, *11*(3), 493–503.
Hayes, B. (2002). Computing science: Statistics of deadly quarrels. *American Scientist*, *90*, 10–14.
He, A. W., Kwok, J. T., & Wan, A. T. (2010). An empirical model of daily highs and lows of west texas intermediate crude oil prices. *Energy Economics*, *32*(6), 1499–1506.
He, K., Yu, L., & Lai, K. K. (2012). Crude oil price analysis and forecasting using wavelet decomposed ensemble model. *Energy*, *46*(1), 564–574.
Hecht, R., & Gandhi, G. (2008). Demand forecasting for preventive AIDS vaccines. *Pharmacoeconomics*, *26*(8), 679–697.
Hedonometer (2020). Hedonometer word list. Accessed on 2020-09-05, https://hedonometer.org/words/labMT-en-v2/.
Heinrich, C. (2014). The mode functional is not elicitable. *Biometrika*, *101*(1), 245–251.

Heinrich, C. (2020). On the number of bins in a rank histogram. *Quarterly Journal of the Royal Meteorological Society*.
Heinrich, C., Hellton, K. H., Lenkoski, A., & Thorarinsdottir, T. L. (2020). Multivariate postprocessing methods for high-dimensional seasonal weather forecasts. *Journal of the American Statistical Association*.
Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, *107*, 49–80.
Hemri, S. (2018). Applications of postprocessing for hydrological forecasts. In *Statistical postprocessing of ensemble forecasts* (pp. 219–240). Elsevier.
Hemri, S., Lisniak, D., & Klein, B. (2015). Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research*, *51*(9), 7436–7451.
Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople's trust in experts in a digital age: The muenster epistemic trustworthiness inventory (METI). *PLoS One*, *10*(10), Article e0139309.
Hendry, D. F. (2001). Modelling UK inflation, 1875-1991. *Journal of Applied Econometrics*, *16*, 255–275.
Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics*, *135*(1–2), 399–426.
Hendry, D. F. (2010). Equilibrium-correction models. In *Macroeconometrics and time series analysis* (pp. 76–89). Springer.
Hendry, D. F. (Ed.), (2015). *Introductory macro-econometrics: A new approach*. London: Timberlake Consultants Press.
Hendry, D. F. (2020). *First-in, First-out: Modelling the UK's CO2 Emissions, 1860–2016*: *Working paper 2020-W02*, Oxford University: Nuffield College.
Hendry, D., & Clements, M. (2001). *Forecasting non-stationary economic time series*. Cambridge, Mass.: MIT Press.
Hendry, D. F., & Doornik, J. A. (2014). *Empirical model discovery and theory evaluation*. Cambridge MA: MIT Press.
Hendry, D. F., Johansen, S., & Santos, C. (2008a). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, *33*, 317–335, Erratum, 337–339.
Hendry, D. F., Johansen, S., & Santos, C. (2008b). Automatic selection of indicators in a fully saturated regression.. *Computational Statistics & Data Analysis*, *33*, 317–335.
Hendry, D. F., & Mizon, G. E. (2012). Open-model forecast-error taxonomies. In X. Chen, & N. R. Swanson (Eds.), *Recent advances and future directions in causality, prediction, and specification analysis* (pp. 219–240). Springer.
Herbst, E., & Schorfheide, F. (2016). *Bayesian estimation of DSGE models* (1st ed.). Princeton University Press.
Herrera, R., & González, N. (2014). The modeling and forecasting of extreme events in electricity spot markets. *International Journal of Forecasting*, *30*(3), 477–490.
Herrera, A. M., Hu, L., & Pastor, D. (2018). Forecasting crude oil price volatility. *International Journal of Forecasting*, *34*(4), 622–635.
Herron, M. C., & Shotts, K. W. (2004). Logical inconsistency in EI-based second-stage regressions. *American Journal of Political Science*, *48*(1), 172–183.
Hertzum, M. (2002). The importance of trust in software engineers' assessment and choice of information sources. *Information and Organization*, *12*(1), 1–18.
Hertzum, M. (2014). Expertise seeking: A review. *Information Processing & Management*, *50*(5), 775–795.
Hevia, C., Gonzalez-Rozada, M., Sola, M., & Spagnolo, F. (2015). Estimating and forecasting the yield curve using a Markov switching dynamic nelson and Siegel model. *Journal of Applied Economics*, *30*(6), 987–1009.
Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, *37*(1), 388–427.
Hii, Y. L., Zhu, H., Ng, N., Ng, L. C., & Rocklöv, J. (2012). Forecast of dengue incidence using temperature and rainfall. *PLoS Neglected Tropical Diseases*, *6*(11), Article e1908.
Hill, C. A., Zhang, G. P., & Miller, K. E. (2018). Collaborative planning, forecasting, and replenishment & firm performance: An empirical evaluation. *International Journal of Production Economics*, *196*, 12–23.
Hillebrand, E., & Medeiros, M. C. (2010). The benefits of bagging for forecast models of realized volatility. *Econometric Reviews*, *29*(5–6), 571–593.
Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. ArXiv:1207.0580.
Hinton, H. L., Jr. (1999). *Defence inventory, continuing challenger in managing inventories and avoiding adverse operational effects*: *Tech. rep.*, US General Accounting Office.
Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Elsevier.
Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, *16*(1), 44–55.
Hobijn, B., Franses, P. H., & Ooms, M. (2004). Generalizations of the KPSS-test for stationarity. *Statistica Neerlandica*, *58*(4), 483–502.
Hodges, P., Hogan, K., Peterson, J. R., & Ang, A. (2017). Factor timing with cross-sectional and time-series predictors. *Journal of Portfolio Management*, *44*(1), 30–43.
Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, 1–16.
Hoem, J. M., Madsen, D., Nielsen, J. L., Ohlsen, E. M., Hansen, H. O., & Rennermalm, B. (1981). Experiments in modelling recent danish fertility curves. *Demography*, *18*(2), 231–244.
Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayeslan model averaging: A tutorial (with discussion). *Statistical Science*, *214*, 382–417.
Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. *The International Journal of Logistics Management*, *29*(2), 739–766.
Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science*, *27*(2), 115–138.
Holly, S., Pesaran, M. H., & Yamagata, T. (2010). *Spatial and temporal diffusion of house prices in the U.K.*: *Discussion Papers 4694*, Institute of Labor Economics (IZA).
Hollyman, R., Petropoulos, F., & Tipping, M. E. (2021). Understanding forecast reconciliation. *European Journal of Operational Research*.
Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10.
Homburg, A., Weiß, C. H., Alwan, L. C., Frahm, G., & Göb, R. (2019). Evaluating approximate point forecasting of count processes. *Econometrics*, *7*(3), 1–28.
Homburg, A., Weiß, C. H., Alwan, L. C., Frahm, G., & Göb, R. (2020). A performance analysis of prediction intervals for count time series. *Journal of Forecasting*.
Hong, W.-C. (2011). Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. *Neurocomputing*, *74*(12–13), 2096–2107.
Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, *32*(3), 914–938.
Hong, Y., Li, H., & Zhao, F. (2004). Out-of-sample performance of discrete-time spot interest rate models. *Journal of Business & Economic Statistics*, *22*(4), 457–473.
Hong, T., & Pinson, P. (2019). Energy forecasting in the big data world. *International Journal of Forecasting*, *35*(4), 1387–1388.
Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, *30*(2), 357–363.
Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, *32*(3), 896–913.
Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, *7*, 376–388.
Hong, T., Xie, J., & Black, J. (2019). Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. *International Journal of Forecasting*, *35*(4), 1389–1399.
Honnibal, M. (2015). spaCy: INdustrial-strength natural language processing (NLP) with python and cython. Accessed on 2020-09-10, https://spacy.io.
Honoré, C., Menut, L., Bessagnet, B., Meleux, F., & Rou, L. (2007). An integrated air quality forecast system for a metropolitan area. *Development in Environmental Science*, *6*, 292–300.
Hooker, R. H. (1901). The suspension of the Berlin produce exchange and its effect upon corn prices. *Journal of the Royal Statistical Society*, *64*(4), 574–613.

Hopman, D., Koole, G., & van der Mei, R. (2021). A machine learning approach to itinerary-level booking prediction in competitive airline markets. ArXiv:2103.08405.

Hora, S. C. (2004). Probability judgments for continuous quantities: Linear combinations and calibration. *Management Science*, *50*(5), 597–604.

Hörmann, S., Horváth, L., & Reeder, R. (2013). A functional version of the ARCH model. *Economic Theory*, *29*(2), 267–288.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, *4*(2), 251–257.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Horrace, W. C., & Schmidt, P. (2000). Multiple comparisons with the best, with economic applications. *Journal of Applied Econometrics*, *15*(1), 1–26.

Horst, E. T., Rodriguez, A., Gzyl, H., & Molina, G. (2012). Stochastic volatility models including open, close, high and low prices. *Quantitative Finance*, *12*(2), 199–212.

Horváth, L., & Kokoszka, P. (2012). *Inference for Functional Data with Applications*. New York: Springer.

Horváth, L., Kokoszka, P., & Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, *179*(1), 66–82.

Horváth, L., Liu, Z., Rice, G., & Wang, S. (2020). A functional time series analysis of forward curves derived from commodity futures. *International Journal of Forecasting*, *36*(2), 646–665.

Hoskins, B. (2013). The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Quarterly Journal of the Royal Meteorological Society*, *139*(672), 573–584.

Hossin, M., & Sulaiman, M. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 1–11.

Hou, Y., Edara, P., & Sun, C. (2014). Traffic flow forecasting for urban work zones. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 1761–1770.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, *33*(5), 2019–2133.

Hsu, J. C. (1981). Simultaneous confidence intervals for all distances from the "best". *The Annals of Statistics*, 1026–1034.

Hu, K., Acimovic, J., Erize, F., Thomas, D. J., & Van Mieghem, J. A. (2019). Forecasting new product life cycle curves: Practical approach and empirical analysis. *Manufacturing & Service Operations Management*, *21*(1), 66–85.

Huang, C., Chen, S., Yang, S., & Kuo, C. (2015). One-day-ahead hourly forecasting for photovoltaic power generation using an intelligent method with weather-based forecasting models. *IET Generation, Transmission and Distribution*, *9*(14), 1874–1882.

Huang, T., Fildes, R., & Soopramanien, D. (2019). Forecasting retailer product sales in the presence of structural change. *European Journal of Operational Research*, *279*(2), 459–470.

Huang, J., Horowitz, J. L., & Wei, F. (2010). Variable selection in nonparametric additive models. *The Annals of Statistics*, *38*(4), 2282–2313.

Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies*, *28*(3), 791–837.

Huard, D., Évin, G., & Favre, A.-C. (2006). BayesIan copula selection. *Computational Statistics & Data Analysis*, *51*(2), 809–822.

Hubáček, O., Šourek, G., & Železnỳ, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, *35*(2), 783–796.

Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*.

Huberty, M. (2015). Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*, *31*(3), 992–1007.

Hubicka, K., Marcjasz, G., & Weron, R. (2018). A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Transactions on Sustainable Energy*, *10*(1), 321–323.

Hughes, M. C. (2001). Forecasting practice: organisational issues. *Journal of the Operational Research Society*, *52*(2), 143–149.

Huh, S.-Y., & Lee, C.-Y. (2014). Diffusion of renewable energy technologies in South Korea on incorporating their competitive interrelationships. *Energy Policy*, *69*, 248–257.

Hui, F. K. C., Warton, D. I., & Foster, S. D. (2015). Tuning parameter selection for the adaptive lasso using ERIC. *Journal of the American Statistical Society*, *110*(509), 262–269.

Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, *44*(1), 215–238.

Hyndman, R. J. (1996). Computing and graphing Highest Density Regions. *The American Statistician*, *50*(2), 120–126.

Hyndman, R. J. (2020). Quality measure for predictive Highest Density Regions. Cross Validated, Accessed on 2020-08-20, URL https://stats.stackexchange.com/q/483882.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, *55*(9), 2579–2589.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). Melbourne, Australia: OTexts.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). Melbourne, Australia: OTexts, URL https://otexts.com/fpp3/.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., et al. (2020). Forecast: Forecasting functions for time series and linear models. R package version 8.12.

Hyndman, R. J., Bashtannyk, D. M., & Grunwald, G. K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, *5*(4), 315–336.

Hyndman, R. J., & Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, *19*(2), 287–290.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Berlin: Springer Verlag.

Hyndman, R. J., Koehler, A. B., Snyder, R., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439–454.

Hyndman, R. J., & Shang, H. L. (2009). Forecasting functional time series (with discussions). *Journal of the Korean Statistical Society*, *38*(3), 199–221.

Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, *51*(10), 4942–4956.

Hyndman, R. J., Zeng, Y., & Shang, H. L. (2021). Forecasting the old-age dependency ratio to determine a sustainable pension age. *Australian & New Zealand Journal of Statistics*, *in press*.

Hyppöla, J., Tunkelo, A., & Törnqvist, L. (Eds.), (1949). Suomen väestöä, sen uusiutumista ja tulevaa kehitystä koskevia laskelmia. *Tilastollisia tiedonantoja*: *vol. 38*, Helsinki: Statistics Finland.

Ibrahim, R., & L'Ecuyer, P. (2013). Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models. *Manufacturing & Service Operations Management*, *15*(1), 72–85.

Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. (2016). Modeling and forecasting call center arrivals: A literature survey and a case study. *International Journal of Forecasting*, *32*(3), 865–874.

IEA, P. (2020). Electricity information: Overview. URL www.iea.org/reports/electricity-information-overview.

IHME COVID-19 health service utilization forecasting team, & Murray, C. J. L. (2020a). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. http://dx.doi.org/10.1101/2020.03.27.20043752, Medrxiv;2020.03.27.20043752v1.

IHME COVID-19 health service utilization forecasting team, & Murray, C. J. L. (2020b). Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European economic area countries. http://dx.doi.org/10.1101/2020.04.21.20074732, Medrxiv;2020.04.21.20074732v1.

Ince, O. (2014). Forecasting exchange rates out-of-sample with panel methods and real-time data. *Journal of International Money and Finance*, *43*(C), 1–18.

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, *196*(1), 55–67.

Inoue, A., & Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. *Journal of the American Statistical Association*, *103*(482), 511–522.
ifo Institute (2020). *ifo business climate index for Germany*. Ifo Institute, Accessed on 2020-09-07, https://www.ifo.de/en/survey/ifo-business-climate-index.
Ioannidis, J. P. A., Cripps, S., & Tanner, M. A. (2020). Forecasting for COVID-19 has failed. *International Journal of Forecasting*.
Irwin, G. A., & Meeter, D. A. (1969). Building voter transition models from aggregate data. *Midwest Journal of Political Science*, *13*(4), 545–566.
Islam, T., & Meade, N. (2000). Modelling diffusion and replacement. *European Journal of Operational Research*, *125*(3), 551–570.
Ivanov, S., & Zhechev, V. (2012). Hotel revenue management – a critical literature review. *Tourism: An International Interdisciplinary Journal*, *60*(2), 175–197.
Jacobs, J. P. A. M., & van Norden, S. (2011). Modeling data revisions: Measurement error and dynamics of 'true' values. *Journal of Econometrics*, *161*, 101–109.
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. New York, USA: Springer.
Jammazi, R., & Aloui, C. (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics*, *34*(3), 828–841.
Janczura, J., Trück, S., Weron, R., & Wolff, R. C. (2013). Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics*, *38*, 96–110.
Janke, T., & Steinke, F. (2019). Forecasting the price distribution of continuous intraday electricity trading. *Energies*, *12*(22), 4262.
Janssen, F. (2018). Advances in mortality forecasting: introduction. *Genus*, *74*(21).
Januschowski, T., Arpin, D., Salinas, D., Flunkert, V., Gasthaus, J., Stella, L., et al. (2018). Now available in Amazon SageMaker: DeepAR algorithm for more accurate time series forecasting. In *AWS machine learning blog*. Accessed on 2020-09-01, https://aws.amazon.com/blogs/machine-learning/now-available-in-amazon-sagemaker-deepar-algorithm-for-more-accurate-time-series-forecasting/.
Januschowski, T., Gasthaus, J., Wang, Y., Rangapuram, S. S., & Callot, L. (2018). Deep learning for forecasting: Current trends and challenges. *Foresight: The International Journal of Applied Forecasting*, *51*, 42–47.
Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., et al. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, *36*(1), 167–177.
Januschowski, T., & Kolassa, S. (2019). A classification of business forecasting problems. *Foresight: The International Journal of Applied Forecasting*, *52*, 36–43.
Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510.
Jennings, W., Lewis-Beck, M., & Wlezien, C. (2020). Election forecasting: Too far out? *International Journal of Forecasting*, *36*(3), 949–962.
Jeon, J., Panagiotelis, A., & Petropoulos, F. (2019). Probabilistic forecast reconciliation with applications to wind power and electric load. *European Journal of Operational Research*.
Jeon, J., & Taylor, J. (2016). Short-term density forecasting of wave energy using ARMA-GARCH models and kernel density estimation. *International Journal of Forecasting*, *32*(3), 991–1004.
Jiang, J. J., Muhanna, W. A., & Pick, R. A. (1996). The impact of model performance history information on users' confidence in decision models: An experimental examination. *Computers in Human Behavior*, *12*(2), 193–207.
Jiao, E. X., & Chen, J. L. (2019). Tourism forecasting: A review of methodological developments over the last decade. *Tourism Economics*, *25*(3), 469–492.
Jing, G., Cai, W., Chen, H., Zhai, D., Cui, C., & Yin, X. (2018). An air balancing method using support vector machine for a ventilation system. *Building and Environment*, *143*, 487–495.
Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London.
Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, *94*(2), 401–419.
Joe, H. (2014). *Dependence Modeling with Copulas*. CRC Press.
Johansen, S., & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In J. Castle, & N. Shephard (Eds.), *The methodology and practice of econometrics: A festschrift in honour of David F. Hendry* (pp. 1–35). Oxford and New York: Oxford University Press.
Johnes, G. (1999). Forecasting unemployment. *Applied Economics Letters*, *6*(9), 605–607.
Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: Initial studies of its effects on risk perception and trust. *Risk Analysis*, *15*(4), 485–494.
Johnston, D. M. (2008). *The historical foundations of world order: The tower and the arena*. Martinus Nijhoff Publishers.
Johnston, R., & Pattie, C. (2000). Ecological inference and entropy-maximizing: An alternative estimation procedure for split-ticket voting. *Political Analysis*, *8*(4), 333–345.
Johnstone, D. J., Jose, V. R. R., & Winkler, R. L. (2011). Tailored scoring rules for probabilities. *Decision Analysis*, *8*(4), 256–268.
Joiner, T. A., Leveson, L., & Langfield-Smith, K. (2002). Technical language, advice understandability, and perceptions of expertise and trustworthiness: The case of the financial planner. *Australian Journal of Management*, *27*(1), 25–43.
Jondeau, E. (2007). *Financial modelling under non-Gaussian distributions* (1st ed.). London: Springer.
Jongbloed, G., & Koole, G. (2001). Managing uncertainty in call centres using Poisson mixtures. *Applied Stochastic Models in Business and Industry*, *17*(4), 307–318.
Jonung, L., & Larch, M. (2006). Improving fiscal policy in the EU: the case for independent forecasts. *Economic Policy*, *21*(47), 491–534.
Jordá, O., Knüppelc, M., & Marcellino, M. (2013). Empirical simultaneous prediction regions for path-forecasts. *International Journal of Forecasting*, *29*(3), 456–468.
Jore, A. S., Mitchell, J., & Vahey, S. P. (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics*, *25*(4), 621–634.
Jose, V. R. R., Grushka-Cockayne, Y., & Lichtendahl, K. C. (2014). Trimmed opinion pools and the crowd's calibration problem. *Management Science*, *60*(2), 463–475.
Jose, V. R. R., Nau, R. F., & Winkler, R. L. (2008). Scoring rules, generalized entropy, and utility maximization. *Operations Research*, *56*(5), 1146–1157.
Jose, V. R. R., & Winkler, R. L. (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, *24*(1), 163–169.
Jose, V. R. R., & Winkler, R. L. (2009). Evaluating quantile assessments. *Operations Research*, *57*(5), 1287–1297.
Joslyn, S. L., Nadav-Greenberg, L., Taing, M. U., & Nichols, R. M. (2009). The effects of wording on the understanding and use of uncertainty information in a threshold forecasting decision. *Applied Cognitive Psychology*, *23*(1), 55–72.
Joslyn, S. L., & Nichols, R. M. (2009). Probability or frequency? Expressing forecast uncertainty in public weather forecasts. *Meteorological Applications*, *16*(3), 309–314.
Julier, S. J., & Uhlmann, J. K. (1997). New extension of the Kalman filter to nonlinear systems. In I. Kadar (Ed.), *3068*, *Signal processing, sensor fusion, and target recognition VI* (pp. 182–193). SPIE, International Society for Optics and Photonics.
Jung, R. C., & Tremayne, A. (2006). Coherent forecasting in integer time series models. *International Journal of Forecasting*, *22*(2), 223–238.
Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, *10*(3), 215–236.
Kaboudan, M. (2001). Compumetric forecasting of crude oil prices. In *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)* (pp. 283–287). IEEE.
Kagraoka, Y. (2016). Common dynamic factors in driving commodity prices: Implications of a generalized dynamic factor model. *Economic Modelling*, *52*, 609–617.
Kahn, K. B. (2002). An exploratory investigation of new product forecasting practices. *Journal of Product Innovation Management*, *19*(2), 133–143.
Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin books.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*(3), 582–91; discusion 592–6.
Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). *Making text count: economic forecasting using newspaper text*: *Tech. Rep. 865*, Bank of England.
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, *82*(1), 35–45.
Kamarianakis, Y., & Prastacos, P. (2005). Space–time modeling of traffic flow. *Computers & Geosciences*, *31*(2), 119–133.
Kamisan, N. A. B., Lee, M. H., Suhartono, S., Hussin, A. G., & Zubairi, Y. Z. (2018). Load forecasting using combination model of multiple linear regression with neural network for Malaysian city. *Sains Malaysiana*, *47*(2), 419–426.
Kang, Y. (2012). Real-time change detection in time series based on growing feature quantization. In *The 2012 international joint conference on neural networks (IJCNN)*(pp. 1–6).
Kang, Y., Belušić, D., & Smith-Miles, K. (2014). Detecting and classifying events in noisy time series. *Journal of the Atmospheric Sciences*, *71*(3), 1090–1104.
Kang, Y., Belušić, D., & Smith-Miles, K. (2015). Classes of structures in the stable atmospheric boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *141*(691), 2057–2069.
Kang, Y., Hyndman, R. J., & Li, F. (2020). GRATIS: GeneRAting time series with diverse and controllable characteristics. *Statistical Analysis and Data Mining*, *13*(4), 354–376.
Kang, Y., Hyndman, R. J., & Smith-Miles, K. (2017). Visualising forecasting algorithm performance using time series instance spaces. *International Journal of Forecasting*, *33*(2), 345–358.
Kang, S. H., Kang, S.-M., & Yoon, S.-M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, *31*(1), 119–125.
Kang, Y., Spiliotis, E., Petropoulos, F., Athiniotis, N., Li, F., & Assimakopoulos, V. (2020). Déjà vu: A Data-centric forecasting approach through time series cross-similarity. *Journal of Business Research*.
Kapetanios, G., Mitchell, J., Price, S., & Fawcett, N. (2015). Generalised density forecast combinations. *Journal of Econometrics*, *188*(1), 150–165.
Kargin, V., & Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, *99*(10), 2508–2526.
Karniouchina, E. V. (2011). Are virtual markets efficient predictors of new product success? The case of the Hollywood stock exchange. *The Journal of Product Innovation Management*, *28*(4), 470–484.
Kascha, C., & Ravazzolo, F. (2010). Combining inflation density forecasts. *Journal of Forecasting*, *29*(1–2), 231–250.
Katz, R. W., & Lazo, J. K. (Eds.), (2011). *The Oxford Handbook of Economic Forecasting*, *Economic value of weather and climate forecasts*. Oxford University Press.
Kaufmann, R., & Juselius, K. (2013). Testing hypotheses about glacial cycles against the observational record. *Paleoceanography*, *28*, 175–184.
Kayacan, E., Ulutas, B., & Kaynak, O. (2010). Grey system theory-based models in time series prediction. *Expert Systems with Applications*, *37*(2), 1784–1789.
Keane, M. P., & Runkle, D. E. (1990). Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review*, *80*(4), 714–735.
Kedia, S., & Williams, C. (2003). Predictors of substance abuse treatment outcomes in Tennessee. *Journal of Drug Education*, *33*(1), 25–47.
Kehagias, A., & Petridis, V. (1997). Time-series segmentation using predictive modular neural networks. *Neural Computation*, *9*(8), 1691–1709.
Keiding, N., & Hoem, J. M. (1976). Stochastic stable population theory with continuous time. I. *Scandinavian Actuarial Journal*, *1976*(3), 150–175.
Kelle, P., & Silver, E. A. (1989). Forecasting the returns of reusable containers. *Journal of Operations Management*, *8*(1), 17–35.
Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, *68*(5), 1721–1756.
Kennedy, W. J., Wayne Patterson, J., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of Production Economics*, *76*(2), 201–215.
Kennedy, R., Wojcik, S., & Lazer, D. (2017). Improving election prediction internationally. *Science*, *355*(6324), 515–520.
Keyfitz, N. (1972). On future population. *Journal of the American Statistical Association*, *67*(338), 347–363.
Keyfitz, N. (1981). The limits of population forecasting. *Population and Development Review*, *7*(4), 579–593.
Khaldi, R., El Afia, A., & Chiheb, R. (2019). Forecasting of weekly patient visits to emergency department: real case study. *Procedia Computer Science*, *148*, 532–541.
Kiesel, R., & Paraschiv, F. (2017). Econometric analysis of 15-minute intraday electricity prices. *Energy Economics*, *64*, 77–90.
Kilian, L., & Inoue, A. (2004). *Bagging time series models*: *Tech. Rep. 110*, Econometric Society.
Kilian, L., & Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, *60*(1), 85–107.
Kim, C.-J., Charles, K. C.-J. N., & Nelson, C. R. (1999). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*. MIT Press.
Kim, T. Y., Dekker, R., & Heij, C. (2017). Spare part demand forecasting for consumer goods using installed base information. *Computers & Industrial Engineering*, *103*, 201–215.
Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, *81*, 361–393.
Kim, H. H., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, *178*, 352–367.
King, G. (1997). *A solution to the ecological inference problem: reconstructing individual behavior from aggregate data*. Princeton University Press.
King, G., Rosen, O., & Tanner, M. A. (1999). Binomial-beta hierarchical models for ecological inference. *Sociological Methods & Research*, *28*(1), 61–90.
King, G., Tanner, M. A., & Rosen, O. (2004). *Ecological inference: New methodological strategies*. Cambridge University Press.
Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. San Diego, Third Annual International Conference on Learning Representations.
Kishor, N. K., & Koenig, E. F. (2012). VAR estimation and forecasting when data are subject to revision. *Journal of Business & Economic Statistics*, *30*(2), 181–190.
Kittichotsatsawat, Y., Jangkrajarng, V., & Tippayawong, K. Y. (2021). Enhancing coffee supply chain towards sustainable growth with big data and modern agricultural technologies. *Sustainability*, *13*(8), 4593.
Klepsch, J., & Klüppelberg, C. (2017). An innovations algorithm for the prediction of functional linear processes. *Journal of Multivariate Analysis*, *155*, 252–271.
Klepsch, J., Klüppelberg, C., & Wei, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econometrics and Statistics*, *1*, 128–149.
Klima, A., Schlesinger, T., Thurner, P. W., & Küchenhoff, H. (2019). Combining aggregate data and exit polls for the estimation of voter transitions. *Sociological Methods & Research*, *48*(2), 296–325.
Klima, A., Thurner, P. W., Molnar, C., Schlesinger, T., & Küchenhoff, H. (2016). Estimation of voter transitions based on ecological inference. *AStA. Advances in Statistical Analysis*, *2*, 133–159.
Kline, D. M. (2004). Methods for multi-step time series forecasting with neural networks. In G. P. Zhang (Ed.), *Neural networks in business forecasting* (pp. 226–250). Information Science Publishing.
Klofstad, C. A., & Bishin, B. G. (2012). Exit and entrance polling: A comparison of election survey methods. *Field Methods*, *24*(4), 429–437.
Knudsen, C., McNown, R., & Rogers, A. (1993). Forecasting fertility: An application of time series methods to parameterized model schedules. *Social Science Research*, *22*(1), 1–23.
Koenig, E. F., Dolmas, S., & Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *The Review of Economics and Statistics*, *85*(3), 618–628.
Koenker, R. (2005). *Econometric Society Monographs*, *Quantile regression*. Cambridge University Press.
Koh, Y.-M., Spindler, R., Sandgren, M., & Jiang, J. (2018). A model comparison algorithm for increased forecast accuracy of dengue fever incidence in Singapore and the auxiliary role of total precipitation information. *International Journal of Environmental Health Research*, *28*(5), 535–552.

Koirala, K. H., Mishra, A. K., D'Antoni, J. M., & Mehlhorn, J. E. (2015). Energy prices and agricultural commodity prices: Testing correlation using copulas method. *Energy*, *81*, 430–436.
Kokoszka, P., & Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, *34*(1), 116–129.
Kokoszka, P., Rice, G., & Shang, H. L. (2017). Inference for the autocovariance of a functional time series under conditional heteroscedasticity. *Journal of Multivariate Analysis*, *162*, 32–50.
Kolasa, M., & Rubaszek, M. (2015a). Forecasting using DSGE models with financial frictions. *International Journal of Forecasting*, *31*(1), 1–19.
Kolasa, M., & Rubaszek, M. (2015b). How frequently should we reestimate DSGE models? *International Journal of Central Banking*, *11*(4), 279–305.
Kolasa, M., Rubaszek, M., & Skrzypczyński, P. (2012). Putting the New Keynesian DSGE model to the real-time forecasting test. *Journal of Money, Credit and Banking*, *44*(7), 1301–1324.
Kolassa, S. (2011). Combining exponential smoothing forecasts using akaike weights. *International Journal of Forecasting*, *27*(2), 238–251.
Kolassa, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting*, *32*(3), 788–803.
Kolassa, S. (2020a). Quality measure for predictive Highest Density Regions. Cross Validated, Accessed on 2020-08-20, URL https://stats.stackexchange.com/q/483878.
Kolassa, S. (2020b). Why the "best" point forecast depends on the error or accuracy measure. *International Journal of Forecasting*, *36*(1), 208–211.
Kolassa, S. (2020c). Will deep and machine learning solve our forecasting problems? *Foresight: The International Journal of Applied Forecasting*, *57*, 13–18.
Kolassa, S., & Siemsen, E. (2016). *Demand forecasting for managers*. Business Expert Press.
Kon Kam King, G., Canale, A., & Ruggiero, M. (2019). BayesIan functional forecasting with locally-autoregressive dependent processes. *Bayesian Analysis*, *14*(4), 1121–1141.
Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. O. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting*, *21*(3), 397–409.
Koop, G. M. (2003). *Bayesian econometrics*. John Wiley & Sons Inc..
Koop, G., & Korobilis, D. (2018). Variational Bayes inference in high-dimensional time-varying parameter models. *Journal of Econometrics*.
Koop, G., & Potter, S. M. (1999). Dynamic asymmetries in U.S. unemployment. *Journal of Business & Economic Statistics*, *17*(3), 298–312.
Kostenko, A. V., & Hyndman, R. J. (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society*, *57*(10), 1256–1257.
Kotchoni, R., Leroux, M., & Stevanovic, D. (2019). Macroeconomic forecast accuracy in a data-rich environment. *Journal of Applied Econometrics*, *34*(7), 1050–1072.
Kourentzes, N., & Athanasopoulos, G. (2019). Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*, *75*, 393–409.
Kourentzes, N., & Athanasopoulos, G. (2020). Elucidate structure in intermittent demand series. *European Journal of Operational Research*.
Kourentzes, N., Barrow, D., & Petropoulos, F. (2019). Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, *209*, 226–235.
Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*, *181, Part A*, 145–153.
Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, *30*(2), 291–302.
Kourentzes, N., Rostami-Tabar, B., & Barrow, D. K. (2017). Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels? *Journal of Business Research*, *78*, 1–9.
Kovalchik, S., & Reid, M. (2019). A calibration method with dynamic updates for within-match forecasting of wins in tennis. *International Journal of Forecasting*, *35*(2), 756–766.
Krishnan, T., Bass, F., & Kummar, V. (2000). Impact of a late entrant on the diffusion of a new product/service. *Journal of Marketing Research*, *37*, 269–278.
Krüger, E., & Givoni, B. (2004). Predicting thermal performance in occupied dwellings. *Energy and Buildings*, *36*(3), 301–307.
Krzysztofowicz, R. (1999). BayesIan theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, *35*(9), 2739–2750.
Krzysztofowicz, R. (2014). Probabilistic flood forecast: Exact and approximate predictive distributions. *Journal of Hydrology*, *517*, 643–651.
Kück, M., Crone, S. F., & Freitag, M. (2016). Meta-learning with neural networks and landmarking for forecasting model selection an empirical evaluation of different feature sets applied to industry data. In *2016 international joint conference on neural networks (IJCNN)* (pp. 1499–1506). IEEE.
Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection*. Taylor & Francis Ltd.
Kulakov, S. (2020). X-model: further development and possible modifications. *Forecasting*, *2*(1), 20–35.
Kulakov, S., & Ziel, F. (2021). The impact of renewable energy forecasts on intraday electricity prices. *Economics of Energy and Environmental Policy*, *10*, 79–104.
Kulkarni, G., Kannan, P. K., & Moe, W. (2012). Using online search data to forecast new product sales. *Decision Support Systems*, *52*(3), 604–611.
Kumar, D. (2015). Sudden changes in extreme value volatility estimator: Modeling and forecasting with economic significance analysis. *Economic Modelling*, *49*, 354–371.
Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, *17*(3), 1217–1241.
Kupiszewski, M., & Kupiszewska, D. (2011). MULTIPOLES: A Revised multiregional model for improved capture of international migration. In J. Stillwell, & M. Clarke (Eds.), *Population dynamics and projection methods* (pp. 41–60). Dordrecht: Springer Netherlands.
Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, *35*, 257–270.
Kusters, U., McCullough, B., & Bell, M. (2006). Forecasting software: Past, present and future. *International Journal of Forecasting*, *22*(3), 599–615.
Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1), 159–178.
Kyriazi, F., Thomakos, D. D., & Guerard, J. B. (2019). Adaptive learning forecasting, with applications in forecasting agricultural prices. *International Journal of Forecasting*, *35*(4), 1356–1369.
La Scalia, G., Micale, R., Miglietta, P. P., & Toma, P. (2019). Reducing waste and ecological impacts through a sustainable and efficient management of perishable food based on the Monte Carlo simulation. *Ecological Indicators*, *97*, 363–371.
Labarere, J., Bertrand, R., & Fine, M. J. (2014). How to derive and validate clinical prediction models for use in intensive care medicine. *Intensive Care Medicine*, *40*(4), 513–527.
Ladiray, D., & Quenneville, B. (2001). *Lecture notes in statistics 158, Seasonal adjustment with the X-11 method*. New York, USA: Springer.
Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, *221*, 386–405.
Lahiri, S. K., & Lahiri, N. (2003). *Resampling methods for dependent data (Springer series in statistics)*. Springer.
Lai, G., Chang, W.-C., Yang, Y., & Liu, H. (2018. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 95–104).
Landon, J., Ruggeri, F., Soyer, R., & Tarimcilar, M. M. (2010). Modeling latent sources in call center arrival data. *European Journal of Operational Research*, *204*(3), 597–603.
Lanne, M., & Saikkonen, P. (2003). Modeling the U.S. short-term interest rate by mixture autoregressive processes. *Journal of Financial Econometrics*, *1*(1), 96–125.
Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*(1), 111–127.

Larson, P. D., Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2001). Designing and managing the supply chain: Concepts, strategies, and case studies. *Journal of Business Logistics*, *22*(1), 259–261.
Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, *75*, 410–423.
Lawrence, M. (2000). What does it take to achieve adoption in sales forecasting? *International Journal of Forecasting*, *16*(2), 147–148.
Lawrence, M., Goodwin, P., & Fildes, R. (2002). Influence of user participation on DSS use and decision accuracy. *Omega*, *30*(5), 381–392.
Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, *22*(3), 493–518.
Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, *43*(2), 172–187.
Lawrence, M., & O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, *8*(1), 15–26.
Layard, R., Nickell, S. J., & Jackman, R. (1991). *Unemployment, macroeconomic performance and the labour market*. Oxford: Oxford University Press.
Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188–1196).
Leadbetter, M. R. (1991). On a basis for 'Peaks over Threshold' modeling. *Statistics & Probability Letters*, *12*(4), 357–362.
Leal, T., Pérez, J. J., Tujula, M., & Vidal, J. P. (2008). Fiscal forecasting: Lessons from the literature and challenges. *Fiscal Studies*, *29*, 347–386.
Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *International Journal of Forecasting*, *5*(2), 231–240.
Ledolter, J. (1991). Outliers in time series analysis: Some comments on their impact and their detection. *Image*.
Lee, A. (1990). *Airline reservations forecasting: Probabilistic and statistical models of the booking process*: *Tech. rep.*, Cambridge, Mass.: Flight Transportation Laboratory, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology.
Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting*, *9*(2), 187–202.
Lee, K. L., & Billings, S. a. (2003). A new direct approach of computing multi-step ahead predictions for non-linear models. *International Journal of Control*, *76*(8), 810–822.
Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, *87*(419), 659–671.
Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, *23*(3), 377–390.
Lee, C.-Y., & Huh, S.-Y. (2017a). Forecasting new and renewable energy supply through a bottom-up approach: The case of South Korea. *Renewable and Sustainable Energy Reviews*, *69*, 207–217.
Lee, C.-Y., & Huh, S.-Y. (2017b). Forecasting the diffusion of renewable electricity considering the impact of policy and oil prices: The case of South Korea. *Applied Energy*, *197*, 29–39.
Lee, J., Milesi-Ferretti, G. M., & Ricci, L. A. (2013). Real exchange rates and fundamentals: A cross-country perspective. *Journal of Money, Credit and Banking*, *45*(5), 845–865.
Lee, H. L., Padmanabhan, V., & Whang, S. (2004). Information distortion in a supply chain: The Bullwhip effect. *Management Science*, *50*, 1875–1886.
Leigh, C., Alsibai, O., Hyndman, R. J., Kandanaarachchi, S., King, O. C., McGree, J. M., et al. (2019). A framework for automated anomaly detection in high frequency water-quality data from in situ sensors. *Science of the Total Environment*, *664*, 885–898.
Lemke, C., & Gabrys, B. (2010). Meta-learning for time series forecasting and forecast combination. *Neurocomputing*, *73*(10–12), 2006–2016.
Lerch, S., Baran, S., Möller, A., Groß, J., Schefzik, R., Hemri, S., et al. (2020). Simulation-based comparison of multivariate ensemble post-processing methods. *Nonlinear Processes in Geophysics*, *27*(2), 349–371.
Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, *33*(3), 183–212.
Leslie, P. H. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika*, *35*(3/4), 213–245.
Leuenberger, D., Haefele, A., Omanovic, N., Fengler, M., Martucci, G., Calpini, B., et al. (2020). Improving high-impact numerical weather prediction with lidar and drone observations. *Bulletin of the American Meteorological Society*, *101*(7), E1036–E1051.
Leva, S., Mussetta, M., & Ogliari, E. (2019). PV module fault diagnosis based on microconverters and day-ahead forecast. *IEEE Transactions on Industrial Electronics*, *66*(5), 3928–3937.
Levine, R., Pickett, J., Sekhri, N., & Yadav, P. (2008). Demand forecasting for essential medical technologies. *American Journal of Law & Medicine*, *34*(2–3), 225–255.
Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review*, *4*(1), 1–44.
Lewis, B., Herbert, R., & Bell, R. (2003). The application of fourier analysis to forecasting the inbound call time series of a call centre. In *Proceedings of the international congress on modeling and simulation (MODSIM03); Townsville, Australia* (pp. 1281–1286). Citeseer.
Lewis-Beck, M. S. (2005). Election forecasting: Principles and practice. *British Journal of Politics and International Relations*, *7*(2), 145–164.
L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access*, *5*, 7776–7797.
Li, D. X. (2000). On default correlation: A copula function approach. *The Journal of Fixed Income*, *9*(4), 43–54.
Li, J. S.-H., & Chan, W.-S. (2011). Time-simultaneous prediction bands: A new look at the uncertainty involved in forecasting mortality. *Insurance: Mathematics & Economics*, *49*(1), 81–88.
Li, W., Han, Z., & Li, F. (2008). Clustering analysis of power load forecasting based on improved ant colony algorithm. In *2008 7th world congress on intelligent control and automation* (pp. 7492–7495).
Li, F., & He, Z. (2019). Credit risk clustering in a business group: which matters more, systematic or idiosyncratic risk? *Cogent Economics & Finance*, Article 1632528.
Li, H., & Hong, Y. (2011). Financial volatility forecasting with range-based autoregressive volatility model. *Finance Research Letters*, *8*(2), 69–76.
Li, M., Huang, L., & Gong, L. (2011). Research on the challenges and solutions of design large-scale call center intelligent scheduling system. *Procedia Engineering*, *15*, 2359–2363.
Li, G., & Jiao, E. (2020). Tourism forecasting research: a perspective article. *Tourism Review*.
Li, F., & Kang, Y. (2018). Improving forecasting performance using covariate-dependent copula models. *International Journal of Forecasting*, *34*(3), 456–476.
Li, X., Kang, Y., & Li, F. (2020). Forecasting with time series imaging. *Expert Systems with Applications*, *160*, Article 113680.
Li, J., Li, G., Liu, M., Zhu, X., & Wei, L. (2020). A novel text-based framework for forecasting agricultural futures using massive online news headlines. *International Journal of Forecasting*.
Li, J., Liao, Z., & Quaedvlieg, R. (2020). Conditional superior predictive ability. SSRN:3536461.
Li, L., Noorian, F., Moss, D. J., & Leong, P. H. (2014). Rolling window time series prediction using MapReduce. In *Proceedings of the 2014 IEEE 15th international conference on information reuse and integration (IEEE IRI 2014)* (pp. 757–764). IEEE.
Li, D., Robinson, P. M., & Shang, H. L. (2020d). *Nonstationary fractionally integrated functional time series*: *Working paper*, University of York.
Li, D., Robinson, P. M., & Shang, H. L. (2020e). Long-range dependent curve time series. *Journal of the American Statistical Association*, *115*(530), 957–971.
Li, D., Robinson, P. M., & Shang, H. L. (2021). Local whittle estimation of long range dependence for functional time series. *Journal of Time Series Analysis*, *In Press*.
Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, *44*(1), 82–99.
Liang, Y., He, D., & Chen, D. (2019). Poisoning attack on load forecasting. In *2019 IEEE innovative smart grid technologies-Asia (ISGT Asia)* (pp. 1230–1235). IEEE.
Liberty, E., Karnin, Z., Xiang, B., Rouesnel, L., Coskun, B., Nallapati, R., et al. (2020). Elastic machine learning algorithms in amazon SageMaker. In *SIGMOD '20*, *Proceedings of the 2020 International Conference on Management of Data* (pp. 731–737). New York, NY, USA: ACM.

Lichtendahl, K. C., Grushka-Cockayne, Y., & Winkler, R. L. (2013). Is it better to average probabilities or quantiles? *Management Science*, *59*(7), 1594–1611.
Lichtendahl Jr, K. C., & Winkler, R. L. (2020). Why do some combinations perform better than others? *International Journal of Forecasting*, *36*(1), 142–149.
Lildholdt, P. M. (2002). *Centre for Analytical Finance*, *Estimation of GARCH models based on open, close, high, and low prices*. Aarhus School of Business, Centre for Analytical Finance, Aarhus School of Business.
Lim, J. S., & O'Connor, M. (1996a). Judgmental forecasting with interactive forecasting support systems. *Decision Support Systems*, *16*(4), 339–357.
Lim, J. S., & O'Connor, M. (1996b). Judgmental forecasting with interactive forecasting support systems. *Decision Support Systems*, *16*(4), 339–357.
Lim, J. S., & O'Connor, M. (1996c). Judgmental forecasting with time series and causal information. *International Journal of Forecasting*, *12*(1), 139–153.
Limaye, V. S., Vargo, J., Harkey, M., Holloway, T., & Patz, J. A. (2018). Climate change and heat-related excess mortality in the eastern USA. *EcoHealth*, *15*(3), 485–496.
Lin, E. M., Chen, C. W., & Gerlach, R. (2012). Forecasting volatility with asymmetric smooth transition dynamic range models. *International Journal of Forecasting*, *28*(2), 384–399.
Lin, J. L., & Granger, C. (1994). Forecasting from non-linear models in practice. *Journal of Forecasting*, *13*, 1–9.
Lin, C.-F. J., & Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, *62*(2), 211–228.
Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied Probability*, *36*(3), 688–705.
Ling, S., Tong, H., & Li, D. (2007). Ergodicity and invertibility of threshold moving-average models. *Bernoulli*, *13*(1), 161–168.
Linnér, L., Eriksson, I., Persson, M., & Wettermark, B. (2020). Forecasting drug utilization and expenditure: ten years of experience in stockholm. *BMC Health Services Research*, *20*, 1–11.
Litsiou, K., Polychronakis, Y., Karami, A., & Nikolopoulos, K. (2019). Relative performance of judgmental methods for forecasting the success of megaprojects. *International Journal of Forecasting*.
Liu, Y. (2005). Value-at-Risk Model Combination Using Artificial Neural Networks. Ermory University Working Papers.
Liu, L., & Wu, L. (2021). Forecasting the renewable energy consumption of the European countries by an adjacent non-homogeneous grey model. *Applied Mathematical Modelling*, *89*, 1932–1948.
Liu, W., Zhu, F., Zhao, T., Wang, H., Lei, X., Zhong, P.-A., et al. (2020). Optimal stochastic scheduling of hydropower-based compensation for combined wind and photovoltaic power outputs. *Applied Energy*, *276*, Article 115501.
Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.
Loaiza-Maya, R., Martin, G. M., & Frazier, D. T. (2020). Focused Bayesian prediction. *Journal of Applied Econometrics*.
Loaiza-Maya, R., & Smith, M. S. (2020). Real-time macroeconomic forecasting with a heteroscedastic inversion copula. *Journal of Business & Economic Statistics*, *38*(2), 470–486.
Loaiza-Maya, R., Smith, M. S., Nott, D. J., & Danaher, P. J. (2020). Fast and accurate variational inference for models with many latent variables. ArXiv:2005.07430.
Locarek-Junge, H., & Prinzler, R. (1998). Estimating value-at-risk using neural networks. In *Informationssysteme in der finanzwirtschaft* (pp. 385–397). Springer Berlin Heidelberg.
Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103.
Lohmann, T., Hering, A. S., & Rebennack, S. (2016). Spatio-temporal hydro forecasting of multireservoir inflows for hydro-thermal scheduling. *European Journal of Operational Research*, *255*(1), 243–258.
Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. ArXiv: Cs/0205028.
López, M., Valero, S., Senabre, C., Aparicio, J., & Gabaldon, A. (2012). Application of SOM neural networks to short-term load forecasting: The spanish electricity market case study. *Electric Power Systems Research*, *91*, 18–27.
López, C., Zhong, W., & Zheng, M. (2017). Short-term electric load forecasting based on wavelet neural network, particle swarm optimization and ensemble empirical mode decomposition. *Energy Procedia*, *105*, 3677–3682.
López Cabrera, B., & Schulz, F. (2016). Volatility linkages between energy and agricultural commodity prices. *Energy Economics*, *54*(C), 190–203.
López-Ruiz, A., Bergillos, R. J., & Ortega-Sánchez, M. (2016). The importance of wave climate forecasting on the decision-making process for nearshore wave energy exploitation. *Applied Energy*, *182*, 191–203.
Lopez-Suarez, C. F., & Rodriguez-Lopez, J. A. (2011). Nonlinear exchange rate predictability. *Journal of International Money and Finance*, *30*(5), 877–895.
Lothian, J. R., & Taylor, M. P. (1996). Real exchange rate behavior: The recent float from the perspective of the past two centuries. *Journal of Political Economy*, *104*(3), 488–509.
Lotka, A. J. (1907). Relation between birth rates and death rates. *Science*, *26*(653), 21–22.
Lotka, A. (1920). Undamped oscillations derived from the law of mass action. *Journal of the American Chemical Society*, *42*, 1595–1599.
Lotka, A. J. (1925). *Elements of physical biology*. Williams & Wilkins.
Lovins, J. (1968). Development of a stemming algorithm. *Mechanical Translation and Computational Linguistics*, *11*(1–2), 22–31.
Lowe, R., Bailey, T. C., Stephenson, D. B., Graham, R. J., Coelho, C. A., Carvalho, M. S., et al. (2011). Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Computers & Geosciences*, *37*(3), 371–381.
Lu, S.-L. (2019). Integrating heuristic time series with modified grey forecasting for renewable energy in Taiwan. *Renewable Energy*, *133*, 1436–1444.
Lu, Y. (2021). The predictive distributions of thinning-based count processes. *Scandinavian Journal of Statistics*, *48*(1), 42–67.
Lu, H., Azimi, M., & Iseley, T. (2019). Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine. *Energy Reports*, *5*, 666–677.
Lübbers, J., & Posch, P. N. (2016). Commodities' common factor: An empirical assessment of the markets' drivers. *Journal of Commodity Markets*, *4*(1), 28–40.
Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, *1*, 19–46.
Lucas, A., Schwaab, B., & Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics*, *32*(2), 271–284.
Ludvigson, S. C., & Ng, S. (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics*, *83*(1), 171–222.
Luo, J., Hong, T., & Fang, S.-C. (2018a). Benchmarking robustness of load forecasting models under data integrity attacks. *International Journal of Forecasting*, *34*(1), 89–104.
Luo, J., Hong, T., & Fang, S.-C. (2018b). Robust regression models for load forecasting. *IEEE Transactions on Smart Grid*, *10*(5), 5397–5404.
Luo, J., Hong, T., & Yue, M. (2018). Real-time anomaly detection for very short-term load forecasting. *Journal of Modern Power Systems and Clean Energy*, *6*(2), 235–243.
Luo, J., Klein, T., Ji, Q., & Hou, C. (2019). Forecasting realized volatility of agricultural commodity futures with infinite hidden Markov HAR models. *International Journal of Forecasting*.
Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer, Berlin, Heidelberg.
Lütkepohl, H. (2011). Forecasting nonlinear aggregates and aggregates with time-varying weights. *JahrbÜCher FÜR NationalÖKonomie Und Statistik*, *231*(1), 107–133.
Lutz, W., Butz, W. P., & Samir, K. C. (2017). *World population and human capital in the twenty-first century: An overview*. Oxford University Press.
Lux, T. (2008). The Markov-switching multifractal model of asset returns. *Journal of Business & Economic Statistics*, *26*(2), 194–210.
Lynn, G. S., Schnaars, S. P., & Skov, R. B. (1999). A survey of new product forecasting practices in industrial high technology and low technology businesses. *Industrial Marketing Management*, *28*(6), 565–571.

Ma, S. (2021). A hybrid deep meta-ensemble networks with application in electric utility industry load forecasting. *Information Sciences*, *544*, 183–196.
Ma, F., Chitta, R., Zhou, J., You, Q., Sun, T., & Gao, J. (2017).
Ma, S., & Fildes, R. (2017). A retail store SKU promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research*, *260*(2), 680–692.
Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, *249*(1), 245–257.
Macaulay, F. R. (1931). The smoothing of time series. In *NBER Books*. National Bureau of Economic Research, Inc.
MacDonald, R. (1998). What determines real exchange rates? The long and the short of it. *Journal of International Financial Markets, Institutions and Money*, *8*(2), 117–153.
MacDonald, R., & Marsh, I. W. (1994). Combining exchange rate forecasts: What is the optimal consensus measure? *Journal of Forecasting*, *13*(3), 313–332.
Madaus, L., McDermott, P., Hacker, J., & Pullen, J. (2020). Hyper-local, efficient extreme heat projection and analysis using machine learning to augment a hybrid dynamical-statistical downscaling technique. *Urban Climate*, *32*, Article 100606.
Maddix, D. C., Wang, Y., & Smola, A. (2018). Deep factors with Gaussian processes for forecasting. ArXiv:1812.00098.
Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301.
Magdon-Ismail, M., & Atiya, A. F. (2003). A maximum likelihood approach to volatility estimation for a Brownian motion using high, low and close price data. *Quantitative Finance*, *3*(5), 376–384.
Mahajan, V., Muller, E., & Bass, F. (1990). New product diffusion models in marketing: a review and directions of future research. *Journal of Marketing*, *54*, 1–26.
Maheu, J. M., & Yang, Q. (2016). An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance*, *38*, 202–220.
Maister, D. H., Galford, R., & Green, C. (2012). *The trusted advisor*. Simon and Schuster.
Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111–153.
Makridakis, S., Bonneli, E., Clarke, S., Fildes, R., Gilliland, M., Hover, J., et al. (2020). The benefits of systematic forecasting for organizations: The UFO project. *Foresight: The International Journal of Applied Forecasting*, *59*, 45–56.
Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., et al. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*(1), 5–22.
Makridakis, S., Fry, C., Petropoulos, F., & Spiliotis, E. (2021). The future of forecasting competitions: Design attributes and principles. *INFORMS Journal on Data Science*.
Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society: Series A (General)*, *142*(2), 97–125.
Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476.
Makridakis, S. G., Hogarth, R. M., & Gaba, A. (2010). *Dance with chance: Making luck work for you*. Oneworld Publications.
Makridakis, S., Hyndman, R. J., & Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, *36*(1), 15–28.
Makridakis, S., Kirkham, R., Wakefield, A., Papadaki, M., Kirkham, J., & Long, L. (2019). Forecasting, uncertainty and risk; perspectives on clinical decision-making in preventive and curative medicine. *International Journal of Forecasting*, *35*(2), 659–666.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS One*, *13*(3), 1–26.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2021). The M5 accuracy competition: Results, findings and conclusions. *International Journal of Forecasting*.
Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Winkler, R. L., et al. (2021). The M5 uncertainty competition: Results, findings and conclusions. *International Journal of Forecasting*.
Makridakis, S., & Winkler, R. L. (1989). Sampling distributions of post-sample forecasting errors. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *38*(2), 331–342.
Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, *7*, 1–15.
Mandal, P., Madhira, S. T. S., Haque, A. U., Meng, J., & Pineda, R. L. (2012). Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. *Procedia Computer Science*, *12*, 332–337.
Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*(4), 394.
Mandelbrot, B. B. (1983). *The fractal geometry of nature*. Henry Holt and Company.
Manders, A., Schaap, M., & Hoogerbrugge, R. (2009). Testing the capability of the chemistry transport model LOTOS-EUROS to forecast PM10 levels in the Netherlands. *Atmospheric Enviroment*, *46*, 4050–4059.
Mangina, E., & Vlachos, I. P. (2005). The changing role of information technology in food and beverage logistics management: beverage network optimisation using intelligent agent technology. *Journal of Food Engineering*, *70*(3), 403–420.
Mankiw, N. G., & Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve. *Quarterly Journal of Economics*, *117*, 1295–1328.
Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectations. *Tech. rep.*, Cambridge MA: National Bureau of Economic Research.
Mann, M. (2018). Have wars and violence declined? *Theory and Society*, *47*(1), 37–60.
Manna, S., Biswas, S., Kundu, R., Rakshit, S., Gupta, P., & Barman, S. (2017). A statistical approach to predict flight delay using gradient boosted decision tree. In *2017 International conference on computational intelligence in data science (ICCIDS)* (pp. 1–5). IEEE.
Manner, H., Türk, D., & Eichler, M. (2016). Modeling and forecasting multivariate electricity price spikes. *Energy Economics*, *60*, 255–265.
Mannes, A. E., Larrick, R. P., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. *Social Judgment and Decision Making*, *297*, 227–242.
Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology*, *107*(2), 276–299.
Manning, C., Schütze, H., & Raghavan, P. (2008). *Introduction to information retrieval*. Cambridge University Press.
Manski, C. F., & Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics*, *28*(2), 219–231.
Mapa, D. (2003). A range-based GARCH model for forecasting volatility. *The Philippine Review of Economics*, *60*(2), 73–90.
Marangon Lima, L. M., Popova, E., & Damien, P. (2014). Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models. *International Journal of Forecasting*, *30*(3), 464–476.
Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1–2), 499–526.
Marchetti, C. (1983). The automobile in a system context: The past 80 years and the next 20 years. *Technological Forecasting and Social Change*, *23*(1), 3–23.
Marchetti, C., & Nakicenovic, N. (1979). *The dynamics of energy systems and the logistic substitution model RR-79-13*, 1–71.
Marcjasz, G., Uniejewski, B., & Weron, R. (2019). On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. *International Journal of Forecasting*, *35*(4), 1520–1532.
Marcjasz, G., Uniejewski, B., & Weron, R. (2020). Beating the Naïve—Combining LASSO with Naïve Intraday Electricity Price Forecasts. *Energies*, *13*(7), 1667.

Marczak, M., & Proietti, T. (2016). Outlier detection in structural time series models: The indicator saturation approach. *International Journal of Forecasting*, *32*(1), 180–202.

Marinakis, V., Doukas, H., Spiliotis, E., & Papastamatiou, I. (2017). Decision support for intelligent energy management in buildings using the thermal comfort model. *International Journal of Computational Intelligence Systems*, *10*, 882–893.

Marinakis, V., Doukas, H., Tsapelas, J., Mouzakitis, S., Sicilia, A., Madrazo, L., et al. (2020). From big data to smart energy services: An application for intelligent energy management. *Future Generation Computer Systems*, *110*, 572–586.

Marinakis, Y., & Walsh, S. (2021). Parameter instability and structural change in s-curve-based technology diffusion forecasting. Working Paper.

Mark, N. C. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review*, *85*(1), 201–218.

Mark, N. C., & Sul, D. (2001). Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton Woods panel. *Journal of International Economics*, *53*(1), 29–52.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

Marron, J. S., & Wand, M. P. (1992). Exact mean integrated squared error. *The Annals of Statistics*, *20*(2), 712–736.

Martin, G. M., Frazier, D. T., & Robert, C. P. (2020). Computing Bayes: Bayesian computation from 1763 to the 21st century. ArXiv:2004.06425.

Martinez, A. B., Castle, J. L., & Hendry, D. F. (2021). Smooth robust multi-horizon forecasts. *Advances in Econometrics*, *Forthcoming*.

Martinez, R., & Sanchez, M. (1970). Automatic booking level control. *10*, In *AGIFORS symposium proc.*

Martinez, E. Z., & Silva, E. A. S. d. (2011). Predicting the number of cases of dengue infection in ribeirão preto, são paulo state, Brazil, using a SARIMA model. *Cadernos de Saude Publica*, *27*, 1809–1818.

Martínez-Álvarez, F., Troncoso, A., Riquelme, J. C., & Aguilar-Ruiz, J. S. (2011). Discovery of motifs to forecast outlier occurrence in time series. *Pattern Recognition Letters*, *32*(12), 1652–1665.

Martinez Alvarez, F., Troncoso, A., Riquelme, J. C., & Aguilar Ruiz, J. S. (2011). Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, *23*(8), 1230–1243.

Masarotto, G. (1990). Bootstrap prediction intervals for autoregressions. *International Journal of Forecasting*, *6*(2), 229–239.

Mat Daut, M. A., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., & Hussin, F. (2017). Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. *Renewable and Sustainable Energy Reviews*, *70*, 1108–1118.

Matte, T. D., Lane, K., & Ito, K. (2016). Excess mortality attributable to extreme heat in new york city, 1997–2013. *Health Security*, *14*(2), 64–70.

Maymin, P. (2019). Wage against the machine: A generalized deep-learning market test of dataset value. *International Journal of Forecasting*, *35*(2), 776–782.

McAlinn, K., Aastveit, K. A., Nakajima, J., & West, M. (2020). Multivariate Bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association*, *115*(531), 1092–1110.

McAlinn, K., & West, M. (2019). Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of Econometrics*, *210*(1), 155–169.

McCabe, B. P., & Martin, G. (2005). BayesIan predictions of low count time series. *International Journal of Forecasting*, *21*(2), 315–330.

McCabe, B. P. M., Martin, G. M., & Harris, D. (2011). Efficient probabilistic forecasts for counts. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *73*(2), 253–272.

McCarthy, C., & Ryan, T. M. (1977). Estimates of voter transition probabilities from the british general elections of 1974. *Journal of the Royal Statistical Society, Series A*, *140*(1), 78–85.

McCoy, T. H., Pellegrini, A. M., & Perlis, R. H. (2018). Assessment of time-series machine learning methods for forecasting hospital discharge volume. *JAMA Network Open*, *1*(7), Article e184087.

McFadden, D. (1977). *Modelling the Choice of residential location*: *Tech. Rep. 477*, Cowles Foundation for Research in Economics, Yale University.

McGill, J., & Van Ryzin, G. (1999). Revenue management: Research overview and prospects. *Transportation Science*, *33*(2), 233–256.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy return predictability? *The Journal of Finance*, *71*(1), 5–32.

McNames, J. (1998). A nearest trajectory strategy for time series prediction. In *Proceedings of the international workshop on advanced black-box techniques for nonlinear modeling* (pp. 112–128). Citeseer.

McNees, S. K. (1990). The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting*, *6*(3), 287–299.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools - revised edition*. Princeton University Press.

Meade, N. (1984). The use of growth curves in forecasting market development - a review and appraisal. *Journal of Forecasting*, *3*(4), 429–451.

Meade, N. (2000). Evidence for the selection of forecasting methods. *Journal of Forecasting*, *19*(6), 515–535.

Meade, N., & Islam, T. (2001). Forecasting the diffusion of innovations: Implications for time-series extrapolation. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 577–595). Boston, MA: Springer US.

Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation – a 25-year review. *International Journal of Forecasting*, *22*, 519–545.

Meade, N., & Islam, T. (2015a). Forecasting in telecommunications and ICT - a review. *International Journal of Forecasting*, *31*(4), 1105–1126.

Meade, N., & Islam, T. (2015b). Modelling European usage of renewable energy technologies for electricity generation. *Technological Forecasting and Social Change*, *90*, 497–509.

Meehl, P. (2013). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Echo Point Books & Media.

Meeran, S., Dyussekeneva, K., & Goodwin, P. (2013). Sales forecasting using a combination of diffusion model and forecast market: an adaptation of prediction/preference markets. In *Proceedings of the 7th IFAC conference on manufacturing modelling, management, and control* (pp. 87–92).

Meeran, S., Jahanbin, S., Goodwin, P., & Quariguasi Frota Neto, J. (2017). When do changes in consumer preferences make forecasts from choice-based conjoint models unreliable? *European Journal of Operational Research*, *258*(2), 512–524.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1–2), 3–24.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*, 983–999.

Meira, E., Cyrino Oliveira, F. L., & Jeon, J. (2020). Treating and pruning: new approaches to forecasting model selection and combination using prediction intervals. *International Journal of Forecasting*.

Melacini, M., Perotti, S., Rasini, M., & Tappia, E. (2018). E-fulfilment and distribution in omni-channel retailing: a systematic literature review. *International Journal of Physical Distribution and Logistics Management*, *48*(4), 391–414.

Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., & Lughi, V. (2020). Advanced methods for photovoltaic output power forecasting: A review. *Applied Sciences*, *10*(2), 487.

Mello, J. (2009). The impact of sales forecast game playing on supply chains. *Foresight: The International Journal of Applied Forecasting*, *13*, 13–22.

Mello, J. (2010). Corporate culture and s&op: Why culture counts. *Foresight: The International Journal of Applied Forecasting*, *16*, 46–49.

Mena-Oreja, J., & Gozalvez, J. (2020). A comprehensive evaluation of deep learning-based techniques for traffic prediction. *IEEE Access*, *8*, 91188–91212.

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., et al. (2016). MLlib: MAchine learning in apache spark. *Journal of Machine Learning Research*, *17*(1), 1235–1241.

Meng, X., & Taylor, J. W. (2020). Estimating value-at-risk and expected shortfall using the intraday low and range data. *European Journal of Operational Research*, *280*(1), 191–202.

Meng, X., Taylor, J. W., Ben Taieb, S., & Li, S. (2020). Scoring functions for multivariate distributions and level sets. ArXiv:2002.09578.

Merkle, E. C., & Steyvers, M. (2013). Choosing a strictly proper scoring rule. *Decision Analysis*, *10*(4), 292–304.

Merrick, J. R. W., Hardin, J. R., & Walker, R. (2006). Partnerships in training. *INFORMS Journal on Applied Analytics*, *36*(4), 359–370.

Merrow, E. W., McDonnell, L. M., & Arguden, R. Y. (1988). *Understanding the outcomes of mega-projects*. RAND Corporation.

Messner, J. W., & Pinson, P. (2018). Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*.

Mestre, G., Portela, J., San Roque, A. M. n., & Alonso, E. (2020). Forecasting hourly supply curves in the Italian day-ahead electricity market with a double-seasonal SARMAHX model. *International Journal of Electrical Power & Energy Systems*, *121*, Article 106083.

Miao, H., Ramchander, S., Wang, T., & Yang, D. (2017). Influential factors in crude oil price forecasting. *Energy Economics*, *68*, 77–88.

Miao, D. W. C., Wu, C. C., & Su, Y. K. (2013). Regime-switching in volatility and correlation structure using range-based models with Markov-switching. *Economic Modelling*, *31*(1), 87–93.

Mikkelsen, L., Moesgaard, K., Hegnauer, M., & Lopez, A. D. (2020). ANACONDA: A new tool to improve mortality and cause of death data. *BMC Medicine*, *18*(1), 1–13.

Milankovitch, M. (1969). *Canon of insolation and the ice-age problem*. Washington, D.C: National Science Foundation, English translation by the Israel Program for Scientific Translations of Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem, Textbook Publishing Company, Belgrade, 1941.

Milas, C., & Rothman, P. (2008). Out-of-sample forecasting of unemployment rates with pooled STVECM forecasts. *International Journal of Forecasting*, *24*(1), 101–121.

Millán-Ruiz, D., & Hidalgo, J. I. (2013). Forecasting call centre arrivals. *Journal of Forecasting*, *32*(7), 628–638.

Miller, R., & Lessard, D. (2007). *Evolving strategy: Risk management and the shaping of large engineering projects*: *Tech. Rep. 37157*, MIT Sloan School of Management, Massachusetts Institute of Technology.

Min, A., & Czado, C. (2011). BayesIan model selection for D-vine pair-copula constructions. *The Canadian Journal of Statistics*, *39*(2), 239–258.

Min, C.-k., & Zellner, A. (1993). BayesIan and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, *56*(1–2), 89–118.

Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. Mincer (Ed.), *Economic forecasts and expectations: analysis of forecasting behavior and performance* (pp. 3–46). National Bureau of Economic Research, Inc.

Mingming, T., & Jinliang, Z. (2012). A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices. *Journal of Economics and Business*, *64*(4), 275–286.

Mirakyan, A., Meyer-Renschhausen, M., & Koch, A. (2017). Composite forecasting approach, application for next-day electricity price forecasting. *Energy Economics*, *66*, 228–237.

Mircetica, D., Rostami-Tabar, B., Nikolicica, S., & Maslarica, M. (2020). *Forecasting hierarchical time series in supply chains: an empirical investigation*. Cardiff University.

Mirko, K., & Kantelhardt, J. W. (2013). Hadoop. TS: Large-scale time-series processing. *International Journal of Computer Applications*, *74*(17).

Mirmirani, S., & Li, H. C. (2004). A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. *Advances in Econometrics*, *19*, 203–223.

Mišić, S., & Radujković, M. (2015). Critical drivers of megaprojects success and failure. *Procedia Engineering*, *122*, 71–80.

Mitchell, T. J., & Beauchamp, J. J. (1988). BayesIan variable selection in linear regression. *Journal of the American Statistical Association*, *83*(404), 1023–1032.

Mitofsky, W. (1991). A short history of exit polls. In P. J. Lavrakas, & J. K. Holley (Eds.), *Polling and presidential election coverage* (pp. 83–99). Newbury Park, CA: Sage.

des Mobilités, L. F. (2020). Motorway traffic in Luxembourg. Accessed on 2020-09-01, URL https://www.kaggle.com/fabmob/motorway-traffic-in-luxembourg?select=datexDataA1.csv.

Modis, T. (1992). *Predictions: Society's telltale signature reveals the past and forecasts the future*. Simon & Schuster.

Modis, T. (1994). Fractal aspects of natural growth. *Technological Forecasting and Social Change*, *47*(1), 63–73.

Modis, T. (1997). Genetic re-engineering of corporations. *Technological Forecasting and Social Change*, *56*(2), 107–118.

Modis, T. (1998). *Conquering uncertainty: Understanding corporate cycles and positioning your company to survive the changing environment*. McGraw-Hill.

Modis, T. (2007). The normal, the natural, and the harmonic. *Technological Forecasting and Social Change*, *74*(3), 391–399.

Modis, T. (2013). *Natural laws in the service of the decision maker: How to use science-based methodologies to see more clearly further into the future*. Growth Dynamics.

Modis, T., & Debecker, A. (1992). Chaoslike states can be expected before and after logistic growth. *Technological Forecasting and Social Change*, *41*(2), 111–120.

Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, *21*(41), 89–93.

Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics*, *32*(5), 1001–1008.

Mohandes, S. R., Zhang, X., & Mahdiyar, A. (2019). A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing*, *340*, 55–75.

Molenaers, A., Baets, H., Pintelon, L., & Waeyenbergh, G. (2012). Criticality classification of spare parts: A case study. *International Journal of Production Economics*, *140*(2), 570–578.

Möller, A., Lenkoski, A., & Thorarinsdottir, T. L. (2013). Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, *139*(673), 982–991.

Molnár, P. (2016). High-low range in GARCH models of stock return volatility. *Applied Economics*, *48*(51), 4977–4991.

Molodtsova, T., & Papell, D. H. (2009). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics*, *77*(2), 167–180.

Monsell, B., Aston, J., & Koopman, S. (2003). Toward x-13? In *Proceedings of the american statistical association, section on business and economic statistics* (pp. 1–8). U.S. Census Bureau.

Montero Jimenez, J. J., Schwartz, S., Vingerhoeds, R., Grabot, B., & Salaün, M. (2020). Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems*, *56*, 539–557.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., & Talagala, T. S. (2020). FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting*, *36*(1), 86–92.

Montero-Manso, P., & Hyndman, R. J. (2020). Principles and algorithms for forecasting groups of time series: Locality and globality. arXiv:2008.00444.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., & Tiao, G. C. (1998). Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association*, *93*, 478–493.

Moon, M. A., Mentzer, J. T., & Smith, C. D. (2003). Conducting a sales forecasting audit. *International Journal of Forecasting*, *19*(1), 5–25.

Moon, S., Simpson, A., & Hicks, C. (2013). The development of a classification model for predicting the performance of forecasting methods for naval spare parts demand. *International Journal of Production Economics*, *143*(2), 449–454.

Moonchai, S., & Chutsagulprom, N. (2020). Short-term forecasting of renewable energy consumption: Augmentation of a modified grey model with a Kalman filter. *Applied Soft Computing*, *87*, Article 105994.

Mori, H., & Yuihara, A. (2001). Deterministic annealing clustering for ANN-based short-term load forecasting. *IEEE Transactions on Power Systems*, *16*(3), 545–551.

Morlidge, S. (2014a). Do forecasting methods reduce avoidable error? Evidence from forecasting competitions. *Foresight: The International Journal of Applied Forecasting*, *32*, 34–39.

Morlidge, S. (2014b). Forecast quality in the supply chain. *Foresight: The International Journal of Applied Forecasting*, *33*, 26–31.

Morlidge, S. (2014c). Using relative error metrics to improve forecast quality in the supply chain. *Foresight: The International Journal of Applied Forecasting*, *34*, 39–46.

Morris, S., & Pratt, D. (2003). Analysis of the Lotka–Volterra competition equations as a technological substitution model. *Technological Forecasting and Social Change*, *77*, 103–133.

Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: A survey of the US public. *Weather and Forecasting*, *23*(5), 974–991.

Morwitz, V. (1997). Why consumers don't always accurately predict their own future behavior. *Marketing Letters*, *8*(1), 57–70.

Moshman, J. (1964). The role of computers in election night broadcasting. In F. L. Alt, & M. Rubinoff (Eds.), *vol. 5*, *Advances in computers* (pp. 1–21). Elsevier.
Moultrie, T., Dorrington, R., Hill, A., Hill, K., Timæ us, I., & Zaba, B. (2013). Tools for demographic estimation. Paris: International Union for the Scientific Study of Population.
Mount, T. D., Ning, Y., & Cai, X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, *28*(1), 62–80.
Mueller, J. (2009a). *Retreat from Doomsday: The obsolescence of major war*. Zip Publishing.
Mueller, J. (2009b). War has almost ceased to exist: An assessment. *Political Science Quarterly*, *124*(2), 297–321.
Mukhopadhyay, S., & Sathish, V. (2019). Predictive likelihood for coherent forecasting of count time series. *Journal of Forecasting*, *38*(3), 222–235.
Mulholland, J., & Jensen, S. T. (2019). Optimizing the allocation of funds of an NFL team under the salary cap. *International Journal of Forecasting*, *35*(2), 767–775.
Muniain, P., & Ziel, F. (2020). Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. *International Journal of Forecasting*, *36*(4), 1193–1210.
Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, *8*(2), 281–293.
Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*, 315–335.
Myrskylä, M., Goldstein, J. R., & Cheng, Y.-H. A. (2013). New cohort fertility forecasts for the developed world: Rises, falls, and reversals. *Population and Development Review*, *39*(1), 31–56.
Nagi, J., Yap, K. S., Nagi, F., Tiong, S. K., & Ahmed, S. K. (2011). A computational intelligence scheme for the prediction of the daily peak load. *Applied Soft Computing*, *11*(8), 4773–4788.
Naish, S., Dale, P., Mackenzie, J. S., McBride, J., Mengersen, K., & Tong, S. (2014). Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC Infectious Diseases*, *14*(1), 1–14.
Nanopoulos, A., Alcock, R., & Manolopoulos, Y. (2001). Feature-based classification of time-series data. In *Information processing and technology* (pp. 49–61). USA: Nova Science Publishers, Inc..
Napierała, J., Hilton, J., Forster, J. J., Carammia, M., & Bijak, J. (2021). Towards an early warning system for monitoring asylum-related migration flows in Europe. *International Migration Review*, in press.
Narajewski, M., & Ziel, F. (2020a). Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets*, *19*, Article 100107.
Narajewski, M., & Ziel, F. (2020b). Ensemble forecasting for intraday electricity prices: Simulating trajectories. *Applied Energy*, *279*, Article 115801.
National Research Council (2000). *Beyond Six Billion: Forecasting the World's Population*. National Academies Press.
National Research Council (2006). *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions using Weather and Climate Forecasts*. National Academies Press.
Neal, P., & Kypraios, T. (2015). Exact Bayesian inference via data augmentation. *Statistics and Computing*, *25*, 333–347.
Neale, W. C. (1964). The peculiar economics of professional sports. *Quarterly Journal of Economics*, *78*(1), 1–14.
Nelsen, R. (2006). *An introduction to copulas*. Springer Verlag.
Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347.
Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, *10*(2), 139–162.
Nespoli, A., Ogliari, E., Leva, S., Massi Pavan, A., Mellit, A., Lughi, V., et al. (2019). Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. *Energies*, *12*(9), 1621.
Neves, M. M., & Cordeiro, C. (2020). Modellling (and forecasting) extremes in time series: a naive approach. In *Atas do XXIII congresso da SPE* (pp. 189–202). Sociedade Portuguesa de Estatística.
Newbold, P., & Granger, C. W. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society: Series A (General)*, *137*(2), 131–146.
Ng, Y. S., Stein, J., Ning, M., & Black-Schaffer, R. M. (2007). Comparison of clinical characteristics and functional outcomes of ischemic stroke in different vascular territories. *Stroke*, *38*(8), 2309–2314.
Nicol-Harper, A., Dooley, C., Packman, D., Mueller, M., Bijak, J., Hodgson, D., et al. (2018). Inferring transient dynamics of human populations from matrix non-normality. *Population Ecology*, *60*(1), 185–196.
Nielsen, J., Mazick, A., Andrews, N., Detsis, M., Fenech, T., Flores, V., et al. (2013). Pooling European all-cause mortality: methodology and findings for the seasons 2008/2009 to 2010/2011. *Epidemiology & Infection*, *141*(9), 1996–2010.
Nielsen, M., Seo, W., & Seong, D. (2019). Inference on the dimension of the nonstationary subspace in functional time series 1420. Queen's Economics Department.
Nikolopoulos, K. (2020). We need to talk about intermittent demand forecasting. *European Journal of Operational Research*.
Nikolopoulos, K., Assimakopoulos, V., Bougioukos, N., Litsa, A., & Petropoulos, F. (2012). The theta model: An essential forecasting tool for supply chain planning. In *Advances in automation and robotics, Vol. 2* (pp. 431–437). Springer Berlin Heidelberg.
Nikolopoulos, K. I., Babai, M. Z., & Bozos, K. (2016). Forecasting supply chain sporadic demand with nearest neighbor approaches. *International Journal of Production Economics*, *177*, 139–148.
Nikolopoulos, K., Goodwin, P., Patelis, A., & Assimakopoulos, V. (2007). Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches. *European Journal of Operational Research*, *180*(1), 354–368.
Nikolopoulos, K., Litsa, A., Petropoulos, F., Bougioukos, V., & Khammash, M. (2015). Relative performance of methods for forecasting special events. *Journal of Business Research*, *68*(8), 1785–1791.
Nikolopoulos, K., & Petropoulos, F. (2018). Forecasting for big data: Does suboptimality matter? *Computers & Operations Research*, *98*, 322–329.
Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2020). Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research*.
Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., & Assimakopoulos, V. (2011). An aggregate - disaggregate intermittent demand approach (ADIDA) to forecasting: An empirical proposition and analysis. *Journal of the Operational Research Society*, *62*(3), 544–554.
Nikolopoulos, K. I., & Thomakos, D. D. (2019). *Forecasting with the theta method: theory and applications*. John Wiley & Sons.
Nogueira, P. J., de Araújo Nobre, M., Nicola, P. J., Furtado, C., & Carneiro, A. V. (2020). Excess mortality estimation during the COVID-19 pandemic: preliminary data from Portugal. *Acta Médica Portuguesa*, *33*(13).
Nordhaus, W. D. (1987). Forecasting efficiency: Concepts and applications. *The Review of Economics and Statistics*, *69*(4), 667–674.
Norton-Taylor, R. (2015). Global armed conflicts becoming more deadly, major study finds. In *the Guardian*. Accessed on 2020-09-12, http://www.theguardian.com/world/2015/may/20/.
Nowotarski, J., & Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, *81*, 1548–1568.
Nsoesie, E., Mararthe, M., & Brownstein, J. (2013). Forecasting peaks of seasonal influenza epidemics. *PLoS Currents*, *5*.
Nunes, B., Viboud, C., Machado, A., Ringholz, C., Rebelo-de Andrade, H., Nogueira, P., et al. (2011). Excess mortality associated with influenza epidemics in Portugal, 1980 to 2004. *PLoS One*, *6*(6), Article e20661.
Nye, J. S. (1990). The changing nature of world power. *Political Science Quarterly*, *105*(2), 177–192.
Nymoen, R., & Sparrman, V. (2015). Equilibrium unemployment dynamics in a panel of OECD countries. *Oxford Bulletin of Economics and Statistics*, *77*(2), 164–190.
Nystrup, P., Lindström, E., Pinson, P., & Madsen, H. (2020). Temporal hierarchies with autocorrelation for load forecasting. *European Journal of Operational Research*, *280*(3), 876–888.
OBR (2019). Long-term economic determinants. London: Office for Budget Responsibility.
Obst, D., Ghattas, B., Claudel, S., Cugliari, J., Goude, Y., & Oppenheim, G. (2019). Textual data for time series forecasting. arXiv:1910.12618.

O'Connor, M., Remus, W., & Griggs, K. (1993). Judgemental forecasting in times of change. *International Journal of Forecasting*, *9*(2), 163–172.
Office for National Statistics (2019). U.K. national accounts, the blue book: 2019. Office for National Statistics.
Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *30*(2), 243–261.
Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27.
Ogliari, E., Dolara, A., Manzolini, G., & Leva, S. (2017). Physical and hybrid methods comparison for the day ahead PV output power forecast. *Renewable Energy*, *113*, 11–21.
Ogliari, E., Niccolai, A., Leva, S., & Zich, R. E. (2018). Computational intelligence techniques applied to the day ahead PV output power forecast: PHANN, SNO and mixed. *Energies*, *11*(6), 1487.
Oh, D. H., & Patton, A. J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, *193*(2), 349–366.
Oh, D. H., & Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business & Economic Statistics*, *36*(2), 181–195.
Oh, H., & Yoon, C. (2020). Time to build and the real-options channel of residential investment. *Journal of Financial Economics*, *135*(1), 255–269.
O'Hagan, A., Buck, C. E., Daneshkhah, A., Richard Eiser, J., Garthwaite, P. H., Jenkinson, D. J., et al. (2006). Uncertain judgements: Eliciting experts' probabilities. Wiley.
O'Hagan, A., & Forster, J. (2004). *Kendall's advanced theory of statistics: Bayesian inference, vol. 2B* (second ed). Arnold.
O'Hagan, A., & West, M. (2010). *The oxford handbook of applied bayesian analysis*. OUP.
O'Hara-Wild, M., & Hyndman, R. (2020). Fasster: Fast additive switching of seasonality, trend and exogenous regressors. R package version 0.1.0.9100.
Ojo, O. O., Shah, S., Coutroubis, A., Jiménez, M. T., & Ocana, Y. M. (2018). Potential impact of industry 4.0 in sustainable food supply chain environment. In *2018 IEEE international conference on technology management, operations and decisions* (pp. 172–177). IEEE.
Oksuz, I., & Ugurlu, U. (2019). Neural network based model comparison for intraday electricity price forecasting. *Energies*, *12*(23), 4557.
Okun, A. M. (1962). Potential GNP: Its measurement and significance. *American Statistical Association, Proceedings of the Business and Economics Statistics Section*, 98–104.
Oliva, R., & Watson, N. (2009). Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *International Journal of Operations & Production Management*, *18*(2), 138–151.
Oliveira, J. M., & Ramos, P. c. (2019). Assessing the performance of hierarchical forecasting methods on the retail sector. *Entropy*, *21*(4).
Oliveira, F. L. C., Souza, R. C., & Marcato, A. L. M. (2015). A time series model for building scenarios trees applied to stochastic optimisation. *International Journal of Electrical Power & Energy Systems*, *67*, 315–323.
Omar, H., Klibi, W., Babai, M. Z., & Ducq, Y. (2021). Basket data-driven approach for omnichannel demand forecasting. working paper.
Önkal, D., & Gönül, M. S. (2005). Judgmental adjustment: A challenge for providers and users of forecasts. *Foresight: The International Journal of Applied Forecasting*, *1*, 13–17.
Önkal, D., Gönül, M. S., & De Baets, S. (2019). Trusting forecasts. *Futures & Foresight Science*, *1*, Article e19.
Önkal, D., Gönül, M. S., Goodwin, P., Thomson, M., & Öz, E. (2017). Evaluating expert advice in forecasting: Users' reactions to presumed vs. experienced credibility. *International Journal of Forecasting*, *33*(1), 280–297.
Önkal, D., Gönül, M. S., & Lawrence, M. (2008). Judgmental adjustments of previously adjusted forecasts. *Decision Sciences*, *39*(2), 213–238.
Önkal, D., Goodwin, P., Thomson, M., Gönül, M. S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*(4), 390–409.
Önkal, D., Sayım, K. Z., & Gönül, M. S. (2013). Scenarios as channels of forecast advice. *Technological Forecasting and Social Change*, *80*(4), 772–788.
Ord, K., & Fildes, R. (2013). *Principles of business forecasting* (1st ed.). South-Western Cengage Learning, Mason, OH and Andover, UK.
Ord, J. K., Fildes, R., & Kourentzes, N. (2017). *Principles of business forecasting* (2nd ed). Wessex Press Publishing Co..
Ordu, M., Demir, E., & Tofallis, C. (2019). A comprehensive modelling framework to forecast the demand for all hospital services. *The International Journal of Health Planning and Management*, *34*(2), e1257–e1271.
Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2020a). Meta-learning framework with applications to zero-shot time-series forecasting. arXiv:2002.02887.
Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2020b). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv:1905.10437.
Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics*, *31*(1), 145–155.
Ozer, M. (2011). Understanding the impacts of product knowledge and product type on the accuracy of intentions-based new product predictions. *European Journal of Operational Research*, *211*(2), 359–369.
Özer, O., Zheng, Y., & Chen, K.-Y. (2011). Trust in forecast information sharing. *Management Science*, *57*(6), 1111–1137.
Paccagnini, A. (2017). Dealing with misspecification in DSGE models: A survey 82914. University Library of Munich, Germany.
Pacheco, J., Millán-Ruiz, D., & Vélez, J. L. (2009). Neural networks for forecasting in a multi-skill call centre. In *International conference on engineering applications of neural networks* (pp. 291–300). Springer.
Pai, J., & Pedersen, H. (1999). Threshold models of the term structure of interest rate. In *Joint day proceedings volume of the XXXth international ASTIN colloquium/9th international AFIR colloquium* (pp. 387–400).
Paillard, D. (2001). Glacial cycles: towards a new paradigm. *Reviews of Geophysics*, *39*, 325–346.
Pal, D., & Mitra, S. K. (2019). Correlation dynamics of crude oil with agricultural commodities: A comparison between energy and food crops. *Economic Modelling*, *82*, 453–466.
Palm, F. C., & Zellner, A. (1992). To combine or not to combine? Issues of combining forecasts. *Journal of Forecasting*, *11*(8), 687–701.
Panagiotelis, A., Athanasopoulos, G., Hyndman, R. J., Jiang, B., & Vahid, F. (2019). Macroeconomic forecasting for Australia using a large number of predictors. *International Journal of Forecasting*, *35*(2), 616–633.
Panagiotelis, A., Czado, C., & Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, *107*(499), 1063–1072.
Panagiotelis, A., Czado, C., Joe, H., & Stöber, J. (2017). Model selection for discrete regular vine copulas. *Computational Statistics & Data Analysis*, *106*, 138–152.
Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., & Hyndman, R. J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, *37*(1), 343–359.
Panahifar, F., Byrne, P. J., & Heavey, C. (2015). A hybrid approach to the study of CPFR implementation enablers. *Production Planning and Control*, *26*(13), 1090–1109.
Panda, C., & Narasimhan, V. (2007). Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling*, *29*(2), 227–236.
Pankratz, A., & Dudley, U. (1987). Forecasts of power-transformed series. *Journal of Forecasting*, *6*(4), 239–248.
Parag, Y., & Sovacool, B. K. (2016). Electricity market design for the prosumer era. *Nature Energy*, *1*(4), 1–6.
Paredes, J., Pedregal, D. J., & Pérez, J. J. (2014). Fiscal policy analysis in the euro area: Expanding the toolkit. *Journal of Policy Modeling*, *36*, 800–823.
Park, B.-J. (2002). An outlier robust GARCH model and forecasting volatility of exchange rate returns. In *Handbook of E*: *Journal of Forecasting*, In *Handbook of E*: *21*(5).381–393,
Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, *3*(2), 246–257.
Park, S. Y., Yun, B.-Y., Yun, C. Y., Lee, D. H., & Choi, D. G. (2016). An analysis of the optimum renewable energy portfolio using the bottom–up model: Focusing on the electricity generation sector in South Korea. *Renewable and Sustainable Energy Reviews*, *53*, 319–329.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, *53*(1), 61–65.
Pastore, E., Alfieri, A., Zotteri, G., & Boylan, J. E. (2020). The impact of demand parameter uncertainty on the bullwhip effect. *European Journal of Operational Research*, *283*(1), 94–107.
Patel, J. (1989). Prediction intervals - a review. *Communications in Statistics. Theory and Methods*, *18*(7), 2393–2465.
Patterson, K. D. (1995). An integrated model of the data measurement and data generation processes with an application to consumers' expenditure. *The Economic Journal*, *105*, 54–76.
Patti, E., Acquaviva, A., Jahn, M., Pramudianto, F., Tomasi, R., Rabourdin, D., et al. (2016). Event-driven user-centric middleware for energy-efficient buildings and public spaces. *IEEE Systems Journal*, *10*(3), 1137–1146.
Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, *47*(2), 527–556.
Patton, A. J. (2006a). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics*, *21*(2), 147–173.
Patton, A. (2013). Copula methods for forecasting multivariate time series. In *Handbook of economic forecasting, vol. 2* (pp. 899–960). Elsevier.
Patton, A. J., & Timmermann, A. (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, *102*, 1172–1184.
Pavía, J. M. (2010). Improving predictive accuracy of exit polls. *International Journal of Forecasting*, *26*(1), 68–81.
Pavia, J. M., Cabrer, B., & Sala, R. (2009). Updating input–output matrices: assessing alternatives through simulation. *Journal of Statistical Computation and Simulation*, *79*(12), 1467–1482.
Pavía, J. M., Gil-Carceller, I., Rubio-Mataix, A., Coll, V., Alvarez-Jareño, J. A., Aybar, C., et al. (2019). The formation of aggregate expectations: wisdom of the crowds or media influence? *Contemporary Social Science*, *14*(1), 132–143.
Pavía, J. M., & Larraz, B. (2012). Nonresponse bias and superpopulation models in electoral polls. *Reis*, *137*(1), 237–264.
Pavía, J. M., & Romero, R. (2021). Improving estimates accuracy of voter transitions. Two new algorithms for ecological inference based on linear programming. *Advance*.
Pavía-Miralles, J. M. (2005). Forecasts from nonrandom samples. *Journal of the American Statistical Association*, *100*(472), 1113–1122.
Pavía-Miralles, J. M., & Larraz-Iribas, B. (2008). Quick counts from non-selected polling stations. *Journal of Applied Statistics*, *35*(4), 383–405.
Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, *92*(2), 382–402.
Peña, I., Martinez-Anido, C. B., & Hodge, B.-M. (2018). An extended IEEE 118-bus test system with high renewable penetration. *IEEE Transactions on Power Systems*, *33*(1), 281–289.
Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
Pearl, R., & Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences of the United States of America*, *6*(6), 275–288.
Pedregal, D. J., & Carmen Carnero, M. (2006). State space models for condition monitoring: a case study. *Reliability Engineering & System Safety*, *91*(2), 171–180.
Pedregal, D. J., García, F. P., & Roberts, C. (2009). An algorithmic approach for maintenance management based on advanced state space systems and harmonic regressions. *Annals of Operations Research*, *166*(1), 109–124.
Pedregal, D. J., & Pérez, J. J. (2010). Should quarterly government finance statistics be used for fiscal surveillance in europe? *International Journal of Forecasting*, *26*, 794–807.
Pedregal, D. J., Pérez, J. J., & Sánchez, A. J. (2014). A toolkit to strengthen government budget surveillance. *Review of Public Economics*, *211*, 117–146.
Peel, D. A., & Speight, A. (2000). Threshold nonlinearities in unemployment rates: Further evidence for the UK and G3 economies. *Applied Economics*, *32*(6), 705–715.
Pegels, C. C. (1969). Exponential forecasting: Some new variations. *Management Sience*, *15*(5), 311–315.
Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics*, *131*(1), 445–473.
Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance*, *12*(3), 30–32.
Pennings, C. L., & van Dalen, J. (2017). Integrated hierarchical forecasting. *European Journal of Operational Research*, *263*(2), 412–418.
Pennings, C. L. P., van Dalen, J., & Rook, L. (2019). Coordinating judgmental forecasting: Coping with intentional biases. *Omega*, *87*, 46–56.
Pereira, L. N. (2016). An introduction to helpful forecasting methods for hotel revenue management. *International Journal of Hospitality Management*, *58*, 13–23.
Perera, H. N., Hurley, J., Fahimnia, B., & Reisi, M. (2019). The human factor in supply chain forecasting: A systematic review. *European Journal of Operational Research*, *274*(2), 574–600.
Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, *27*, 91–106.
Pesaran, M. H., Pick, A., & Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, *177*(2), 134–152.
Pesaran, M. H. M. H., Pick, A., & Timmermann, A. (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, *164*(250), 173–187.
Pesaran, M. H., Shin, Y., & Smith, R. P. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association*, *94*(446), 621–634.
Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference*. MIT Press.
Petropoulos, F. (2015). Forecasting support systems: Ways forward. *Foresight: The International Journal of Applied Forecasting*, *39*, 5–11.
Petropoulos, F., Fildes, R., & Goodwin, P. (2016). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? *European Journal of Operational Research*, *249*(3), 842–852.
Petropoulos, F., Goodwin, P., & Fildes, R. (2017). Using a rolling training approach to improve judgmental extrapolations elicited from forecasters with technical knowledge. *International Journal of Forecasting*, *33*(1), 314–324.
Petropoulos, F., Hyndman, R. J., & Bergmeir, C. (2018a). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, *268*(2), 545–554.
Petropoulos, F., & Kourentzes, N. (2014). Improving forecasting via multiple temporal aggregation. *Foresight: The International Journal of Applied Forecasting*, *34*, 12–17.
Petropoulos, F., & Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, *66*(6), 914–924.
Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E. (2018b). Judgmental selection of forecasting models. *Journal of Operations Management*, *60*, 34–46.
Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS One*, *15*(3), Article e0231236.
Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). 'Horses for courses' in demand forecasting. *European Journal of Operational Research*, *237*(1), 152–163.
Petropoulos, F., Makridakis, S., & Stylianou, N. (2020). COVID-19: Forecasting confirmed cases and deaths with a simple time-series model. *International Journal of Forecasting*.
Pettenuzzo, D., & Ravazzolo, F. (2016). Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics*, *31*(7), 1312–1332.
Pfann, G. A., Schotman, P. C., & Tschernig, R. (1996). Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics*, *74*(1), 149–176.
Phillips, A. W. H. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica*, *25*, 283–299.
Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, *55*(2), 277–301.
Phillips, D. E., Adair, T., & Lopez, A. D. (2018). How useful are registered birth statistics for health and social policy? A global systematic assessment of the availability and quality of birth registration data. *Population Health Metrics*, *16*(1), 1–13.

Pierce, M. A., Hess, E. P., Kline, J. A., Shah, N. D., Breslin, M., Branda, M. E., et al. (2010). The chest pain choice trial: a pilot randomized trial of a decision aid for patients with chest pain in the emergency department. *Trials*, *11*(1), 1–8.

Piironen, J., & Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, *27*(3), 711–735.

Pinheiro Neto, D., Domingues, E. G., Coimbra, A. P., de Almeida, A. T., Alves, A. J., & Calixto, W. P. (2017). Portfolio optimization of renewable energy assets: Hydro, wind, and photovoltaic energy in the regulated market in Brazil. *Energy Economics*, *64*, 238–250.

Pinker, S. (2011). *The better angels of our nature: the decline of violence in history and its causes*. Penguin UK.

Pinker, S. (2018). *Enlightenment now: the case for reason, science, humanism, and progress*. Penguin.

Pinson, P. (2012). Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, 555–576.

Pinson, P., Chevallier, C., & Kariniotakis, G. (2007). Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transaction on Power Systems*, *22*(3), 1148–1156.

Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., & Klöckl, B. (2009). From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, *12*(1), 51–62.

Pinson, P., & Makridakis, S. (2020). Pandemics and forecasting: The way forward through the Taleb-Ioannidis debate. *International Journal of Forecasting*.

Pinson, P., Reikard, G., & Bidlot, J.-R. (2012). Probabilistic forecasting of the wave energy flux. *Applied Energy*, *93*, 364–370.

Pinson, P., & Tastu, J. (2013). Discrimination ability of the energy score. *Technical University of Denmark (DTU)*.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675.

Pitt, M., Chan, D., & Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, *93*(3), 537–554.

Plescia, C., & De Sio, L. (2018). An evaluation of the performance and suitability of r × c methods for ecological inference with known true values. *Quality & Quantity*, *52*(2), 669–683.

Plott, C., & Chen, K.-Y. (2002). Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem, 1131. California Institute of Technology, Division of the Humanities and Social Sciences.

Poccia, D. (2019). Amazon forecast – now generally available. In *AWS news blog*. https://aws.amazon.com/blogs/aws/amazon-forecast-now-generally-available/ (Accessed on 01 September 2020).

Politi, M. C., Han, P. K., & Col, N. F. (2007). Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making*, *27*(5), 681–695.

Politis, D. N., & Romano, J. P. (1992). A circular block-resampling procedure for stationary data. *Exploring the Limits of Bootstrap*, *2635270*.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313.

Polk, C., Haghbin, M., & de Longis, A. (2020). Time-series variation in factor premia: The influence of the business cycle. *Journal of Investment Management*, *18*(1).

Polson, N. G., & Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C (Emerging Technologies)*, *79*, 1–17.

Porras, E., & Dekker, R. (2008). An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods. *European Journal of Operational Research*, *184*(1), 101–132.

Powell, W. B. (2019). A unified framework for stochastic optimization. *European Journal of Operational Research*, *275*(3), 795–821.

Poynting, J. H. (1884). A comparison of the fluctuations in the price of wheat and in the cotton and silk imports into great britain. *Journal of the Statistical Society of London*, *47*(1), 34–74.

Pradeepkumar, D., & Ravi, V. (2017). Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. *Applied Soft Computing*, *58*, 35–52.

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702.

Prak, D., Teunter, R., & Syntetos, A. (2017). On the calculation of safety stocks when demand is forecasted. *European Journal of Operational Research*, *256*(2), 454–461.

Preston, S., Heuveline, P., & Guillot, M. (2000). *Demography: Measuring and modeling population processes*. Wiley.

Prestwich, S. D., Tarim, S. A., Rossi, R., & Hnich, B. (2014). Forecasting intermittent demand by hyperbolic-exponential smoothing. *International Journal of Forecasting*, *30*(4), 928–933.

Pretis, F. (2020). Econometric modelling of climate systems: The equivalence of energy balance models and cointegrated vector autoregressions. *Journal of Econometrics*, *214*(1), 256–273.

Pretis, F., & Kaufmann, R. K. (2018). Out-of-sample paleo-climate simulations: Testing hypotheses about the mid-brunhes event, the stage 11 paradox, and orbital variations. Canada: University of Victoria.

Pretis, F., & Kaufmann, R. K. (2020). Managing carbon emissions to avoid the next ice age. Canada: University of Victoria.

Pretis, F., Reade, J., & Sucarrat, G. (2017). gets: GEneral-to-specific (GETS) modelling and indicator saturation methods. R package version 0.12.

Pretis, F., Reade, J., & Sucarrat, G. (2018). Automated general-to-specific (GETS) regression modeling and indicator saturation for outliers and structural breaks. *Journal of Statistical Software*, *86*(3).

Pretis, F., Schneider, L., & Smerdon, J. (2016). Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation. *Journal of Economic Surveys*, *30*(3), 403–429.

Pritularga, K. F., Svetunkov, I., & Kourentzes, N. (2021). Stochastic coherency in forecast reconciliation. *International Journal of Production Economics*, *240*, Article 108221, URL https://www.sciencedirect.com/science/article/pii/S0925527321001973.

Programme, U. N. D. (2019). Population facts no. 2019/6, december 2019: How certain are the united nations global population projections?. (2019/6), Department of Economic and Social Affairs, Population Division.

Proietti, T. (2003). Forecasting the US unemployment rate. *Computational Statistics & Data Analysis*, *42*, 451–476.

Promprou, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2006). Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA models. *Dengue Bulletin*, *30*, 99–106.

Prudêncio, R. B., & Ludermir, T. B. (2004). Meta-learning approaches to selecting time series models. *Neurocomputing*, *61*, 121–137.

Psaradellis, I., & Sermpinis, G. (2016). Modelling and trading the u.s. implied volatility indices. Evidence from the VIX, VXN and VXD indices. *International Journal of Forecasting*, *32*(4), 1268–1283.

Puig, X., & Ginebra, J. (2015). Ecological inference and spatial variation of individual behavior: National divide and elections in catalonia. *Geographical Analysis*, *47*(3), 262–283.

Qiao, Z., Wu, X., Ge, S., & Fan, W. (2019). MNN: multimodal attentional neural networks for diagnosis prediction. *Extraction*, *1*, A1.

Qu, X., Kang, X., Zhang, C., Jiang, S., & Ma, X. (2016). Short-term prediction of wind power based on deep long short-term memory. In *2016 IEEE PES asia-pacific power and energy engineering conference* (pp. 1148–1152).

Quaedvlieg, R. (2019). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, 1–14.

Quiroz, M., Nott, D. J., & Kohn, R. (2018). Gaussian variational approximation for high-dimensional state space models. arXiv:1801.07873.

R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rabanser, S., Januschowski, T., Flunkert, V., Salinas, D., & Gasthaus, J. (2020). The effectiveness of discretization in forecasting: An empirical study on neural time series models. arXiv:2005.10111.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics*, *99*(1), 39–61.

Raftery, A. E. (1993). BayesIan model selection in structural equation models. In K. Bollen, & J. Long (Eds.), *Testing structural equation models* (pp. 163–180). Newbury Park, CA: Sage.

Raftery, A. E. (2016). Use and communication of probabilistic forecasts. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *9*(6), 397–410.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). BayesIan model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191.

Rahman, S., & Serletis, A. (2012). Oil price uncertainty and the Canadian economy: Evidence from a VARMA, GARCH-in-mean, asymmetric BEKK model. *Energy Economics*, *34*(2), 603–610.
Rajvanshi, V. (2015). Performance of range and return based volatility estimators: evidence from Indian crude oil futures market. *Global Economy and Finance Journal*, *8*(1), 46–66.
Ramos, M.-H., Mathevet, T., Thielen, J., & Pappenberger, F. (2010). Communicating uncertainty in hydro-meteorological forecasts: mission impossible? *Meteorological Applications*, *17*(2), 223–235.
Ramos, P., & Oliveira, J. M. (2016). A procedure for identification of appropriate state space and ARIMA models based on time-series cross-validation. *Algorithms*, *9*(4), 76.
Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, *34*, 151–163.
Ranawana, R., & Palade, V. (2006). Optimized precision-a new measure for classifier performance evaluation. In *2006 IEEE international conference on evolutionary computation* (pp. 2254–2261). IEEE.
Rangapuram, S. S., de Bezenac, E., Benidis, K., Stella, L., & Januschowski, T. (2020). Normalizing Kalman filters for multivariate time series analysis. In *Advances in neural information processing systems* (pp. 7785–7794).
Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., & Januschowski, T. (2018). Deep state space models for time series forecasting. In *Advances in neural information processing systems* (pp. 7785–7794).
Rangarajan, P., Mody, S. K., & Marathe, M. (2019). Forecasting dengue and influenza incidences using a sparse representation of google trends, electronic health records, and time series data. *PLoS Computational Biology*, *15*(11), Article e1007518.
Ranjan, R., & Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *72*(1), 71–91.
Rao, J. K., Anderson, L. A., Sukumar, B., Beauchesne, D. A., Stein, T., & Frankel, R. M. (2010). Engaging communication experts in a delphi process to identify patient behaviors that could enhance communication in medical encounters. *BMC Health Services Research*, *10*, 97.
Rao, K., & Kishore, V. (2010). A review of technology diffusion models with special reference to renewable energy technologies. *Renewable and Sustainable Energy Reviews*, *14*(3), 1070–1078.
Rao, Y., & McCabe, B. (2016). Real-time surveillance for abnormal events: the case of influenza outbreaks. *Statistics in Medicine*, *35*(13), 2206–2220.
Rapach, D. E., & Strauss, J. K. (2009). Differences in housing price forecastability across U.S. states. *International Journal of Forecasting*, *25*(2), 351–372.
Rapach, D. E., Strauss, J. K., Tu, J., & Zhou, G. (2019). Industry return predictability: A machine learning approach. *The Journal of Financial Data Science*, *1*(3), 9–28.
Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, *23*(2), 821–862.
Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the United States? *The Journal of Finance*, *68*(4), 1633–1662.
Rapach, D. E., & Zhou, G. (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. In E. Jurczenko (Ed.), *Machine learning for asset management: New developments and financial applications* (pp. 1–34). Hoboken, NJ: Wiley.
Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689.
Ravishanker, N., yen Wu, L. S., & Glaz, J. (1991). Multiple prediction intervals for time series: comparison of simultaneous and marginal intervals. *Journal of Forecasting*, *10*(5), 445–463.
Raymer, J., & Wiśniowski, A. (2018). Applying and testing a forecasting model for age and sex patterns of immigration and emigration. *Population Studies*, *72*(3), 339–355.
Reade, J., Singleton, C., & Brown, A. (2020). Evaluating strange forecasts: The curious case of football match scorelines. *Scottish Journal of Political Economy*.
Rebollo, J., & Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. *Transportation Research Part C (Emerging Technologies)*, *44*, 231–241.
Rees, P. H., & Wilson, A. G. (1973). Accounts and models for spatial demographic analysis I: Aggregate population. *Environment & Planning A*, *5*(1), 61–90.
Reggiani, P., & Boyko, O. (2019). A Bayesian processor of uncertainty for precipitation forecasting using multiple predictors and censoring. *Monthly Weather Review*, *147*(12), 4367–4387.
Reid, D. (1972). A comparison of forecasting techniques on economic time series. *Forecasting in Action. Operational Research Society and the Society for Long Range Planning*.
Reikard, G., Pinson, P., & Bidlot, J.-R. (2011). Forecasting ocean wave energy: The ECMWF wave model and time series methods. *Ocean Engineering*, *38*(10), 1089–1099.
Reikard, G., Robertson, B., Buckham, B., Bidlot, J.-R., & Hiles, C. (2015). Simulating and forecasting ocean wave energy in western Canada. *Ocean Engineering*, *103*, 223–236.
Reimers, S., & Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting*, *27*(4), 1196–1214.
Rendall, M. S., Handcock, M. S., & Jonsson, S. H. (2009). BayesIan estimation of hispanic fertility hazards from survey and population data. *Demography*, *46*(1), 65–83.
Renzl, B. (2008). Trust in management and knowledge sharing: The mediating effects of fear and knowledge documentation. *Omega*, *36*(2), 206–220.
Riahi, N., Hosseini-Motlagh, S.-M., & Teimourpour, B. (2013). A three-phase hybrid times series modeling framework for improved hospital inventory demand forecast. *International Journal of Hospital Research*, *2*(3), 133–142.
Rice, G., Wirjanto, T., & Zhao, Y. (2020). Tests for conditional heteroscedasticity of functional data. *Journal of Time Series Analysis*, *41*(6), 733–758.
Richardson, L. F. (1948). Variation of the frequency of fatal quarrels with magnitude. *Journal of the American Statistical Association*, *43*(244), 523–546.
Richardson, L. F. (1960). *Statistics of deadly quarrels*. Boxwood Press.
Riedel, K. (2021). The value of the high, low and close in the estimation of Brownian motion. *Statistical Inference for Stochastic Processes*, *24*, 179–210.
Rios, I., Wets, R. J.-B., & Woodruff, D. L. (2015). Multi-period forecasting and scenario generation with limited data. *Computational Management Science*, *12*, 267–295.
Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., et al. (2020). Coronavirus pandemic (COVID-19). https://ourworldindata.org/coronavirus (Accessed on 20 August 2020).
Riveiro, M., Helldin, T., Falkman, G., & Lebram, M. (2014). Effects of visualizing uncertainty on decision-making in a target identification scenario. *Computers & Graphics*, *41*, 84–98.
Roberts, J. M. (2001). Estimates of the productivity trend using time-varying parameter techniques. *The BE Journal of Macroeconomics*, *1*(1).
Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357.
Rodriguez, J. C. (2007). Measuring financial contagion: A copula approach. *Journal of Empirical Finance*, *14*(3), 401–423.
Rodríguez-Sanz, A., Comendador, F., Valdés, R., Pérez-Castán, J., Montes, R. B., & Serrano, S. (2019). Assessment of airport arrival congestion and delay: Prediction and reliability. *Transportation Research Part C (Emerging Technologies)*, *98*, 255–283.
Rogers, A. (1975). *Introduction to multiregional mathematical demography*. New York: Wiley.
Rogers, L. C. G., & Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *Annals of Applied Probability*, *1*(4), 504–512.
Romero, D., Olivero, J., Real, R., & Guerrero, J. C. (2019). Applying fuzzy logic to assess the biogeographical risk of dengue in south america. *Parasites & Vectors*, *12*(1), 1–13.
Romero, R., Pavía, J. M., Martín, J., & Romero, G. (2020). Assessing uncertainty of voter transitions estimated from aggregated data. Application to the 2017 french presidential election. *Journal of Applied Statistics*, *47*(13–15), 2711—2736.
Rosen, O., Jiang, W., King, G., & Tanner, M. A. (2001). BayesIan and frequentist inference for ecological inference: The rxc case. *Statistica Neerlandica*, *55*(2), 134–156.
Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, *23*, 470–472.

Rossi, B. (2005). Testing long-horizon predictive ability with high persistence, and the meese–rogoff puzzle. *International Economic Review*, *46*(1), 61–92.
Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, *51*(4), 1063–1119.
Rossi, B., & Sekhposyan, T. (2016). Forecast rationality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. *Journal of Applied Econometrics*, *31*(3), 507–532.
Rostami-Tabar, B., Babai, M. Z., Syntetos, A., & Ducq, Y. (2013). Demand forecasting by temporal aggregation. *Naval Research Logistics*, *60*(6), 479–498.
Rostami-Tabar, B., & Ziel, F. (2020). Anticipating special events in emergency department forecasting. *International Journal of Forecasting*.
Rothman, P. (1998). Forecasting asymmetric unemployment rates. *The Review of Economics and Statistics*, *80*(1), 164–168.
Rothschild, D. (2009). Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opinion Quarterly*, *73*(5), 895–916.
Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy*, *64*(3), 242–258.
Roulin, E., & Vannitsem, S. (2019). Post-processing of seasonal predictions – case studies using EUROSIP hindcast data base. *Nonlinear Processes in Geophysics*.
Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. F. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.
Rowe, G., & Wright, G. (1999). The delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375.
Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: The role of the delphi technique. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 125–144). Boston, MA: Springer US.
Royer, J. F. (1993). Review of recent advances in dynamical extended range forecasting for the extratropics. In J. Shukla (Ed.), *Prediction of interannual climate variations* (pp. 49–69). Berlin, Heidelberg: Springer.
Ruano, A. E., Crispim, E. M., Conceiçao, E. Z., & Lúcio, M. M. J. (2006). Prediction of building's temperature using neural networks models. *Energy and Buildings*, *38*(6), 682–694.
Rubaszek, M. (2020). Forecasting crude oil prices with DSGE models. *International Journal of Forecasting*.
Ruddiman, W. (2005). *Plows, plagues and petroleum: how humans took control of climate*. Princeton: Princeton University Press.
Rycroft, R. S. (1993). Microcomputer software of interest to forecasters in comparative review: An update. *International Journal of Forecasting*, *9*(4), 531–575.
Sa, J. (1987). Reservations forecasting in airline yield management. (Ph.D. thesis), Massachusetts Institute of Technology.
Sacheti, A., Gregory-Smith, I., & Paton, D. (2014). Uncertainty of outcome or strengths of teams: An economic analysis of attendance demand for international cricket. *Applied Economics*, *46*(17), 2034–2046.
Sah, S., Moore, D. A., & MacCoun, R. J. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes*, *121*(2), 246–255.
Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, the Netherlands: D. Reidel*, *81*.
Sakata, S., & White, H. (1998). High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica*, *66*(3), 529–568.
Sakia, R. M. (1992). The box-cox transformation technique: A review. *Journal of the Royal Statistical Society: Series D (the Statistician)*, *41*(2), 169–178.
Saksornchai, T., Lee, W.-J., Methaprayoon, K., Liao, J. R., & Ross, R. J. (2005). Improve the unit commitment scheduling by using the neural-network-based short-term load forecasting. *IEEE Transactions on Industry Applications*, *41*(1), 169–179.
Salinas, D., Bohlke-Schneider, M., Callot, L., Medico, R., & Gasthaus, J. (2019). High-dimensional multivariate forecasting with low-rank Gaussian copula processes. In *Advances in neural information processing systems* (pp. 6827–6837).
Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2019). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*.
Salway, R., & Wakefield, J. (2004). A common framework for ecological inference in epidemiology, political science and sociology. In *Ecological inference: new methodological strategies* (pp. 303–332). Cambridge University Press.
Sanders, N. R., & Manrodt, K. B. (2003). Forecasting software in practice: Use, satisfaction, and performance. *Interfaces*, *33*(5), 90–93.
Sanderson, J. (2012). Risk, uncertainty and governance in megaprojects: A critical discussion of alternative explanations. *International Journal of Project Management*, *30*(4), 432–443.
Santos, M. S., Abreu, P. H., Garca-Laencina, P. J., Simão, A., & Carvalho, A. (2015). A new cluster-based oversampling method for improving survival prediction of hepatocellular carcinoma patients. *Journal of Biomedical Informatics*, *58*, 49–59.
Sardinha-Lourenço, A., Andrade-Campos, A., Antunes, A., & Oliveira, M. S. (2018). Increased performance in the short-term water demand forecasting through the use of a parallel adaptive weighting strategy. *Journal of Hydrology*, *558*, 392–404.
Savin, S., & Terwiesch, C. (2005). Optimal product launch times in a duopoly: balancing life-cycle revenues with product cost. *Operations Research*, *53*, 26–47.
Scerri, M., De Goumoens, P., Fritsch, C., Van Melle, G., Stiefel, F., & So, A. (2006). The INTERMED questionnaire for predicting return to work after a multidisciplinary rehabilitation program for chronic low back pain. *Joint Bone Spine*, *73*(6), 736–741.
Schäfer, A. M., & Zimmermann, H. G. (2006). Recurrent neural networks are universal approximators. In *Artificial Neural Networks – ICANN 2006* (pp. 632–640). Springer Berlin Heidelberg.
Scharpf, A., Schneider, G., Nöh, A., & Clauset, A. (2014). Forecasting the risk of extreme massacres in Syria. *European Review of International Studies*, *1*(2), 50–68.
Schefzik, R., Thorarinsdottir, T. L., & Gneiting, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, *28*(4), 616–640.
Scheuerer, M., & Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, *143*(4), 1321–1334.
Scheuren, F. J., & Alvey, W. (2008). *Elections and exit polling*. Hoboken (New Yersey): John Wiley & Sons.
Schmertmann, C. P. (2003). A system of model fertility schedules with graphically intuitive parameters. *Demographic Research*, *9*, 81–110.
Schmertmann, C., Zagheni, E., Goldstein, J. R., & Myrskylä, M. (2014). BayesIan forecasting of cohort fertility. *Journal of the American Statistical Association*, *109*(506), 500–513.
Schnaars, S. P., & Topol, M. T. (1987). The use of multiple scenarios in sales forecasting: An empirical test. *International Journal of Forecasting*, *3*(3), 405–419.
Schoemaker, P. J. H. (1991). When and how to use scenario planning: A heuristic approach with illustration. *Journal of Forecasting*, *10*(6), 549–564.
Schoen, R. (1987). *Modeling multigroup populations*. Springer Science & Business Media.
Schönbucher, P. J. (2003). *Credit derivatives pricing models: models, pricing and implementation*. John Wiley & Sons.
Schubert, S., & Rickard, R. (2011). Using forecast value added analysis for data-driven forecasting improvement. IBF Best Practices Conference.
Schwanenberg, D., Fan, F. M., Naumann, S., Kuwajima, J. I., Montero, R. A., & Assis dos Reis, A. (2015). Short-term reservoir optimization for flood mitigation under meteorological and hydrological forecast uncertainty. *Water Resources Management*, *29*(5), 1635–1651.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.
Scott Armstrong, J. (2006). Should the forecasting process eliminate face-to-face meetings? *Foresight: The International Journal of Applied Forecasting*, *5*, 3–8.
Seaman, B. (2018). Considerations of a retail forecasting practitioner. *International Journal of Forecasting*, *34*(4), 822–829.
Seifert, D. (2003). *Collaborative planning, forecasting, and replenishment: How to create a supply chain advantage*. New York: AMACOM.

Semenoglou, A.-A., Spiliotis, E., Makridakis, S., & Assimakopoulos, V. (2021). Investigating the accuracy of cross-learning time series forecasting methods. *International Journal of Forecasting*, *37*(3), 1072–1084, URL https://www.sciencedirect.com/science/article/pii/S0169207020301850.

Semero, Y. K., Zhang, J., & Zheng, D. (2020). EMD–PSO–ANFIS-Based hybrid approach for short-term load forecasting in microgrids. *IET Generation, Transmission and Distribution*, *14*(3), 470–475.

Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, *38*(7), 608–625.

Serletis, A., & Rangel-Ruiz, R. (2004). Testing for common features in North American energy markets. *Energy Economics*, *26*(3), 401–414.

Setel, P., AbouZahr, C., Atuheire, E. B., Bratschi, M., Cercone, E., Chinganya, O., et al. (2020). Mortality surveillance during the COVID-19 pandemic. *Bulletin of the World Health Organization*, *98*(6), 374.

Setzler, H., Saydam, C., & Park, S. (2009). EMS call volume predictions: A comparative study. *Computers & Operations Research*, *36*(6), 1843–1851.

Shackleton, M. B., Taylor, S. J., & Yu, P. (2010). A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking & Finance*, *34*(11), 2678–2693.

Shah, I., & Lisi, F. (2020). Forecasting of electricity price through a functional prediction of sale and purchase curves. *Journal of Forecasting*, *39*(2), 242–259.

Shahriari, M., & Blumsack, S. (2018). The capacity value of optimal wind and solar portfolios. *Energy*, *148*, 992–1005.

Shale, E. A., Boylan, J. E., & Johnston, F. R. (2006). Forecasting for intermittent demand: the estimation of an unbiased average. *Journal of the Operational Research Society*, *57*(5), 588–592.

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, *109*(50), 20425–20430.

Shang, H. L., & Booth, H. (2020). Synergy in fertility forecasting: Improving forecast accuracy through model averaging. *Genus*, *76*.

Shang, H. L., Booth, H., & Hyndman, R. J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research*, *25*, 173–214.

Shang, H. L., & Haberman, S. (2020a). Forecasting age distribution of death counts: An application to annuity pricing. *Annals of Actuarial Science*, *14*(1), 150–169.

Shang, H. L., & Haberman, S. (2020b). Retiree mortality forecasting: A partial age-range or a full age-range model? *Risks*, *8*(3), 69.

Shang, H. L., & Hyndman, R. J. (2017). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, *26*(2), 330–343.

Shang, J., Ma, T., Xiao, C., & Sun, J. (2019). Pre-training of graph augmented transformers for medication recommendation. arXiv:1906.00346.

Shang, G., McKie, E. C., Ferguson, M. E., & Galbreth, M. R. (2020). Using transactions data to improve consumer returns forecasting. *Journal of Operations Management*, *66*(3), 326–348.

Shang, H. L., & Xu, R. (2021). Change point detection for COVID-19 excess deaths in Belgium. *Journal of Population Research*, *in press*.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425.

Shen, H. (2009). On modeling and forecasting time series of smooth curves. *Technometrics*, *51*(3), 227–238.

Shen, H., & Huang, J. Z. (2005). Analysis of call centre arrival data using singular value decomposition. *Applied Stochastic Models in Business and Industry*, *21*(3), 251–263.

Shen, H., & Huang, J. Z. (2008a). Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management. *The Annals of Applied Statistics*, *2*(2), 601–623.

Shen, H., & Huang, J. Z. (2008b). Interday forecasting and intraday updating of call center arrivals. *Manufacturing & Service Operations Management*, *10*(3), 391–410.

Shen, H., Huang, J. Z., & Lee, C. (2007). Forecasting and dynamic updating of uncertain arrival rates to a call center. In *2007 IEEE International Conference on Service Operations and Logistics, and Informatics* (pp. 1–6). IEEE.

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., et al. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution.. In *IJCAI* (pp. 3838–3844).

Sheng, C., Zhao, J., Leung, H., & Wang, W. (2013). Extended Kalman filter based echo state network for time series prediction using MapReduce framework. In *2013 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Networks* (pp. 175–180). IEEE.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika*, *81*, 115–131.

Shi, Q., Yin, J., Cai, J., Cichocki, A., Yokota, T., Chen, L., et al. (2020). Block Hankel tensor ARIMA for multiple short time series forecasting. In *AAAI* (pp. 5758–5766).

Shishkin, J., Young, A. H., & Musgrave, J. C. (1967). The x-11 variant of the census II method seasonal adjustment program. (15), Bureau of the Census, US Department of Commerce.

Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer.

Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The hadoop distributed file system. In *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)* (pp. 1–10). IEEE.

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation – a review on the statistical data driven approaches. *European Journal of Operational Research*, *213*(1), 1–14.

Sideratos, G., Ikonomopoulos, A., & Hatziargyriou, N. D. (2020). A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks. *Electric Power Systems Research*, *178*, Article 106025.

Silvapulle, P., & Moosa, I. A. (1999). The relationship between spot and futures prices: evidence from the crude oil market. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *19*(2), 175–193.

Simon, H., & Sebastian, K.-H. (1987). Diffusion and advertising: The german telephone campaign. *Management Science*, *33*(4), 451–466.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *13*(2), 238–241.

Sims, C. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48.

Sims, C. (2002). Solving linear rational expectations models. *Computational Economics*, *20*(1–2), 1–20.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*, 665–690.

Singer, P. W., & Friedman, A. (2014). *Cybersecurity: What Everyone Needs To Know*. OUP USA.

Singleton, C., Reade, J., & Brown, A. (2019). Going with your gut: The (in)accuracy of forecast revisions in a football score prediction game. *Journal of Behavioral and Experimental Economics*, Article 101502.

Sinnathamby, M. A., Whitaker, H., Coughlan, L., Bernal, J. L., Ramsay, M., & Andrews, N. (2020). All-cause excess mortality observed by age group and regions in the first wave of the COVID-19 pandemic in England. *Eurosurveillance*, *25*(28), Article 2001239.

Sisson, S. A., Fan, Y., & Beaumont, M. (2019). *Handbook of Approximate Bayesian Computation*. Chapman & Hall/CRC.

Smets, F., Warne, A., & Wouters, R. (2014). Professional forecasters and real-time forecasting with a DSGE model. *International Journal of Forecasting*, *30*(4), 981–995.

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian dsge approach. *American Economic Review*, *97*(3), 586–606.

Smith, D. R. (2002). Markov-Switching and stochastic volatility diffusion models of short-term interest rates. *Journal of Business & Economic Statistics*, *20*(2), 183–197.

Smith, M. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, *105*(492), 1467–1479.

Smith, J. C. (2011). The ins and outs of UK unemployment. *The Economic Journal*, *121*, 402–444.

Smith, M., & Khaled, M. (2012). Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association*, *107*(497), 290–303.

Smith, B., Leimkuhler, J., & Darrow, R. (1992). Yield management at American airlines. *Interfaces*, *22*(1), 8–31.

Smith, M. S., & Maneesoonthorn, W. (2018). Inversion copulas from nonlinear state space models with an application to inflation forecasting. *International Journal of Forecasting*, *34*(3), 389–407.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., et al. (2020). North atlantic climate far more predictable than models imply. *Nature*, *583*(7818), 796–800.

Smith, M. S., & Vahey, S. P. (2016). Asymmetric forecast densities for US macroeconomic variables from a Gaussian copula model of cross-sectional and serial dependence. *Journal of Business & Economic Statistics*, *34*(3), 416–434.

Smith, J., & Wallis, K. F. (2009a). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, *71*(3), 331–355.

Smith, J., & Wallis, K. F. (2009b). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Etatistics*, *71*(3), 331–355.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, *36*(1), 75–85.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, *43*(1), 1–28.

Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, *28*(2), 485–496.

Sobhani, M., Hong, T., & Martin, C. (2020). Temperature anomaly detection for electric load forecasting. *International Journal of Forecasting*, *36*(2), 324–333.

Sobotka, T., & Beaujouan, E. (2018). Late motherhood in low-fertility countries: Reproductive intentions, trends and consequences. In D. Stoop (Ed.), *Preventing Age Related Fertility Loss* (pp. 11–29). Cham: Springer International Publishing.

Sobri, S., Koohi-Kamali, S., & Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. *Energy Conversion & Management*, *156*, 459–497.

Soebiyanto, R. P., Adimi, F., & Kiang, R. K. (2010). Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One*, *5*(3), Article e9450.

Sohst, R. R., Tjaden, J., de Valk, H., & Melde, S. (2020). The future of migration to europe: A systematic review of the literature on migration scenarios and forecasts. Geneva: International Organization for Migration.

Sommer, B., Pinson, P., Messner, J. W., & Obst, D. (2020). Online distributed learning in wind power forecasting. *International Journal of Forecasting*.

Son, N., Yang, S., & Na, J. (2019). Hybrid forecasting model for short-term wind power prediction using modified long short-term memory. *Energies*, *12*(20), 3901.

Song, H., & Li, G. (2021). Editorial: Tourism forecasting competition in the time of COVID-19. *Annals of Tourism Research*, Article 103198.

Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the annals of tourism research curated collection on tourism demand forecasting. *Annals of Tourism Research*, *75*, 338–362.

Song, H., Witt, S. F., & Li, G. (2008). *The Advanced Econometrics of Tourism Demand*. Routledge.

Sopadjieva, E., Dholakia, U. M., & Benjamin, B. (2017). A study of 46,000 shoppers shows that omnichannel retailing works. *Harvard Business Review*, *Reprint H03D7A*.

Sorensen, D. (2020). Strategic IBP: Driving profitable growth in complex global organizations. *Foresight: The International Journal of Applied Forecasting*, *56*, 36–45.

Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., & Lendasse, A. (2007). Methodology for long-term prediction of time series. *Neurocomputing*, *70*(16–18), 2861–2869.

Sorjamaa, A., & Lendasse, A. (2006). Time series prediction using dirrec strategy. In M. Verleysen (Ed.), *ESANN, European Symposium on Artificial Neural Networks, European Symposium on Artificial Neural Networks* (pp. 143–148). Citeseer, European Symposium on Artificial Neural Networks.

Sornette, D. (2003). Critical market crashes. *Physics Reports*, *378*(1), 1–98.

Soule, D., Grushka-Cockayne, Y., & Merrick, J. R. W. (2020). A heuristic for combining correlated experts. *SSRN:3680229*.

Souza, R. C., Marcato, A. L. M., Dias, B. H., & Oliveira, F. L. C. (2012). Optimal operation of hydrothermal systems with hydrological scenario generation through bootstrap and periodic autoregressive models. *European Journal of Operational Research*, *222*(3), 606–615.

Soyer, R., & Tarimcilar, M. M. (2008). Modeling and analysis of call center arrival data: A Bayesian approach. *Management Science*, *54*(2), 266–278.

Spagat, M., Mack, A., Cooper, T., & Kreutz, J. (2009). Estimating war deaths: An arena of contestation. *The Journal of Conflict Resolution*, *53*(6), 934–950.

Sparkes, J. R., & McHugh, A. K. (1984). Awareness and use of forecasting techniques in british industry. *Journal of Forecasting*, *3*(1), 37–42.

Spencer, J. (1904). On the graduation of the rates of sickness and mortality presented by the experience of the manchester unity of oddfellows during the period 1893-97. *Journal of the Institute of Actuaries*, *38*(4), 334–343.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, *333*(6048), 1393–1400.

Spiliotis, E., Assimakopoulos, V., & Makridakis, S. (2020). Generalizing the theta method for automatic forecasting. *European Journal of Operational Research*, *284*(2), 550–558.

Spiliotis, E., Assimakopoulos, V., & Nikolopoulos, K. (2019). Forecasting with a hybrid method utilizing data smoothing, a variation of the theta method and shrinkage of seasonal factors. *International Journal of Production Economics*, *209*, 92–102.

Spiliotis, E., Kouloumos, A., Assimakopoulos, V., & Makridakis, S. (2020). Are forecasting competitions data representative of the reality? *International Journal of Forecasting*, *36*(1), 37–53.

Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2019). Improving the forecasting performance of temporal hierarchies. *PLoS One*, *14*(10), Article e0223422.

Spiliotis, E., Petropoulos, F., Kourentzes, N., & Assimakopoulos, V. (2020). Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. *Applied Energy*, *261*, Article 114339.

Spiliotis, E., Raptis, A., & Assimakopoulos, V. (2015). Off-the-shelf vs. Customized forecasting support systems. *Foresight: The International Journal of Applied Forecasting, Issue 43*, 42–48.

Spithourakis, G., Petropoulos, F., Babai, M. Z., Nikolopoulos, K., & Assimakopoulos, V. (2011). Improving the performance of popular supply chain forecasting techniques. *Supply Chain Forum, An International Journal*, *12*(4), 16–25.

Spithourakis, G., Petropoulos, F., Nikolopoulos, K., & Assimakopoulos, V. (2014). A systemic view of ADIDA framework. *IMA Journal of Management Mathematics*, *25*, 125–137.

Squire, P. (1988). Why the 1936 literary digest poll failed. *Public Opinion Quarterly*, *52*(1), 125–133.

Sridhar, S., & Govindarasu, M. (2014). Model-based attack detection and mitigation for automatic generation control. *IEEE Transactions on Smart Grid*, *5*(2), 580–591.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Stadlober, E., Hormann, S., & Pfeiler, B. (2018). Quality and performance of a PM10 daily forecasting model. *Atmospheric Enviroment*, *42*, 1098–1109.

Stanford NLP Group (2013). Code for deeply moving: Deep learning for sentiment analysis. https://nlp.stanford.edu/sentiment/code.html (Accessed on 05 September 2020).

Staszewska-Bystrova, A. (2011). Bootstrap prediction bands for forecast paths from vector autoregressive models. *Journal of Forecasting*, *30*(8), 721–735.

Steckley, S. G., Henderson, S. G., & Mehrotra, V. (2005). Performance measures for service systems with a random arrival rate. In *Proceedings of the Winter Simulation Conference, 2005.*. IEEE.

Steurer, J. (2011). The delphi method: an efficient procedure to generate knowledge. *Skeletal Radiology*, *40*(8), 959–961.

Stillwell, J., & Clarke, M. (Eds.), (2011). Population dynamics and projection methods. Springer, Dordrecht.

Stock, J. H., & Watson, M. W. (1998). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. (6607), National Bureau of Economic Research, Inc.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, *30*, 481–493.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics*, *32*(4), 1339–1342.
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *36*(2), 111–133.
Strähl, C., & Ziegel, J. (2017). Cross-calibration of probabilistic forecasts. *Electronic Journal of Statistics*, *11*(1), 608–639.
Strauch, R., Hallerberg, M., & Hagen, J. (2004). Budgetary forecasts in europe – the track record of stability and convergence programmes. *ECB Working Paper 307*.
Strijbosch, L. W. G., & Moors, J. J. A. (2005). The impact of unknown demand parameters on (r,s)-inventory control performance. *European Journal of Operational Research*, *162*(3), 805–815.
Su, Y.-K., & Wu, C.-C. (2014). A new range-based regime-switching dynamic conditional correlation model for minimum-variance hedging. *Journal of Mathematical Finance*, *04*(03), 207–219.
Sugeno, M. (1985). *Industrial Applications of Fuzzy Control*. Elsevier Science Inc.
Sun, S., Sun, Y., Wang, S., & Wei, Y. (2018). Interval decomposition ensemble approach for crude oil price forecasting. *Energy Economics*, *76*, 274–287.
Sun, J., Sun, Y., Zhang, X., & McCabe, B. (2021). Model averaging of integer-valued autoregressive model with covariates. https://Ssrn.Com.
Sundquist, E. T., & Keeling, R. F. (2009). The mauna loa carbon dioxide record: Lessons for long-term earth observations. *Geophysical Monograph Series*, *183*, 27–35.
Surowiecki, J. (2005). *The Wisdom of Crowds: Why the Many are Smarter than the Few* (New). Abacus.
Svensson, A., Holst, J., Lindquist, R., & Lindgren, G. (1996). Optimal prediction of catastrophes in autoregressive moving-average processes. *Journal of Time Series Analysis*, *17*(5), 511–531.
Svetunkov, I., & Boylan, J. E. (2020). State-space ARIMA for supply-chain forecasting. *International Journal of Productions Research*, *58*(3), 818–827.
Swanson, N. R., & Xiong, W. (2018). Big data analytics in economics: What have we learned so far, and where should we go from here? *Canadian Journal of Economics*, *51*(3), 695–746.
Sweeney, C., Bessa, R. J., Browell, J., & Pinson, P. (2019). The future of forecasting for renewable energy. *Wiley Interdisciplinary Reviews: Energy and Environment*.
Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*(1), 1–26.
Syntetos, A. A., Babai, M. Z., & Luo, S. (2015). Forecasting of compound erlang demand. *Journal of the Operational Research Society*, *66*(12), 2061–2074.
Syntetos, A. A., & Boylan, J. E. (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics*, *71*(1), 457–466.
Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, *21*(2), 303–314.
Syntetos, A. A., & Boylan, J. E. (2006). On the stock control performance of intermittent demand estimators. *International Journal of Production Economics*, *103*(1), 36–47.
Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495–503.
Syntetos, A. A., Kholidasari, I., & Naim, M. M. (2016). The effects of integrating management judgement into OUT levels: In or out of context? *European Journal of Operational Research*, *249*(3), 853–863.
Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., & Goodwin, P. (2009). The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, *118*(1), 72–81.
Syntetos, A. A., Zied Babai, M., & Gardner, E. S. (2015). Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. *Journal of Business Research*, *68*(8), 1746–1752.
Syring, N., & Martin, R. (2020). Gibbs posterior concentration rates under sub-exponential type losses. arXiv:2012.04505.
Szozda, N. (2010). Analogous forecasting of products with a short life cycle. *Decision Making in Manufacturing and Services*, *4*(1–2), 71–85.
Taillardat, M., Mestre, O., Zamo, M., & Naveau, P. (2016). Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Monthly Weather Review*, *144*(6), 2375–2393.
Talagala, T. (2015). Distributed lag nonlinear modelling approach to identify relationship between climatic factors and dengue incidence in colombo district, Sri Lanka. *Epidemiology, Biostatistics and Public Health*, *12*(4).
Talagala, T. S., Hyndman, R. J., & Athanasopoulos, G. (2018). Meta-learning how to forecast time series. (6/18), Monash University, Department of Econometrics and Business Statistics.
Talagala, P. D., Hyndman, R. J., Leigh, C., Mengersen, K., & Smith-Miles, K. (2019). A feature-based procedure for detecting technical outliers in water-quality data from in situ sensors. *Water Resources Research*, *55*(11), 8547–8568.
Talagala, P. D., Hyndman, R. J., & Smith-Miles, K. (2020). Anomaly detection in high dimensional data. *Journal of Computational and Graphical Statistics*, *in press*, 1–32.
Talagala, P. D., Hyndman, R. J., Smith-Miles, K., Kandanaarachchi, S., & Muñoz, M. A. (2020). Anomaly detection in streaming nonstationary temporal data. *Journal of Computational and Graphical Statistics*, *29*(1), 13–27.
Talavera-Llames, R. L., Pérez-Chacón, R., Martínez-Ballesteros, M., Troncoso, A., & Martínez-Álvarez, F. (2016). A nearest neighbours-based algorithm for big time series data forecasting. In *International Conference on Hybrid Artificial Intelligence Systems* (pp. 174–185). Springer.
Taleb, N. N. (2008). *The Black Swan: The Impact of the Highly Improbable* (New Edition). Penguin.
Taleb, N. N. (2020). *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*. STEM Academic Press.
Taleb, N. N., Bar-Yam, Y., & Cirillo, P. (2020). On single point forecasts for fat tailed variables. *International Journal of Forecasting*.
Tam Cho, W. K. (1998). Iff the assumption fits...: A comment on the king ecological inference solution. *Political Analysis*, *7*, 143–163.
Tan, B. K., Panagiotelis, A., & Athanasopoulos, G. (2019). Bayesian inference for the one-factor copula model. *Journal of Computational and Graphical Statistics*, *28*(1), 155–173.
Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction To Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc..
Tandberg, D., Easom, L. J., & Qualls, C. (1995). Time series forecasts of poison center call volume. *Journal of Toxicology: Clinical Toxicology*, *33*(1), 11–18.
Tanir, O., & Booth, R. J. (1999). Call center simulation in bell Canada. *2*, In *WSC'99. 1999 Winter Simulation Conference Proceedings.'simulation-a Bridge To the Future'(Cat. No. 99CH37038)* (pp. 1640–1647). IEEE.
Tarun, G., & Bryan, K. (2019). Factor momentum everywhere. *Journal of Portfolio Management*, *45*(3), 13–36.
Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, *16*(4), 437–450.
Tashman, L. J., & Leach, M. L. (1991). Automatic forecasting software: A survey and evaluation. *International Journal of Forecasting*, *7*(2), 209–230.
Tay, A. S., & Wallis, K. F. (2000). Density forecasting: a survey. *Journal of Forecasting*, *19*(4), 235–254.
Taylor, S. (1986a). *Modelling Financial Time Series*. Wiley.
Taylor, J. M. (1986b). The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association*, *81*(393), 114–118.
Taylor, J. W. (2003a). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, *19*(4), 715–725.
Taylor, J. W. (2003b). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, *54*(8), 799–805.
Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research*, *178*(1), 154–167.
Taylor, J. W. (2008). A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Science*, *54*(2), 253–265.

Taylor, J. W. (2010). Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles. *International Journal of Forecasting*, *26*(4), 627–646.

Taylor, J. W. (2012). Density forecasting of intraday call center arrivals using models based on exponential smoothing. *Management Science*, *58*(3), 534–549.

Taylor, J. W., & Bunn, D. W. (1999). A quantile regression approach to generating prediction intervals. *Management Science*, *45*(2), 131–295.

Taylor, J., & Jeon, J. (2018). Probabilistic forecasting of wave height for offshore wind turbine maintenance. *European Journal of Operational Research*, *267*(3).

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37–45.

Taylor, J. W., McSharry, P. E., & Buizza, R. (2009). Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, *24*(3), 775–782.

Taylor, M. P., & Peel, D. A. (2000). Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals. *Journal of International Money and Finance*, *19*(1), 33–53.

Taylor, J. W., & Snyder, R. D. (2012). Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing. *Omega*, *40*(6), 748–757.

Taylor, P. F., & Thomas, M. E. (1982). Short term forecasting: Horses for courses. *Journal of the Operational Research Society*, *33*(8), 685–694.

Tenti, P. (1996). Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*, *10*(6), 567–582.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, *89*(425), 208–218.

Teräsvirta, T., Tjostheim, D., & Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. OUP Oxford.

Teunter, R. H., & Duncan, L. (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, *60*(3), 321–329.

Teunter, R. H., Syntetos, A. A., & Zied Babai, M. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, *214*(3), 606–615.

Tewari, D. D. (1990). Energy-price impacts modelling in the agriculture sector. *Energy Economics*, *12*(2), 147–158.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin Books.

The Conference Board (2020). Global business cycle indicators. In *the Conference Board*. https://conference-board.org/data/bcicountry.cfm?cid=1 (Accessed on 2020-09-07).

The R. F. E. Working Group Report (2015). Risk in the front end of megaprojects. European Cooperation in Science and Technology.

Theocharis, Z., & Harvey, N. (2019). When does more mean worse? Accuracy of judgmental forecasting is nonlinearly related to length of data series. *Omega*, *87*, 10–19.

Theocharis, Z., Smith, L. A., & Harvey, N. (2018). The influence of graphical format on judgmental forecasting accuracy: Lines versus points. *Futures & Foresight Science*, *13*, Article e7.

Theodosiou, M. (2011). Disaggregation & aggregation of time series components: A hybrid forecasting approach using generalized regression neural networks and the theta method. *Neurocomputing*, *74*(6), 896–905.

Thomakos, D., & Nikolopoulos, K. (2012). Fathoming the theta method for a unit root process. *IMA Journal of Management Mathematics*, *25*(1), 105–124.

Thomakos, D. D., & Nikolopoulos, K. (2015). Forecasting multivariate time series with the theta method: Multivariate theta method. *Journal of Forecasting*, *34*(3), 220–229.

Thomé, A. M. T., Hollmann, R. L., & Scavarda do Carmo, L. F. R. R. (2014). Research synthesis in collaborative planning forecast and replenishment. *Industrial Management & Data Systems*, *114*(6), 949–965.

Thomé, A. M. T., Scavarda, L. F., Fernandez, N. S., & Scavarda, A. J. (2012). Sales and operations planning: A research synthesis. *International Journal of Production Economics*, *138*(1), 1–13.

Thomson, W., Jabbari, S., Taylor, A., Arlt, W., & Smith, D. (2019). Simultaneous parameter estimation and variable selection via the logit-normal continuous analogue of the spike-and-slab prior. *Journal of the Royal Society Interface*, *16*(150).

Thorarinsdottir, T. L., Scheuerer, M., & Heinz, C. (2016). Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *Journal of Computational and Graphical Statistics*, *25*(1), 105–122.

Thorarinsdottir, T. L., & Schuhen, N. (2018). Verification: assessment of calibration and accuracy. In *Statistical Postprocessing of Ensemble Forecasts* (pp. 155–186). Elsevier.

Tian, J., & Anderson, H. M. (2014). Forecast combinations under structural break uncertainty. *International Journal of Forecasting*, *30*(1), 161–175.

Tian, F., Yang, K., & Chen, L. (2017). Realized volatility forecasting of agricultural commodity futures using the HAR model with time-varying sparsity. *International Journal of Forecasting*, *33*(1), 132–152.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *58*, 267–288.

Timmermann, A. (2000). Moments of Markov switching models. *Journal of Econometrics*, *96*(1), 75–111.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *vol. 1*, *Handbook of Economic Forecasting* (pp. 135–196). Amsterdam: Elsevier.

Timmermann, A., & Zhu, Y. (2019). Comparing forecasting performance with panel data. *SSRN:3380755*.

Tiwari, A. K., Nasreen, S., Shahbaz, M., & Hammoudeh, S. (2020). Time-frequency causality and connectedness between international prices of energy, food, industry, agriculture and metals. *Energy Economics*, *85*, Article 104529.

Todini, E. (1991). Coupling real-time forecasting in the aswan dam reservoir management. In *Workshop on Monitoring, Forecasting and Simulation of River Basins for Agricultural Production*. Rome: Land and Water Development Division.

Todini, E. (1999). Using phase-space modelling for inferring forecasting uncertainty in non-linear stochastic decision schemes. *Journal of Hydroinformatics*, *1*(2), 75–82.

Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management*, *6*(2), 123–137.

Todini, E. (2016). Predictive uncertainty assessment and decision making. In V. P. Singh (Ed.), *Handbook of Applied Hydrology* (pp. 26.1–26.16). New York: McGraw Hill.

Todini, E. (2017). Flood forecasting and decision making in the new millennium. Where are we? *Water Resources Management*, *31*(10), 3111–3129.

Todini, E. (2018). Paradigmatic changes required in water resources management to benefit from probabilistic forecasts. *Water Security*, *3*, 9–17.

Toktay, L. B. (2003). Forecasting product returns. In V. D. R. Guide, Jr., & L. N. van Wassenhove (Eds.), *Business Aspects of Closed-Loop Supply Chains* (pp. 203–209). Pittsburgh: Carnegie Mellon University Press.

Toktay, L. B., Wein, L. M., & Zenios, S. A. (2000). Inventory management of remanufacturable products. *Management Science*, *46*(11), 1412–1426.

Tolman, H. L. (2008). A mosaic approach to wind wave modeling. *Ocean Modelling*, *25*(1–2), 35–47.

Tong, H. (1978). On a threshold model. In C. Chen (Ed.), *NATO ASI Series E: Applied Sc.*, *Pattern Recognition and Signal Processing* (pp. 575–586). Netherlands: Sijthoff & Noordhoff.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Clarendon Press.

Toth, Z., & Buizza, R. (2019). Weather forecasting: What sets the forecast skill horizon? In *Sub-Seasonal To Seasonal Prediction* (pp. 17–45). Elsevier.

Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, *158*, 1533–1543.

Tracy, M., Cerdá, M., & Keyes, K. M. (2018). Agent-based modeling in public health: current applications and future directions. *Annual Review of Public Health*, *39*, 77–94.

Tran, M.-N., Nott, D. J., & Kohn, R. (2017). Variational Bayes with intractable likelihood. *Journal of Computational and Graphical Statistics*, *26*(4), 873–882.

Tran, T., Phung, D., Luo, W., Harvey, R., Berk, M., & Venkatesh, S. (2013). An integrated framework for suicide risk prediction. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1410–1418).

Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, *66*(2), 299–307.

Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, *29*(2), 234–243.

Triguero, I., Peralta, D., Bacardit, J., García, S., & Herrera, F. (2015). MRPR: A MapReduce solution for prototype reduction in big data classification. *Neurocomputing*, *150*, 331–345.

Trivedi, P. K., & Zimmer, D. M. (2007). *Copula Modeling: An Introduction for Practitioners*. Now Publishers Inc.

Tsai, S.-B., Xue, Y., Zhang, J., Chen, Q., Liu, Y., Zhou, J., et al. (2017). Models for forecasting growth trends in renewable energy. *Renewable and Sustainable Energy Reviews*, *77*, 1169–1178.

Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, *81*(393), 132–141.

Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, *20*(3), 351–362.

Tsyplakov, A. (2013). Evaluation of probabilistic forecasts: proper scoring rules and moments. *SSRN:2236605*.

Tu, Y., Ball, M., & Jank, W. (2008). Estimating flight departure delay distributions – a statistical approach with long-term trend and short-term pattern. *Journal of the American Statistical Association*, *103*(481), 112–125.

Tuljapurkar, S., & Boe, C. (1999). Validation, probability-weighted priors, and information in stochastic forecasts. *International Journal of Forecasting*, *15*(3), 259–271.

Turkman, M. A. A., & Turkman, K. F. (1990). Optimal alarm systems for autoregressive processes: A Bayesian approach. *Computational Statistics & Data Analysis*, *10*(3), 307–314.

Turkmen, A. C., Wang, Y., & Januschowski, T. (2019). Intermittent demand forecasting with deep renewal processes. arXiv:1911.10416.

Turner, D. S. (1990). The role of judgement in macroeconomic forecasting. *Journal of Forecasting*, *9*(4), 315–345.

Turner, L., & Boulhol, H. (2011). Recent trends and structural breaks in the US and EU15 labour productivity growth. *Applied Economics*, *43*(30), 4769–4784.

Turner, R., & Zolin, R. (2012). Forecasting success on large projects: Developing reliable scales to predict multiple perspectives by multiple stakeholders over multiple time frames. *Project Management Journal*, *43*(5), 87–99.

Turnovsky, S. J., & Wachter, M. L. (1972). A test of the "expectations hypothesis" using directly observed wage and price expectations. *The Review of Economics and Statistics*, *54*(1), 47–54.

Twyman, M., Harvey, N., & Harries, C. (2008). Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication. *Judgment and Decision Making*, *3*(1), 111–120.

Tych, W., Pedregal, D. J., Young, P. C., & Davies, J. (2002). An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system. *International Journal of Forecasting*, *18*(4), 673–695.

Tziafetas, G. (1986). Estimation of the voter transition matrix. *Optimization*, *17*(2), 275–279.

Uematsu, H., Kunisawa, S., Sasaki, N., Ikai, H., & Imanaka, Y. (2014). Development of a risk-adjusted in-hospital mortality prediction model for community-acquired pneumonia: a retrospective analysis using a Japanese administrative database. *BMC Pulmonary Medicine*, *14*(1), 203.

Ugurlu, U., Oksuz, I., & Tas, O. (2018). Electricity price forecasting using recurrent neural networks. *Energies*, *11*(5), 1255.

Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, *145*(10), 1280–1297.

Unwin, A. (2019). Multivariate outliers and the O3 plot. *Journal of Computational and Graphical Statistics*, *28*(3), 635–643.

Vaks, A., Mason, A. J., Breitenbach, S. F. M., et al. (2019). Palaeoclimate evidence of vulnerable permafrost during times of low sea ice. *Nature*, *577*, 221–225.

Van de Ven, A., & Delbeco, A. L. (1971). Nominal versus interacting group processes for committee decision-making effectiveness. *Academy of Management Journal. Academy of Management*, *14*(2), 203–212.

Van den Broeke, M., De Baets, S., Vereecke, A., Baecke, P., & Vanderheyden, K. (2019). Judgmental forecast adjustments over different time horizons. *Omega*, *87*, 34–45.

van den Hengel, G., & Franses, P. H. (2020). Forecasting social conflicts in africa using an epidemic type aftershock sequence model. *Forecasting*, *2*(3), 284–308.

Van der Auweraer, S., & Boute, R. (2019). Forecasting spare part demand using service maintenance information. *International Journal of Production Economics*, *213*, 138–149.

van der Bles, A. M., van der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L., et al. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, *6*(5), Article 181870.

van der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, *117*(14), 7672–7683.

van der Laan, E., van Dalen, J., Rohrmoser, M., & Simpson, R. (2016). Demand forecasting and order planning for humanitarian logistics: An empirical assessment. *Journal of Operations Management*, *45*, 114–122.

van der Mark, L. B., van Wonderen, K. E., Mohrs, J., van Aalderen, W. M., ter Riet, G., & Bindels, P. J. (2014). Predicting asthma in preschool children at high risk presenting in primary care: development of a clinical asthma prediction score. *Primary Care Respiratory Journal*, *23*(1), 52–59.

Van Dijk, D., Franses, P. H., & Lucas, A. (1999). Testing for smooth transition nonlinearity in the presence of outliers. *Journal of Business & Economic Statistics*, *17*(2), 217–235.

Van Heerde, H. J., Leeflang, P. S., & Wittink, D. R. (2002). How promotions work: Scan* PRO-based evolutionary model building. *Schmalenbach Business Review*, *54*(3), 198–220.

van Ours, J. C. (2021). Common international trends in football stadium attendance. *PLoS One*, *16*(3), Article e0247761.

Van Reeth, D. (2019). Forecasting tour de France TV audiences: A multi-country analysis. *International Journal of Forecasting*, *35*(2), 810–821.

Van Schaeybroeck, B., & Vannitsem, S. (2018). Postprocessing of long-range forecasts. In S. Vannitsem, D. S. Wilks, & J. W. Messner (Eds.), *Statistical Postprocessing of Ensemble Forecasts* (pp. 267–290). Amsterdam, Netherlands: Elsevier.

van Wingerden, E., Basten, R. J. I., Dekker, R., & Rustenburg, W. D. (2014). More grip on inventory control through improved forecasting: A comparative study at three companies. *International Journal of Production Economics*, *157*, 220–237.

Vannitsem, S., Wilks, D. S., & Messner, J. (2018). *Statistical Postprocessing of Ensemble Forecasts*. Elsevier.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28.

Vaughan Williams, L., & Reade, J. (2016). Prediction markets, social media and information efficiency. *Kyklos*, *69*(3), 518–556.

Venkatramanan, S., Lewis, B., Chen, J., Higdon, D., Vullikanti, A., & Marathe, M. (2018). Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics*, *22*, 43–49.

Venter, J. H., De Jongh, P. J., & Griebenow, G. (2005). NIG-Garch models based on open, close, high and low prices. *South African Statistical Journal*, *39*(2), 79–101.

Verhulst, P. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathématique Et Physique*, *10*, 113–121.

Verhulst, P. F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux MÉMoires de L'AcadÉMie Royale Des Sciences Et Belles-Lettres de Bruxelles*, *18*, 14–54.

Vermue, M., Seger, C. R., & Sanfey, A. G. (2018). Group-based biases influence learning about individual trustworthiness. *Journal of Experimental Social Psychology*, *77*, 36–49.

Veronesi, P., & Yared, F. (1999). Short and long horizon term and inflation risk premia in the US term structure: Evidence from an integrated model for nominal and real bond prices under regime shifts. *CRSP Working Paper*, *508*.

Vestergaard, L. S., Nielsen, J., Richter, L., Schmid, D., Bustos, N., Braeye, T., et al. (2020). Excess all-cause mortality during the COVID-19 pandemic in europe–preliminary pooled estimates from the euromomo network, march to april 2020. *Eurosurveillance*, *25*(26), Article 2001214.

Vile, J. L., Gillard, J. W., Harper, P. R., & Knight, V. A. (2012). Predicting ambulance demand using singular spectrum analysis. *Journal of the Operational Research Society*, *63*(11), 1556–1565.

Villegas, M. A., & Pedregal, D. J. (2018). Supply chain decision support systems based on a novel hierarchical forecasting approach. *Decision Support Systems*, *114*, 29–36.

Vipul, & Jacob, J. (2007). Forecasting performance of extreme-value volatility estimators. *Journal of Futures Markets*, *27*(11), 1085–1105.

Vitart, F., Robertson, A. W., & Anderson, D. L. T. (2012). Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, *61*(2), 23.

Vlahogianni, E. I., Golias, J. C., & Karlaftis, M. G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews*, *24*(5), 533–557.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C (Emerging Technologies)*, *43*, 3–19.

Volterra, V. (1926a). Fluctuations in the abundance of a species considered mathematically. *Nature*, *118*, 558–560.

Volterra, V. (1926b). Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Memoria Della Reale Accademia Nazionale Dei Lincei*, *2*, 31–113.

Wakefield, J. (2004). Ecological inference for 2x2 tables (with discussion). *Journal of the Royal Statistical Society, Series A*, *167*(3), 385–445.

Wallentin, G., Kaziyeva, D., & Reibersdorfer-Adelsberger, E. (2020). COVID-19 intervention scenarios for a long-term disease management. *International Journal of Health Policy and Management*.

Wallström, P., & Segerstedt, A. (2010). Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics*, *128*(2), 625–636.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.

Wang, C., Jiang, B., Fan, J., Wang, F., & Liu, Q. (2014). A study of the dengue epidemic and meteorological factors in guangzhou, China, by using a zero-inflated Poisson regression model. *Asia Pacific Journal of Public Health*, *26*(1), 48–57.

Wang, X., Kang, Y., Hyndman, R. J., & Li, F. (2020). Distributed ARIMA models for ultra-long time series. arXiv:2007.09577.

Wang, X., Kang, Y., Petropoulos, F., & Li, F. (2021). The uncertainty estimation of feature-based forecast combinations. *Journal of the Operational Research Society*.

Wang, H., Li, B., & Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *71*(3), 671–683.

Wang, P., Liu, B., & Hong, T. (2016). Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting*, *32*(3), 585–597.

Wang, S. L., & McPhail, L. (2014). Impacts of energy shocks on US agricultural productivity growth and commodity prices—A structural VAR analysis. *Energy Economics*, *46*(C), 435–444.

Wang, C.-N., Nhieu, N.-L., Chung, Y.-C., & Pham, H.-T. (2021). Multi-objective optimization models for sustainable perishable intermodal multi-product networks with delivery time window. *Mathematics*, *9*(4), 379.

Wang, W., Pedrycz, W., & Liu, X. (2015). Time series long-term forecasting model based on information granules and fuzzy clustering. *Engineering Applications of Artificial Intelligence*, *41*, 17–24.

Wang, X., & Petropoulos, F. (2016). To select or to combine? The inventory performance of model and expert forecasts. *International Journal of Productions Research*, *54*(17), 5271–5282.

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980–991.

Wang, X., Smith-Miles, K., & Hyndman, R. J. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, *13*(3), 335–364.

Wang, X., Smith-Miles, K., & Hyndman, R. J. (2009). Rule induction for forecasting method selection: meta-learning the characteristics of univariate time series. *Neurocomputing*, *72*(10–12), 2581–2594.

Wang, Y., Smola, A., Maddix, D., Gasthaus, J., Foster, D., & Januschowski, T. (2019). Deep factors for forecasting. In *International Conference on Machine Learning* (pp. 6607–6617).

Wang, W., & Syntetos, A. A. (2011). Spare parts demand: Linking forecasting to equipment maintenance. *Transportation Research Part E: Logistics and Transportation Review*, *47*(6), 1194–1209.

Wang, J., Wang, Z., Li, X., & Zhou, H. (2019). Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*.

Wang, Z., Wang, W., Liu, C., Wang, Z., & Hou, Y. (2017). Probabilistic forecast for multiple wind farms based on regular vine copulas. *IEEE Transactions on Power Systems*, *33*(1), 578–589.

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random forest based hourly building energy prediction. *Energy and Buildings*, *171*, 11–25.

Wang, J., Yang, W., Du, P., & Niu, T. (2018). A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. *Energy Conversion and Management*, *163*, 134–150.

Wang, J., Yang, W., Du, P., & Niu, T. (2020). Outlier-robust hybrid electricity price forecasting model for electricity market management. *Journal of Cleaner Production*, *249*, Article 119318.

Wang, H., & Yeung, D.-Y. (2016). A survey on Bayesian deep learning. arXiv:1604.01662.

Warne, A., Coenen, G., & Christoffel, K. (2010). Forecasting with DSGE models. (1185), European Central Bank.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133.

Weatherford, L., Gentry, T., & Wilamowski, B. (2003). Neural network forecasting for airlines: A comparative analysis. *Journal of Revenue and Pricing Management*, *1*(4), 319–331.

Weaver, W. T. (1971). The delphi forecasting method. *The Phi Delta Kappan*, *52*(5), 267–271.

Webby, R., O'Connor, M., & Edmundson, B. (2005). Forecasting support systems for the incorporation of event information: An empirical investigation. *International Journal of Forecasting*, *21*(3), 411–423.

Wei, W., & Hansen, M. (2006). An aggregate demand model for air passenger traffic in the hub-and-spoke network. *Transportation Research Part A: Policy and Practice*, *40*(10), 841–851.

Wei, W., & Held, L. (2014). Calibration tests for count data. *Test*, *23*, 787–805.

Wei, N., Li, C., Peng, X., Zeng, F., & Lu, X. (2019). Conventional models and artificial intelligence-based models for energy consumption forecasting: A review. *Journal of Petroleum Science and Engineering*, *181*, Article 106187.

Wei Su, C., Wang, X.-Q., Tao, R., & Oana-Ramona, L. (2019). Do oil prices drive agricultural commodity prices? Further evidence in a global bio-energy context. *Energy*, *172*, 691–701.

Weinberg, J., Brown, L. D., & Stroud, J. R. (2007). BayesIan forecasting of an inhomogeneous Poisson process with applications to call center data. *Journal of the American Statistical Association*, *102*(480), 1185–1198.

Weiß, C. H., Homburg, A., Alwan, L. C., Frahm, G., & Göb, R. (2021). Efficient accounting for estimation uncertainty in coherent forecasting of count processes. *Journal of Applied Statistics*, 1–22.

Weiß, G. N., & Supper, H. (2013). Forecasting liquidity-adjusted intraday value-at-risk with vine copulas. *Journal of Banking & Finance*, *37*(9), 3334–3350.

Wen, R., Torkkola, K., Narayanaswamy, B., & Madeka, D. (2017). A multi-horizon quantile recurrent forecaster. arXiv:1711.11053.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*(4), 1030–1081.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, 1067–1084.

White, H. (2000). A reality check for data snooping. *Econometrica*, *68*(5), 1097–1126.

Whitt, W., & Zhang, X. (2019). Forecasting arrivals and occupancy levels in an emergency department. *Operations Research for Health Care*, *21*, 1–18.

Whittaker, J., Garside, S., & Lindveld, K. (1997). Tracking and predicting a network traffic process. *International Journal of Forecasting*, *13*(1), 51–61.

Wicke, L., Dhami, M. K., Önkal, D., & Belton, I. K. (2019). Using scenarios to forecast outcomes of a refugee crisis. *International Journal of Forecasting*.

Wickham, R. (1995). Evaluation of forecasting techniques for short-term demand of air transportation. (Ph.D. thesis), Massachusetts Institute of Technology.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, *114*(526), 804–819.

Wilde, J., Chen, W., & Lohmann, S. (2020). COVID-19 and the future of US fertility: what can we learn from google?. (WP-2020-034), Max Planck Institute for Demographic Research, Rostock, Germany.

Wilkd, D. S. (2005). *Science & Technology*, *Statistical Methods in the Atmospheric Sciences* (2nd). Elsevier.

Wilks, D. S. (2004). The minimum spanning tree histogram as verification tool for multidimensional ensemble forecasts. *Montly Weather Review*, *132*, 1329–1340.

Wilks, D. S. (2019). Indices of rank histogram flatness and their sampling properties. *Monthly Weather Review*, *147*(2), 763–769.

Willekens, F. (2018). Towards causal forecasting of international migration. *Vienna Yearbook of Population Research*, *16*, 199–218.

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, *20*(3), 375–387.

Williams, L. V., & Reade, J. J. (2016). Forecasting elections. *Journal of Forecasting*, *35*(4).

Wilms, I., Rombouts, J., & Croux, C. (2021). Multivariate volatility forecasts for stock market indices. *International Journal of Forecasting*, *37*(2), 484–499.

Wind, Y. (Ed.), (1981). New product forecasting: models and applications. Lexington, MA: Lexington Books.

Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association*, *67*(337), 187–191.

Winkler, R. L., Grushka-Cockayne, Y., Lichtendahl, K. C., & Jose, V. R. R. (2019). Probability forecasts and their combination: A research perspective. *Decision Analysis*, *16*(4), 239–260.

Winkler, R. L., Muñoz, J., Cervera, J. L., Bernardo, J. M., Blattenberger, G., Kadane, J. B., et al. (1996). Scoring rules and the evaluation of probabilities. *Test*, *5*(1), 1–60.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342.

Wiśniowski, A., Smith, P. W. F., Bijak, J., Raymer, J., & Forster, J. J. (2015). BayesIan population forecasting: Extending the lee-carter method. *Demography*, *52*(3), 1035–1059.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. Krishnajah (Ed.), *Multivariate Analysis* (pp. 391–420). New York: Academic Press.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, *18*(2), 107–126.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.

Wolters, M. H. (2015). Evaluating point and density forecasts of DSGE models. *Journal of Applied Econometrics*, *30*(1), 74–96.

Wong, B. K., Bodnovich, T. A., & Selvi, Y. (1995). A bibliography of neural network business applications research: 1988-september 1994. *Expert Systems*, *12*(3), 253–261.

Wong-Fupuy, C., & Haberman, S. (2004). Projecting mortality trends: recent developments in the United Kingdom and the United States. *North American Actuarial Journal*, *8*(2), 56–83.

Woodford, M. (2002). Imperfect common knowledge and the effects of monetary policy. In P. Aghion, R. Frydman, J. Stiglitz, & M. Woodford (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: in Honor of Edmund Phelps* (pp. 25–58). Princeton University Press.

Wright, G., & Goodwin, P. (1999). Future-focussed thinking: combining scenario planning with decision analysis. *Journal of Multi-Criteria Decision Analysis*, *8*(6), 311–321.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, *25*(4), 813–825.

Wright, M. J., & Stern, P. (2015). Forecasting new product trial with analogous series. *Journal of Business Research*, *68*(8), 1732–1738.

Wu, S., & Chen, R. (2007). Threshold variable determination and threshold variable driven switching autoregressive models. *Statistica Sinica*, *17*(1), 241–S38.

Wu, C. C., & Liang, S. S. (2011). The economic value of range-based covariance between stock and bond returns with dynamic copulas. *Journal of Empirical Finance*, *18*(4), 711–727.

Wu, W., Ma, X., Zeng, B., Wang, Y., & Cai, W. (2019). Forecasting short-term renewable energy consumption of China using a novel fractional nonlinear grey Bernoulli model. *Renewable Energy*, *140*, 70–87.

Wu, Y., Yu, W., Cui, Y., & Lu, C. (2020). Data integrity attacks against traffic modeling and forecasting in M2m communications. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)* (pp. 1–6). IEEE.

Xiao, Y., & Han, J. (2016). Forecasting new product diffusion with agent-based models. *Technological Forecasting and Social Change*, *105*, 167–178.

Xie, Y. (2000). Demography: Past, present, and future. *Journal of the American Statistical Association*, *95*(450), 670–673.

Xie, T., & Ding, J. (2020). Forecasting with multiple seasonality. arXiv:2008.12340.

Xie, J., & Hong, T. (2016). Gefcom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. *International Journal of Forecasting*, *32*(3), 1012–1016.

Xie, W., Yu, L., Xu, S., & Wang, S. (2006). A new method for crude oil price forecasting based on support vector machines. In *International Conference on Computational Science* (pp. 444–451). Springer.

Xiong, T., Li, C., Bao, Y., Hu, Z., & Zhang, L. (2015). A combination method for interval forecasting of agricultural commodity futures prices. *Knowledge-Based Systems*, *77*, 92–102.

Xu, W. (1999). Long range planning for call centers at fedex. *The Journal of Business Forecasting*, *18*(4), 7.

Xu, Y., Liu, H., & Long, Z. (2020). A distributed computing framework for wind speed big data forecasting on apache spark. *Sustainable Energy Technologies and Assessments*, *37*, Article 100582.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*(3), 645–678.

Yaffee, R. A., Nikolopoulos, K., Reilly, D. P., Crone, S. F., Wagoner, K. D., Douglas, R. J., et al. (2011). An experiment in epidemiological forecasting: A comparison of forecast accuracies among different methods of forecasting deer mouse population densities in montana. In *Federal Forecaster's Brown Bag Lunch*.

Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Reconciling solar forecasts: Sequential reconciliation. *Solar Energy*, *179*, 391–397.

Yan, X. S., & Zheng, L. (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies*, *30*(4), 1382–1423.

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., et al. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174.

Yang, D., & Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business*, *73*(3), 477–491.

Yassine, A., Shirehjini, A. N., & Shirmohammadi, S. (2015). Smart meters big data: Game theoretic model for fair data sharing in deregulated smart grids. *IEEE Access*, *3*, 2743–2754.

Yelland, P., Baz, Z. E., & Serafini, D. (2019). Forecasting at scale: The architecture of a modern retail forecasting system. *Foresight: The International Journal of Applied Forecasting*, *55*, 10–18.

Yue, M. (2017). An integrated anomaly detection method for load forecasting data under cyberattacks. In *2017 IEEE Power & Energy Society General Meeting* (pp. 1–5). IEEE.

Yue, M., Hong, T., & Wang, J. (2019). Descriptive analytics-based anomaly detection for cybersecure load forecasting. *IEEE Transactions on Smart Grid*, *10*(6), 5964–5974.

Yusupova, A., Pavlidis, N. G., & Pavlidis, E. G. (2019). Adaptive dynamic model averaging with an application to house price forecasting. arXiv:1912.04661.
Yusupova, A., Pavlidis, E., Paya, I., & Peel, D. (2020). UK housing price uncertainty index (HPU). UK Housing Observatory, Dept. of Economics, Lancaster University Management School.
Zadeh, L. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.
Zagdański, A. (2001). Prediction intervals for stationary time series using the sieve bootstrap method. *Demonstratio Mathematica*, *34*(2), 257–270.
Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., et al. (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM*, *59*(11), 56–65.
Zaidi, A., Harding, A., & Williamson, P. (Eds.), (2009). New frontiers in microsimulation modelling. Farnham: Ashgate.
Zailani, S., Jeyaraman, K., Vengadasan, G., & Premkumar, R. (2012). Sustainable supply chain management (SSCM) in Malaysia: A survey. *International Journal of Production Economics*, *140*(1), 330–340.
Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, *12*(3), 372–390.
Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*(5), 931–955.
Zang, H., Cheng, L., Ding, T., Cheung, K. W., Liang, Z., Wei, Z., et al. (2018). Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Generation, Transmission and Distribution*, *12*(20), 4557–4567.
Zarnowitz, V. (1985). Rational expectations and macroeconomic forecasts. *Journal of Business & Economic Statistics*, *3*(4), 293–311.
Zelterman, D. (1993). A semiparametric bootstrap technique for simulating extreme order statistics. *Journal of the American Statistical Association*, *88*(422), 477–485.
Zhang, S., Bauer, N., Yin, G., & Xie, X. (2020). Technology learning and diffusion at the global and local scales: A modeling exercise in the REMIND model. *Technological Forecasting and Social Change*, *151*, Article 119765.
Zhang, J. L., & Bryant, J. (2019). Combining multiple imperfect data sources for small area estimation: a Bayesian model of provincial fertility rates in cambodia. *Statistical Theory and Related Fields*, *3*(2), 178–185.
Zhang, J.-L., Chen, J., & Lee, C.-Y. (2008). Joint optimization on pricing, promotion and inventory control with stochastic demand. *International Journal of Production Economics*, *116*(2), 190–198.
Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62.
Zhang, X., & Hutchinson, J. (1994). Simple architectures on fast machines: practical issues in nonlinear time series prediction. In A. S. Weigend, & N. A. Gershenfeld (Eds.), *Time Series Prediction Forecasting the Future and Understanding the Past* (pp. 219–241). Addison-Wesley, Santa Fe Institute.
Zhang, & Ming (2008). *Artificial higher order neural networks for economics and business*. IGI Global.
Zhang, Y., & Nadarajah, S. (2018). A review of backtesting for value at risk. *Communications in Statistics. Theory and Methods*, *47*(15), 3616–3639.
Zhang, X., Peng, Y., Zhang, C., & Wang, B. (2015). Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *Journal of Hydrology*, *530*, 137–152.
Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2), 501–514.
Zhang, W., Qi, Y., Henrickson, K., Tang, J., & Wang, Y. (2017). Vehicle traffic delay prediction in ferry terminal based on Bayesian multiple models combination method. *Transportmetrica A: Transport Science*, *13*(5), 467–490.
Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. *Annals of Tourism Research*, *87*, Article 103149.
Zhang, Y., & Wang, J. (2018). A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without directaccess to off-site information. *IEEE Transactions on Power Systems*, *33*(5), 5714–5726.
Zhang, Y., Wang, J., & Wang, X. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, *32*, 255–270.
Zhang, J., Wei, Y.-M., Li, D., Tan, Z., & Zhou, J. (2018). Short term electricity load forecasting using a hybrid model. *Energy*, *158*, 774–781.
Zhang, L., Zhou, W.-D., Chang, P.-C., Yang, J.-W., & Li, F.-Z. (2013). Iterated time series prediction with multiple support vector regression models. *Neurocomputing*, *99*, 411–422.
Zhang, W., Zou, Y., Tang, J., Ash, J., & Wang, Y. (2016). Short-term prediction of vehicle waiting queue at ferry terminal based on machine learning method. *Journal of Marine Science and Technology*, *21*(4), 729–741.
Zhao, H.-X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592.
Zheng, J., Xu, C., Zhang, Z., & Li, X. (2017). Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network. In *2017 51st Annual Conference on Information Sciences and Systems (CISS)* (pp. 1–6). IEEE.
Zhou, Z., & Matteson, D. S. (2016). Predicting melbourne ambulance demand using kernel warping. *The Annals of Applied Statistics*, *10*(4), 1977–1996.
Zhou, C., & Viswanathan, S. (2011). Comparison of a new bootstrapping method with parametric approaches for safety stock determination in service parts inventory systems. *International Journal of Production Economics*, *133*(1), 481–485.
Zhou, L., Zhao, P., Wu, D., Cheng, C., & Huang, H. (2018). Time series model for forecasting the number of new admission inpatients. *BMC Medical Informatics and Decision Making*, *18*(1), 39.
Zhu, S., Dekker, R., van Jaarsveld, W., Renjie, R. W., & Koning, A. J. (2017). An improved method for forecasting spare parts demand using extreme value theory. *European Journal of Operational Research*, *261*(1), 169–181.
Ziegel, J. F., & Gneiting, T. (2014). Copula calibration. *Electronic Journal of Statistics*, *8*(2), 2619–2638.
Ziel, F., & Berk, K. (2019). Multivariate forecasting evaluation: On sensitive and strictly proper scoring rules. arXiv:1910.07325.
Ziel, F., & Steinert, R. (2016). Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Economics*, *59*, 435–454.
Ziel, F., & Steinert, R. (2018). Probabilistic mid-and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews*, *94*, 251–266.
Ziel, F., & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, *70*, 396–420.
Zipf, G. K. (2016). *Human Behavior and the Principle of Least Effort: An Introduction To Human Ecology*. Ravenio Books.
Žmuk, B., Dumičić, K., & Palić, I. (2018). Forecasting labour productivity in the European union member states: Is labour productivity changing as expected? *Interdisciplinary Description of Complex Systems*, *16*(3-B), 504–523.
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *67*, 301–320.
Zwijnenburg, J. (2015). Revisions of quarterly GDP in selected OECD countries. *OECD Statistics Briefing*, *July 2015 - No. 22*, 1–12.