

Luck versus Skill Over Time: Time Varying Performance in the Cross-Section of Mutual Fund Returns

Pouliot, William

License:

None: All rights reserved

Document Version

Early version, also known as pre-print

Citation for published version (Harvard):

Pouliot, W 2018 'Luck versus Skill Over Time: Time Varying Performance in the Cross-Section of Mutual Fund Returns'.

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Luck versus skill over time: Time Varying Performance in the Cross-Section of Mutual Fund Returns

MARCO G. ERCOLANI* WILLIAM POULIOT*[†]
JOANNE S. ERCOLANI*[‡]

January 2015

Abstract

We extend the analysis of Fama and French (2010, Journal of Finance) by allowing for time-variation in the parameters of Carhart's (1997) Four Factor Capital Asset Pricing Model. As such, we address the "thorny problems" highlighted by Fama and French (2010, p.1925). We use recently developed statistical procedures that test whether subsets of the parameters in time-series regression models vary over time. We also use fixed-design bootstraps that are valid even in the presence of this time variation. Our results reveal that time variation has significant effects on the cross section distribution of alpha estimates and on our simulations. As in Fama and French (2010), we find that actively managed funds generate net returns that are no better than those from a passive benchmark portfolio. However, this evidence for a lack of skill in active fund management is weaker when we allow for time varying parameters compared to when we keep parameters fixed over time.

Keywords and Phrases: Mutual Funds, Capital Asset Pricing Model; CUSUM test; Linear regression models; Stochastic processes; *U*-Statistics, Bootstrap.

JEL codes: C12, C22, C52.

*Department of Economics, University of Birmingham.

[†]Corresponding Author: Department of Economics, Birmingham Business School, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom, email: W.Pouliot@bham.ac.uk, tel. +44 789 8844 258.

[‡]We acknowledge the helpful comments of attendees at the Fourth International Conference of the Financial Engineering and Banking Society 2014 at the University of Surrey and at the First Conference on Recent Developments in Financial Econometrics and Applications of the Journal of Banking and Finance 2014 at Deakin University.

1 Introduction

There are three recurring themes in the literature on mutual fund performance: whether average risk adjusted abnormal fund performance is negative, positive or zero; whether abnormal performance can be identified ex-ante; and for how long it persists. Assessing fund performance is further complicated by the fact that there is more than one measure of performance: conditional or unconditional. Choice of measure has an important effect on the inference made. For example Elton *et al* (1992), using unconditional measures, find that they indicate poor average fund performance. Ferson and Schadt (1996), however, using conditional measures, conclude that funds performance is neutral. More recently, Fama and French (2010), hereafter FF, look at fund performance using unconditional measures and conclude that few funds produce benchmark-adjusted expected net returns sufficient to cover transaction costs and management expenses. When FF add back management expenses to net returns, they find some evidence of inferior and superior fund performance.

FF point out that capturing time variation in the regression slopes poses thorny problems for their analysis but leave answering this question to future research. This research addresses the “thorny problems” left unanswered by FF (p.1925) using novel tests developed by Olmo and Pouliot (2014). Using their testing procedure, we test for time-varying parameters in the Four Factor Capital Asset Pricing Model, hereafter 4F-CAPM; these tests also provide information on which parameter, slope or intercept, has changed over time. Information on the timing of these changes allows us to develop a bootstrap procedure that remains valid even in the presence of time variation. Using returns on actively managed US mutual funds, we document that there is much time variation in these parameters (8% of the funds analysed have some form of time variation in parameters of the 4F-CAPM), a result which contrasts sharply with research of Baras *et al.* (2010) who find time variation has little effect on their simulations. We show that time variation of estimates of parameters of the 4F-CAPM are important and this has significant effects on our parameter estimates. In particular, it leads to much larger $t(\alpha)$ estimates, the measure of mutual fund performance generally employed, than without time variation. This is important as it can greatly alter the conclusion about the existence of superior performing mutual funds. Our bootstrap simulations remain valid in the presence of time variations and show evidence for superior performing funds is weak, but not as weak as when bootstrap simulations are carried out without time variation.

The paper is structured as follows: Section 2 explains the methodology implemented by FF, Section 3 discusses the methodology implemented here; Section 4 discusses time varying mutual fund performance and the returns data used in the simulations; and Section 5 discusses a recently developed testing procedure. Section 6 discusses the bootstrap simulations devised and implemented here; and Section 7 concludes. The Appendix reports the distribution table needed for the three statistics developed in Section 5.

2 Fama and French (2010) methodology

FF approach the assessment of mutual fund performance from a perspective they call equilibrium accounting. An implication of this perspective is that if passive investors earn passive returns this would imply active investment must be a zero-sum game. They find the value-weighted portfolio

of active mutual funds that invest primarily in US equities is close to the market portfolio; and that the α of the 4F-CAPM, when estimated before expenses, is close to zero. This result concerning performance on the aggregate level implies that active mutual funds with positive true α must be offset by active funds with negative true α . Their goal then is to make inferences regarding the cross-section of true α for active funds and to determine whether this cross-section suggests a world where the true α is zero for all funds or whether a nonzero true α exists for some of these funds.

To separate funds with true α greater than zero from so-called ‘lucky’ funds, FF estimate the 4F-CAPM on long time-series of fund returns and compare the results to bootstrap simulations. In the bootstraps a counter-factual world is simulated where no inferior or superior funds exist, the only variation is that one would expect to see is due to the occurrence of lucky and unlucky funds. The counter-factual data is created by subtracting each fund’s estimated α in the 4F-CAPM. Then 10,000 bootstrap simulation runs are used to produce a distribution of t -statistics on the estimated α , henceforth $t(\alpha)$, the measure of fund performance in a world where the true α is zero. Implementation of their bootstrap procedures is to randomly sample monthly fund returns.

There is a difficulty, however, with this bootstrap sampling procedure. By randomly sampling months, FF lose any possibility of analysing the time-variation in the estimated parameters of the 4F-CAPM. Added to this was the problem that, at the time of writing, the literature on testing for multiple structural breaks in sub-sets of parameters was not well developed:

“Because we randomly sample months, we also lose any effects of variation through time in the regression slopes Capturing time variation in the regression slopes poses thorny problems, and we leave this potentially important issue for future research.” FF (p.1925)

3 Our methodology in context

Our contribution is to confront the two thorny problems highlighted by FF of how to test for multiple structural breaks and how to conduct bootstrap simulations without randomly sampling months. The first problem is to find systematic changes in fund performance by identifying structural breaks in the parameters of the 4F-CAPM. This is a difficult problem because we wish to use an automatic testing procedure capable of identifying more than one break and breaks in subsets of the intercept α or slope β parameters. We are particularly interested in allowing for changes in the α performance parameters. To this end, we use a testing procedure recently developed by Olmo and Pouliot (2014).

The tests in Olmo and Pouliot (2014) are developments on traditional CUSUM tests for structural breaks. Viewing time variation of the parameters of the 4F-CAPM as structural breaks is a non-parametric means of modeling time-variation in fund performance. Traditional CUSUM tests, like other tests for structural change that are popular in the statistical/econometrics literature, are not designed to distinguish changes in intercept from changes in slope parameters. Devising tests that are informative on the nature of the break in parameters allows us to incorporate time-variation into the bootstrapping procedure developed here. In doing so, we are better able to address the issue of time-variation than previous attempts. CUSUM tests have been

around for many years and a large number of such tests are now widely available. For example, Kuan and Hornick (1995) develop generalized fluctuation tests, Andrews (1993) and Andrews and Ploberger (1994) construct a class of tests based on Wald, Lagrange Multiplier and Likelihood Ratio statistics. More recent contributions have been made by Bai and Perron (1998), Altissimo and Corradi (2003) and Kristensen (2012). These tests, however, are not devised to distinguish changes in intercept from slope in regression models and as such are ill suited for our purposes.

Kapar and Pouliot (2013) and Olmo and Pouliot (2014) detail some practical difficulties associated with implementation of the above mentioned tests and document how these difficulties negatively affect their power. For example, many of these tests require trimming which impairs their power and hinders consistent estimation of the date of parameter change. Simulations studies also carried out by Kapar and Pouliot (2013) and Olmo and Pouliot (2014) confirm these tests exhibit little power to detect changes in the α intercept parameter of linear regression models. Unlike existing CUSUM tests, the Olmo and Pouliot (2014) testing procedure is able to disentangle breaks in intercept from breaks in the slope parameters of regression models. Their test statistic is constructed from a bivariate U -statistic type process that can accommodate weight functions. Weight functions allow significant improvements to the power of structural break tests against changes that occur early and later on in the sample period and simplify the requirements necessary to establish some of the functional limit theorems establish here regarding these statistics. The asymptotic theory is based on functionals of a Brownian bridge and therefore critical values used in our tests of change in intercept or slope can be easily tabulated.

The second problem is how to carry out the bootstrap simulations in a situation where we cannot randomly sample observations across time. FF can bootstrap by randomly sampling monthly observations because their 4F-CAPM estimates does not take into account time variation of the parameters in the model and, hence, the time-sequence does not matter. We cannot bootstrap by randomly sampling monthly observations because we do find that some parameters of the 4F-CAPM change over time. To allow for this variation the time-sequence of the data cannot be ignored. We therefore carry out ‘fixed design’ bootstraps where the simulated counter-factual fund returns are constructed by using in-sequence fitted 4F-CAPM predictions and adding randomly selected out-of sequence residuals. We assume that the regression residuals respect the classical assumptions of white noise. In all other senses our bootstrap simulations replicate the approach of FF by using 10,000 replications and by generating a counter-factual distribution where any systematic performance differences are eliminated by subtracting each fund’s estimated α in the 4F-CAPM. Like FF, we also base our testing on the distribution of t-statistics for α s, rather than the distribution of the α parameters themselves.

4 Time varying mutual fund performance

The tests developed in Section 5 indicate that 8% of the parameters, in the 4F-CAPMs estimated on the 5,785 time-series on mutual fund returns, exhibited time variation. 2% of these parameter changes are due to time variation in the intercept only, 5% due to slope only changes and a further 1% due to both slope and intercept changes. When these tests were applied to detect additional time variation in these parameters, it was found that a further 1% of the mutual funds displayed two changes in intercept only or slope only. Applying this testing procedure a third time indicated no further time variation in these parameters.

Early research into fund performance did not allow for time variation (*cf.* Jensen (1968)). By the 1980s this began to change. Admati and Ross (1985), assume fund managers maximize expected utility where utility takes a Constant Absolute Risk Aversion (CARA) form and all random variables are normally distributed. Under these assumptions, portfolio weights are linear functions of these random variables and the portfolio market risk β is also a linear function of these variables. Empirical research has to some extent found time-varying β in the 4F-CAPM. Studies examining the historical stability of the β s include Blume (1971), Levy (1971), Sharpe and Cooper (1972), and Black, Jensen & Scholes (1972). Brendt (1990, p.35) summarizes these studies as finding that the β s in these models are relatively stable:

“Quite frequently, monthly data are employed that are based on returns from the New York Stock Exchange. Econometrics studies based on such data have found that in many studies, β s [in the 4F-CAPM] have tended to be relatively stable over a five-year time span.”

He lists a couple of reasons for this variation: conditions in an industry may change abruptly causing risks to change. Such was the case for Oil company stocks, they had a market risk β below unity before the 1973 Oil price shock. Since then, however, they have typically been above one. A similar result occurred in 1978 when the US deregulated the airline industry, market risk β s for most major US airline companies rose.

More recent research on evaluating fund performance has developed a conditional version of the 4F-CAPM. For example Ferson and Schadt (1996) advocate use of conditional performance measures when evaluating fund performance and to do so they extend a multiple-factor asset pricing model to accommodate time-varying risk parameters. Their model, a reduced form, expresses slope parameters, only, as a linear function of observed variables whilst the intercept remains time invariant. Interestingly, their assumption of linearity is made for illustrative purposes only. According to Ferson and Schadt (1996, p.429),

“They use a linear specification to illustrate the conditional approach, the correct specification of this relationship is left as an empirical issue.”

A second paper, Baras *et al.* (2010), estimates conditional four factor models that allow slope parameters in the 4F-CAPM to be linearly related to a set of conditioning variables. These variables consist of yields on different securities. They, however, assume no variation in the intercept parameter, the parameter used to assess fund performance. After estimating both conditional and unconditional versions of the CAPM, Baras *et al.* (2010 p.195-196) conclude that

“Introducing time-varying market betas provides similar results. In further tests shown in the Internet Appendix, we find that using the unconditional or conditional version of the four-factor model has no material impact on our main results. For brevity ... we present only results from the unconditional four-factor model.”

This result seems at odds with the above statement of Brendt (1990) and Ferson and Schadt (1996). Tests developed in Section 5, when applied to fund returns, indicate a significant proportion of funds have changes in at least one of the parameters of the 4F-CAPM. Results produced in

Section 6 detail the impact this has on the bootstrap simulation. The second paper, Cuthbertson *et al.* (2008), estimates conditional β and conditional α -and- β versions of the 4F-CAPM. The former model allows intercept as well as slope to be linearly related to predetermined variables, while the latter model allows only the slope parameters to depend linearly on predetermined variables.

The specifications of these models are, however, parametric: both intercept and slope parameters are linearly related to predetermined variables. Linear specification can only be justified under very simple assumptions: investors maximize a CARA expected utility function defined over normally distributed random variables (cf. Admati and Ross (1985)). Linearity may not hold or, if it does, it may do so for a different set of conditioning variables than those used in these papers. As a result of these issues, it seems better to avoid a linear specification. For the purposes of this paper, we remain agnostic on the precise nature of this relationship and opt for a non-parametric specification. Allowing for time variation without imposing restrictions on this relation is an added novelty of our approach and is consistent with Ferson and Schadt (1996).

4.1 The data

Following FF, the mutual funds used here must invest primarily in US common stocks. To focus on actively managed funds, as in FF, we include funds in our analysis only if they fall into one of three categories: aggressive, growth and value. Even though FF use a somewhat different process to categorize funds than ours, the funds FF used fall into one of these three categories listed above. Index funds are excluded, as they are in FF, because they do not advertise generating returns superior to benchmark portfolios. With these restrictions noted, the mutual fund data used here consists of time-series on adjusted-closed prices for 6,178 US equity funds collected from January 1984 to October 2014. FF exclude from their analysis funds with less than 8 months of data a practice we follow here as well. After these deletions, the final number of funds totalled 5,785.

There are two important issues regarding the data that need to be addressed: survivor bias and incubation bias. Since the data used here consists of surviving funds only, it does not adjust for survivor bias; the use of $t(\alpha)$ in our analysis helps mitigate, but does not eliminate this problem (cf. Brown *et al.* 1992). Use of $t(\alpha)$ only partially offsets this bias, its impact on our analysis will be to skew the cross-section distribution of $t(\alpha)$ towards larger values. The results reported in Table II of Section 6, on the cross-section distribution of actual $t(\alpha)$, does not show this pattern.

FF raise the issue of incubation bias. Incubation bias occurs because funds include their pre-release returns in the mutual fund database only if these returns turn out to be positive. To lessen its effect on their results, FF exclude funds that have not reached 5 million 2006 US dollars in assets under management. To analyse the effects of incubation bias on their simulations, they use NASDAQ ticker symbol start dates to replicate their tests on this data and then compare them to results produced using CRSP start dates for new funds. They report that switching to ticker start dates has only trivial effects on their results and conclude incubation bias is probably unimportant for their results (cf. FF p.1924). As there is little evidence suggesting that by limiting their analysis to funds that have reached \$5 million assets affects their results, there seems no obvious reason to exclude funds with less than \$5 million from our bootstrap simulations. Support for our position can be found in two recent publications, Baras *et al.* (2010) and Cuthbertson *et al.* (2008), in which neither mention this issue and therefore make no special adjustment for it.

4.2 Regression framework

The benchmark model that will be used to evaluate fund performance is Carhart's (1997) 4F-CAPM. We will, however, allow all parameters of this model to change over time. This is all represented in the following regression model:

$$R_{it} - R_{ft} = \alpha_i(t) + \beta_{1i}(t)(R_{Mt} - R_{ft}) + \beta_{2i}(t)SMB_t + \beta_{3i}(t)HML_t + \beta_{4i}(t)MOM_t + \varepsilon_{it}. \quad (1)$$

Following the notation used by FF, we define R_{it} as the return on fund i for month t , R_{ft} as the risk-free rate and R_{Mt} as the market return (the return on the value-weight portfolio of NYSE, Amex, and NASDAQ stocks). SMB_t and HML_t are the size returns and value-growth returns of Fama and French (1993), while MOM_t is the momentum return of Carhart (1997). Times-series on all four factors can be obtained from Kenneth French's web site. The parameter $\alpha_i(t)$ is the average return left unexplained by the benchmark model. The $\alpha_i(t)$, along with the β (this is a 4×1 vector of parameters) are now time varying.

The more interesting aspect of equation (1), at least from an investment perspective, is that the β on the returns describes a diversified portfolio of passive benchmarks that replicates exposure to common factors in returns. This implies that α measures the average return provided by a fund in excess of the return on a comparable passive portfolio. Justification for interpreting a positive α as superior fund performance lies in Theorem 5 of Dybvig and Roth (1985). When $\alpha > 0$, or in Dybvig and Roth's terminology, a portfolio plots above the CAPM equation, a mean-variance superior portfolio can be constructed that is superior to the benchmark portfolio.

What can be concluded from equation (1)? This depends on whether returns are net of costs or gross. Any test for superior performance depends on whether managers have skill that causes expected returns to differ from those of the passive benchmark. For this, according to FF (2010, p.1921)

"one would like fund returns measured before all non-return revenues. This would put funds on the same pure return basis as the benchmark portfolio and the regression in (1) and allow for a proper test of manager's skill."

The returns calculated for the purpose of this analysis are gross which allows us to test whether managers have skill to generate returns in excess of returns on a passive benchmark portfolio that is represented in the equation (1). For FF, returns are net of trading costs and as such they are restricted to testing whether managers have sufficient skill to cover trading costs not whether they have skill to better the returns on passive benchmark portfolios.

To estimate the β based on time-series data of individual mutual fund returns, it must be assumed that, for a particular mutual fund, the β is stable over time. When monthly data are used, as they are here, it has been noted that the β is stable over a five-year time span. Since our data have a span of 30 years, there is scope for many structural breaks. To address this issue, we design some novel tests specifically for this purpose. More specifically, we construct tests for simultaneous as well as for joint changes in the parameters of linear regression models. These tests have the novelty of allowing one to distinguish a change in β from a change in α . Distinguishing a change in β from a change in α is important for the coming analysis because α is the measure employed here to evaluate fund performance. Hence it is important to determine when these parameters change and to incorporate this additional variation into the analysis. Our tests were applied three times to detect more than one change: we find no changes beyond two in any one of the α or β parameters.

Table I: 4F-CAPM estimates for Equal Weight Portfolio of Gross Returns

Variable	Our results				
	$12 * \hat{\alpha}$	$R_{Mt} - R_{ft}$	SMB_t	HML_t	MOM_t
<i>Coef</i>	-0.50	0.95	0.17	0.06	-0.02
<i>t-stat</i> [†]	-0.42	52.63	6.37	1.87	-1.05
<i>t-stat</i> [‡]		-2.77			

Variable	Fama and French (2010, Table II, EW returns)				
	$12 * \hat{\alpha}$	$R_{Mt} - R_{ft}$	SMB_t	HML_t	MOM_t
<i>Coef</i>	-0.39	0.98	0.18	-0.00	-0.00
<i>t-stat</i> [†]	-0.90	87.22 [§]	16.01	-0.25	-0.14
<i>t-stat</i> [‡]		-1.78			

[†] Standard *t*-statistics for coefficient being equal to zero.

[‡] Test of whether parameter $\beta_1 = 1$ on the variable $R_{Mt} - R_{ft}$.

[§] This value is not reported in Fama and French (2010, Table II).

4.3 Regression results for the equal weight portfolio

Table I reports parameter estimates for equation (1) using gross returns of the equal weight (EW) portfolio as the dependent variable. The equal weight portfolio weights equally individual monthly fund returns by averaging across funds' gross monthly returns over the period 1984 to 2014. EW fund returns are informative on whether funds on average produce returns different from those implied by their exposures to common factors in returns. Table I also lists results from FF (*cf.* Table II) for the same four-factor model. The results produced here bear a strong resemblance to those of FF. Whilst our estimate of α is larger in absolute value than the α estimate of FF, it is statistically insignificant as it is in FF. The difference in value is not unusual given that FF use returns net of trading costs while we use gross returns. Our estimate of market risk is similar to that produced by FF (0.95 compared to 0.98 respectively) and FF find marginal support for the market risk parameter being equal to one (*p*-value of 0.08), whereas our test rejects this hypothesis. Our parameter estimates on the remaining three factors are similar in magnitude to those reported by FF.

We interpret Table I in same way FF as interpret it. They find that, on average, there is little evidence that active mutual funds produce gross returns (gross returns for FF deduct trading costs) above or below that of passive benchmarks. We also find, on average, no evidence to suggest active mutual funds produce gross returns better than passive benchmarks. We argue, along the lines of FF, that this result indicates how active fund managers who are able to outperform the benchmark portfolio are more than offset by inferior managers who under-perform the benchmark. This similarity between conclusions is reassuring.

5 A novel test for structural breaks in regressions

To test for structural breaks in the 4F-CAPM we use the test of Olmo and Pouliot (2014), henceforth OP. This test has some superior properties over other break tests, that are useful to us here. Firstly, within a linear regression model, such as the one in equation (1), it can distinguish between a change in intercept and a change in slope. This is because the test statistic is constructed from two other statistics in which one is only capable of detecting a change intercept and the other is only capable of detecting a change in slope. Secondly, the test has power against parameter changes that occur early or late in the sample period. In this regard, the test complements and improves upon the approaches of Andrews (1993) and Andrews and Ploberger (1994), which do not have this property. Hence, this section provides sufficient detail for readers to understand why the test is useful and how it can be implemented. For those interested in the technical details, please refer to the original OP paper.

OP construct their test by assuming a random sample $\{(Y_t, \mathbf{X}_t)\}_{t=1}^T$ that satisfies equation (2) with an unknown break parameter at time t^* :

$$Y_t = \begin{cases} \alpha^{(1)} + \mathbf{X}_t \boldsymbol{\beta}^{(1)} + \sigma \varepsilon_t, & 1 \leq t \leq t^*, \\ \alpha^{(2)} + \mathbf{X}_t \boldsymbol{\beta}^{(2)} + \sigma \varepsilon_t, & t^* < t \leq T, \end{cases} \quad (2)$$

where the ε_t s are independent and identically distributed (iid) random variables satisfying the following moment conditions:

$$\mathbb{E}\varepsilon_t = 0, \mathbb{E}\varepsilon_t^2 = 1 \text{ and } \mathbb{E}|\varepsilon_t|^4 < \infty, \quad t = 1, \dots, T. \quad (3)$$

Note that equation (1) can be expressed in this way. The terms $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ are $K \times 1$ parameter vectors, \mathbf{X}_t is a $1 \times K$ vector of explanatory variables. It is assumed that all components of \mathbf{X}_t and the dependent variable Y_t are stationary¹. This assumption is required to establish weak convergence results regarding the processes considered here.

The issue of detecting parameter instability can be represented by the hypothesis

$$H_0 : t^* \geq T$$

versus the alternative hypothesis of at-most-one change in intercept or slope

$$H_1 : 1 \leq t^* < T.$$

The null hypothesis does not rule out structural breaks occurring at some point in time, but it does rule them out for the period covered by the sample of data. So if a break occurs, it happens beyond the sample. Under the alternative hypothesis of a break in either intercept or slope parameters, we assume that at least $\alpha^{(1)} \neq \alpha^{(2)}$ or $\boldsymbol{\beta}^{(1)} \neq \boldsymbol{\beta}^{(2)}$.

As in the original OP paper, discussion of the test is now divided into two cases: i) when parameters in (2) are known and ii) when parameters are unknown and must be estimated. Of course, the more realistic setting is the latter, however, the asymptotic distributions of our test

¹We can allow for non-stationary Y_t and \mathbf{X}_t as long as (2) represents a cointegrating relationship among the non-stationary variables.

statistics are easier to establish under assumption i). Nevertheless, it can be shown that the statistics derived under assumption i) are asymptotically equivalent to the statistics when under assumption ii). This equivalence permits us to apply any results established under assumption i) to the tests statistics while under assumption ii).

5.1 Parameters Known

To illustrate the roots of their proposed statistic, OP construct a process that has its origins on a U -statistic. They define process $M_T(\tau)$ to be a function of τ where $\tau \in (0, 1)$, such that

$$M_T(\tau) := T^{-1/2} \left\{ \sum_{t=1}^{[(T+1)\tau]} (Y_t - \alpha^{(1)} - \mathbf{X}_t \boldsymbol{\beta}^{(1)})^2 - \tau \sum_{t=1}^T (Y_t - \alpha^{(1)} - \mathbf{X}_t \boldsymbol{\beta}^{(1)})^2 \right\}. \quad (4)$$

This process can be interpreted as comparing the variance before a change in parameters to the variance after a change in parameters. Interest centres on how large this process is across the τ range, with a large value suggesting that the variance has changed. A suitable functional that captures this is the supremum, leading to the test statistic

$$\sup_{0 < \tau < 1} |M_T(\tau)|. \quad (5)$$

A useful by-product of this statistic is that it also indirectly yields an estimator of the break point τ^* . Gombay et al. (1996) show that this statistic converges to the supremum of a Brownian bridge. As it is, this statistic cannot distinguish between rejections in the intercept or the slope. Therefore OP construct two auxiliary U -statistic type processes defined as

$$M_T^{(A)}(\tau) := T^{-1/2} \left\{ \sum_{t=1}^{[(T+1)\tau]} (Y_t - \alpha - \mathbf{X}_t \boldsymbol{\beta}^{(1)})^2 - \tau \sum_{t=1}^T (Y_t - \alpha - \mathbf{X}_t \boldsymbol{\beta}^{(1)})^2 \right\} \quad (6)$$

$$M_T^{(B)}(\tau) := T^{-1/2} \left\{ \sum_{t=1}^{[(T+1)\tau]} (Y_t - \mathbf{X}_t \boldsymbol{\beta}) - \tau \sum_{t=1}^T (Y_t - \mathbf{X}_t \boldsymbol{\beta}) \right\}, \quad (7)$$

where in $M_T^{(A)}(\tau)$ set

$$\alpha = \begin{cases} \alpha^{(1)}, & t \leq t^* \\ \alpha^{(2)}, & t > t^* \end{cases}$$

and in $M_T^{(B)}(\tau)$ set

$$\boldsymbol{\beta} = \begin{cases} \boldsymbol{\beta}^{(1)}, & t \leq t^* \\ \boldsymbol{\beta}^{(2)}, & t > t^*. \end{cases}$$

The first statistic is sensitive to a one-time change in the $\boldsymbol{\beta}^{(1)}$ that is robust to a change in the intercept, should it occur. The second is sensitive to a one-time change in the intercept that is

robust to a one-time change in slope, if one should occur. In order to maximize the power of this test to detect a break, OP construct the test statistics

$$\sup_{0 < \tau < 1} \frac{|M_T^{(i)}(\tau)|}{q(\tau)}, \quad (8)$$

for $i = A, B$, in which $q(\tau)$ is a weight function devised to improve the power of the tests for detecting changes in parameters that occur in sample. Other similar structural break tests have power only in a compact interval within $(0, 1)$ but the interest here is to develop weight functions that can be used to improve the power of the test over the whole $(0, 1)$ range, in particular close to the points 0 and 1 such that they are sensitive to change points that are early and late in the evaluation period. In order to do this, $q(\tau)$ must satisfy the following two assumptions:

A.1: The function $q(\cdot)$ defined on $(0, 1)$ is such that $\inf_{\delta \leq \tau \leq 1-\delta} q(\tau) > 0$ for all $\tau \in (0, 1)$ and $\delta \in (0, 1/2)$.

A.2: $I(q, c) = \int_0^1 \frac{1}{\tau(1-\tau)} \exp^{-\frac{cq^2(\tau)}{(\tau(1-\tau))}} d\tau < \infty$ for some constant $c > 0$.

One family of weight functions that has received some attention, see Gombay et al. (1996), depends on a tuning parameter ν , and is given by

$$q(\tau) = q(\tau; \nu) := \{(\tau(1-\tau))^\nu; 0 \leq \nu < 1/2\}. \quad (9)$$

This class of functions satisfies A.1 and A.2 for all $c > 0$, and is sensitive to a change in parameters of linear regression models that occurs both early and late in the sample.

OP go on to derive the asymptotic distribution of the processes in (8). They show that, under H_0 , if we let the process (2) satisfy the conditions detailed in (3) and let $q(\tau)$ satisfy A.1 and A.2 then as $T \rightarrow \infty$, we can make the following two statements:

(i) If in A.2 the integral holds for all $c > 0$ rather than for some $c > 0$, a sequence of Brownian bridges $\{\mathcal{B}_T(\tau)\}$ can be constructed such that the following result holds:

$$\sup_{0 < \tau < 1} \frac{\left| \frac{1}{\Delta^{(i)}} M_T^{(i)}(\tau) - \mathcal{B}_T(\tau) \right|}{q(\tau)} = o_P(1);$$

(ii) and if in A.2 the integral holds for some $c > 0$ rather than for all $c > 0$, then a sequence of Brownian bridges $\{\mathcal{B}_T(\tau)\}$ can be constructed such that the following result holds:

$$\sup_{0 < \tau < 1} \frac{\left| \frac{1}{\Delta^{(i)}} M_T^{(i)}(\tau) \right|}{q(\tau)} \xrightarrow{\mathcal{D}} \sup_{0 < \tau < 1} \frac{|\mathcal{B}_T(\tau)|}{q(\tau)},$$

where $\Delta^{(A)} = \sigma^2 \sqrt{\text{Var}(\varepsilon_1^2)}$, $\Delta^{(B)} = \sigma$ and $\mathcal{B}_T(t) := \frac{W(T\tau)}{\sqrt{T}} - \tau \frac{W(T)}{\sqrt{T}}$. The proof of this can be found in OP, page 7.

As the interest here is with a bivariate process formed out of the two processes given in (6) and (7), OP go on to show that, under H_0 , if we let the process (2) satisfy conditions detailed in (3) and let $q(\tau)$ satisfy A.2 for all $c > 0$, then as $T \rightarrow \infty$,

$$\begin{bmatrix} \sup_{0 < \tau < 1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} \\ \sup_{0 < \tau < 1} \frac{1}{\Delta^{(B)}} \frac{|M_T^{(B)}(\tau)|}{q(\tau)} \end{bmatrix} \xrightarrow{\mathcal{D}} \begin{bmatrix} \sup_{0 < \tau < 1} \frac{|\mathcal{B}^{(A)}(\tau)|}{q(\tau)} \\ \sup_{0 < \tau < 1} \frac{|\rho \mathcal{B}^{(A)}(\tau) + (1-\rho^2)^{1/2} \mathcal{B}^{(B)}(\tau)|}{q(\tau)} \end{bmatrix}, \quad (10)$$

with $\xrightarrow{\mathcal{D}}$ denoting convergence in distribution and $\mathcal{B}^{(A)}(\tau)$ and $\mathcal{B}^{(B)}(\tau)$ representing two independent copies of Brownian bridges and $\rho = \frac{\mathbb{E}[\varepsilon_1^3]}{\sqrt{\text{Var}(\varepsilon_1^2)}}$. Again proof can be found in OP, page 8.

These distributions depend on unknown parameters, *i.e.* the variance, skewness and kurtosis of the error term. Under symmetry of the distribution of the errors of the regression model, when $\rho = 0$,² then the bivariate process in (10) converges in distribution to two independent copies of a weighted Brownian bridge. In this case the testing framework for parameter changes in intercept or slope consists in comparing the test statistics in (6) and (7) to the critical value at a γ significance level b_γ , obtained from the corresponding tabulated asymptotic distribution. For example, if

$$\frac{1}{\Delta^{(A)}} \sup_{0 < \tau < 1} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} > b_\gamma > \sup_{0 < \tau < 1} \frac{1}{\Delta^{(B)}} \frac{|M_T^{(B)}(\tau)|}{q(\tau)},$$

the test detects a break only in the slope parameter. If

$$\frac{1}{\Delta^{(A)}} \sup_{0 < \tau < 1} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} > b_\gamma \text{ and } \frac{1}{\Delta^{(B)}} \sup_{0 < \tau < 1} \frac{|M_T^{(B)}(\tau)|}{q(\tau)} > b_\gamma$$

both intercept and slope parameters have changed. Finally, if the critical value, at a γ significance level is greater than each of the two statistics then there is no evidence to reject the null hypothesis of no structural break in the model that underlies the sample.

OP consider the case when the error has an asymmetric distribution and this is useful for our analysis of mutual funds. Given the dependence of the two processes under study, it is not clear how to construct the relevant asymptotic critical values. To solve this problem OP show that by reformulating the statistics such that, under the same assumptions as before, as $T \rightarrow \infty$,

$$\begin{bmatrix} \sup_{0 < \tau < 1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} \\ \sup_{0 < \tau < 1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} \end{bmatrix} \xrightarrow{\mathcal{D}} \begin{bmatrix} \sup_{0 < \tau < 1} \frac{|\mathcal{B}^{(A)}(\tau)|}{q(\tau)} \\ \sup_{0 < \tau < 1} \frac{|\mathcal{B}^{(B)}(\tau)|}{q(\tau)} \end{bmatrix} \quad (11)$$

where $\mathcal{M}(\tau) = -\rho((1-\rho^2)\sigma^4 \text{Var}(\varepsilon_1^2))^{-\frac{1}{2}} M_T^{(A)}(\tau) + ((1-\rho^2)\sigma^2)^{-\frac{1}{2}} M_T^{(B)}(\tau)$. This bivariate process and the corresponding asymptotic theory enable the construction of two different test statistics for the null hypothesis of no change in parameters of the linear regression model given in equation (2). The statistic in the first row of the bivariate process detailed in (11) remains unchanged by a change in α , *i.e.* it is robust to a change in α of our regression model. The

² $\rho = 0$ if and only if the distribution of the errors in our equation (2) is symmetrically distributed about 0.

second test statistic, however, will be sensitive to a change in either intercept or slope of our regression and is therefore used to define a joint hypothesis

$$H_0 : \alpha^{(1)} = \alpha^{(2)} \text{ and } \beta^{(1)} = \beta^{(2)}. \quad (12)$$

The alternative hypothesis in our joint test corresponds to a change in at least one parameter, either α or β , of the regression model. A value of the test statistic greater than b_γ implies the rejection of H_0 . This test statistic will be referred to as a simultaneous test because it runs simultaneously the two test statistics in equation (11). The element in the first row of this vector tests the null hypothesis $H_{0,\text{slope}} : \beta^{(1)} = \beta^{(2)}$, while the test statistic in the second row tests the null hypothesis H_0 . This simultaneous test is more informative than traditional tests for changes in parameters. It can provide information on which parameter has changed: whether it is α or β . As a test for changes in the parameters, it has one failure. If there is a change in β , it is no longer informative on a possible change in α . This, however, is easily circumvented and more will be said regarding this.

Let us explore further possible outcomes of this test. It is possible, though unlikely, that

$$\sup_{0 < \tau < 1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} > b_\gamma > \sup_{0 < \tau < 1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)}. \quad (13)$$

In this situation, we can conclude only that a change in slope has occurred and as such H_0 is rejected. If, however, the more likely situation should occur, that

$$\begin{aligned} b_\gamma &< \sup_{0 < \tau < 1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} \\ b_\gamma &> \sup_{0 < \tau < 1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} \end{aligned}$$

then we reject H_0 and conclude that there has been a change in α . If, on the other hand

$$\begin{aligned} b_\gamma &< \sup_{0 < \tau < 1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} \\ b_\gamma &< \sup_{0 < \tau < 1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)}. \end{aligned}$$

then we can conclude only that a change in β has occurred. Here, the test is not informative about a change in α . It is necessary to run an auxiliary test based on the statistic $M_T^{(B)}(\tau)/q(\tau)$ to determine whether α has changed as well.

5.2 Parameters Unknown

The processes defined in (6) and (7) depend on unknown parameters. Ordinary Least Squares (OLS) will produce consistent estimators of α_i and β_i for $i = 1, 2$ under H_0 and H_1 . Let these sequences of estimators be denoted $\{\hat{\alpha}_T^{(i)}\}_{T=1}^\infty$ and $\{\hat{\beta}_T^{(i)}\}_{T=1}^\infty$ for $i = 1, 2$. When these sample estimates are substituted for the population parameters, this produces the following slightly altered sequence of partial sum processes:

$$\widehat{M}_T^{(A)}(\tau) := T^{-1/2} \left\{ \sum_{t=1}^{[(T+1)\tau]} (Y_t - \widehat{\alpha}_T - \mathbf{X}_t \widehat{\beta}_T)^2 - \tau \sum_{t=1}^T (Y_t - \widehat{\alpha}_T - \mathbf{X}_t)^2 \widehat{\beta}_T \right\} \quad (14)$$

$$\widehat{M}_T^{(B)}(\tau) := T^{-1/2} \left\{ \sum_{t=1}^{[(T+1)\tau]} (Y_t - \mathbf{X}_t \widehat{\beta}_T) - \tau \sum_{t=1}^T (Y_t - \mathbf{X}_t \widehat{\beta}_T) \right\}. \quad (15)$$

In $\widehat{M}_T^{(A)}(\tau)$ set

$$\widehat{\alpha}_T = \begin{cases} \widehat{\alpha}_T^{(1)}, & t \leq \widehat{t}^* \\ \widehat{\alpha}_T^{(2)}, & t > \widehat{t}^*, \end{cases}$$

and $\widehat{\beta}_T = \widehat{\beta}_T^{(1)}$ and in process $\widehat{M}_T^{(B)}(\tau)$ set

$$\widehat{\beta}_T = \begin{cases} \widehat{\beta}_T^{(1)}, & t \leq \widehat{t}^* \\ \widehat{\beta}_T^{(2)}, & t > \widehat{t}^*, \end{cases}$$

where \widehat{t}^* is some consistent estimator of t^* . One consistent estimator of t^* that has been widely studied in the literature is defined as follows:

$$\widehat{t}^* := \frac{1}{T} \min \left\{ k : \frac{|\widehat{M}_T(\frac{k}{T})|}{q(\frac{k}{T})} = \max_{1 \leq i < T} \frac{|\widehat{M}_T(\frac{i}{T})|}{q(\frac{i}{T})} \right\} \quad (16)$$

$$\widehat{M}_T(t) = T^{-1/2} \left\{ \sum_{t=1}^{[(T+1)\tau]} (Y_t - \widehat{\alpha}_{LS} - \widehat{\beta}'_{LS} \mathbf{X}_t)^2 - \tau \sum_{t=1}^T (Y_t - \widehat{\alpha}_{LS} - \widehat{\beta}'_{LS} \mathbf{X}_t)^2 \right\},$$

where the subscript LS refers to the least squares estimator of α and β using all T observations. The asymptotic properties of this estimator have been studied by Antoch *et al.* (1995). They also show that the bootstrap approximation to this distribution is asymptotically valid. For more on this, we refer those interested to their paper.

OP establish the asymptotic equivalence between test statistics based on $M_T^{(i)}(\tau)$ and test statistics based on $\widehat{M}_T^{(i)}(\tau)$, for $i = A, B$. They show that by substituting estimators for the population parameters in the regression models, all properties established regarding the processes $M_T^{(i)}(\tau)$ continue to hold for $\widehat{M}_T^{(i)}(\tau)$, for $i = A, B$. They also continue to hold when parameters ρ , σ and $Var(\varepsilon_1^2)$ are replaced by any sequence of consistent estimators. Hence they show that if we assume $\{\widehat{\alpha}_T\}_{T=1}^\infty$ and $\{\widehat{\beta}_T\}_{T=1}^\infty$ are sequences of consistent estimators of the parameters in (2), then under the same conditions as before

$$\sup_{0 < \tau < 1} \frac{|M_T^{(i)}(\tau) - \widehat{M}_T^{(i)}(\tau)|}{q(\tau)} = o_P(1),$$

for $i = A, B$, as $T \rightarrow \infty$, where proof can be found in OP, Lemma 2.1 page 11.

5.3 Under the alternative hypothesis

So far results on the asymptotic properties of the statistics detailed in formulae (6) and (7) have been under the null hypothesis of no time variation in the parameters of equation (1). No discussion regarding the power of these statistics to detect such time variation has been mentioned. OP discuss their asymptotic power by way of Corollaries 3.1 and 3.2. Here the asymptotic consistency of these statistics to one-time changes in these parameters is established. In particular, these corollaries advertise that tests based on these statistics are asymptotically consistent. To be more specific, these statistics will detect time varying parameters in equation (1) when at least one of the parameters changes with probability approaching one.

Now that the important features of the structural break tests have been provided, we can continue with the discussion of the simulations in which these tests are used.

6 Simulation

The purpose of this section is to see whether inferences made regarding the cross-section of true α for funds that advertise superior performance, change when the distribution of actual $t(\alpha)$ is adjusted for time-variation in the parameters of equation (1) and this affects our bootstrap simulations. Incorporating time variation allows for more accurate estimation of the distribution of cross-section $t(\alpha)$ estimates for actively managed US funds, and to determine whether this distribution suggests a world where true α is zero for all funds or whether some funds possess nonzero true α . The test used here to answer this question follows closely that employed by FF with adjustments made for time variation in parameters. Their test compares long histories of individual fund returns to bootstrap simulations of these returns. Returns used in the simulations have the same properties as actual fund returns except the true α is set to zero in the return population for which the bootstrap samples are drawn. This is achieved by subtracting a fund's estimated α from its monthly returns. When FF estimate equation (1) on returns for each fund, they obtain a cross-section of $t(\alpha)$ estimates that can be ordered into a cumulative distribution function (CDF) of $t(\alpha)$ estimates for actual fund returns. A simulation run also involves estimating equation (1) but produces a cross-section of $t(\alpha)$ estimates for a world in which true α is zero.

To alter their method to account for time variation in parameters of equation (1), we begin by estimating this equation on each of the 5,785 funds and obtaining the residuals from each estimated model. Next, the simultaneous test detailed in Section 5 is applied to the residuals to determine whether there has been changes in one of the parameters of equation (1). If the simultaneous test does not reject the H_0 of no change in α or β s then equation (1) is estimated assuming $\beta^{(1)} = \beta^{(2)}$ and $\alpha^{(1)} = \alpha^{(2)}$, i.e. one model is estimated:

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}(R_{Mt} - R_{ft}) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{it}. \quad (17)$$

If, however, it rejects H_0 then we engage in further tests, as outlined in Section 5, to determine which parameters have changed: only α , only β or both. If a change in one parameter is detected, equation (1) is separated into two equations and the first is estimated on returns before the change and the second equation is estimated on returns after the change. To be more specific, suppose a change in α is detected by our tests at some unknown time, the return series originally modelled as equation (1) is now separated into two regressions models. The second model departs from

the first only to the extent that α is allowed to change at \hat{t}^* , our estimator of the unknown time of change detailed in equation (16). As there is a change in α only, we estimate the equations described in (18) which are given below:

$$R_{it} - R_{ft} = \begin{cases} \alpha_i^{(1)} + \beta_{1i}(R_{Mt} - R_{fg}) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{it} & 1984 \leq t \leq \hat{t}^* \\ \alpha_i^{(2)} + \beta_{1i}(R_{Mt} - R_{fg}) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{it} & \hat{t}^* + 1 \leq t \leq 2014. \end{cases} \quad (18)$$

If, instead of a change in α , our specially designed tests indicate a change in β only, then the equations described in (19) are estimated:

$$R_{it} - R_{ft} = \begin{cases} \alpha_i^{(1)} + \beta_{1i}^{(1)}(R_{Mt} - R_{fg}) + \beta_{2i}^{(1)}SMB_t + \beta_{3i}^{(1)}HML_t + \beta_{4i}^{(1)}MOM_t + \varepsilon_{it} & 1984 \leq t \leq \hat{t}^* \\ \alpha_i^{(1)} + \beta_{1i}^{(2)}(R_{Mt} - R_{fg}) + \beta_{2i}^{(2)}SMB_t + \beta_{3i}^{(2)}HML_t + \beta_{4i}^{(2)}MOM_t + \varepsilon_{it} & \hat{t}^* + 1 \leq t \leq 2014. \end{cases} \quad (19)$$

Our second model allows only β to change should our tests detect a change in these parameters at time \hat{t}^* .

For the last possibility, suppose there is a change to α as well as to β at \hat{t}_1^* and \hat{t}_2^* respectively, we then estimate the following version of equation (1):

$$R_{it} - R_{ft} = \begin{cases} \alpha_i^{(1)} + \beta_{1i}^{(1)}(R_{Mt} - R_{fg}) + \beta_{2i}^{(1)}SMB_t + \beta_{3i}^{(1)}HML_t + \beta_{4i}^{(1)}MOM_t + \varepsilon_{it} & 1984 \leq t \leq \hat{t}_1^* \\ \alpha_i^{(2)} + \beta_{1i}^{(1)}(R_{Mt} - R_{fg}) + \beta_{2i}^{(1)}SMB_t + \beta_{3i}^{(1)}HML_t + \beta_{4i}^{(1)}MOM_t + \varepsilon_{it} & \hat{t}_1^* + 1 \leq t \leq \hat{t}_2^* \\ \alpha_i^{(2)} + \beta_{1i}^{(2)}(R_{Mt} - R_{fg}) + \beta_{2i}^{(2)}SMB_t + \beta_{3i}^{(2)}HML_t + \beta_{4i}^{(2)}MOM_t + \varepsilon_{it} & \hat{t}_2^* + 1 \leq t \leq 2014 \end{cases} \quad (20)$$

A similar process is followed when we encounter two changes in a parameter. As it is more difficult to describe, for brevity we leave out listing five possible cases that would need to be accounted for. Lastly, if an estimated location of change in one of the parameters should occur very near to the start or end of the sample, such that it prevents reliable estimation of one of the models detailed in equations (18) to (20), then we assume no change has occurred and estimate one model.

For FF, a simulation run consists of a sample drawn randomly from the 273 months of returns. Rather than using this method, we employ a fixed-design bootstrap. This requires estimating the 4F-CAPM model on each of the bootstrapped adjusted returns series allowing for time-varying parameters. The returns series used in the simulation were generated by randomly sampling residuals produced from estimating one of many versions of equation (1), some of which are detailed in equations (17) to (20) on the original returns series. We find this method easier to implement than the bootstrap method implemented by FF. This method is well described in Chapter 9 of Efron and Tibshirani (1993); we refer those interested in more information

on this method to this chapter. Following FF, we also use 10,000 simulation runs to produce two distributions of t -statistics generated from estimating α : one distribution corresponding to estimating equation (1) assuming no time variation in parameters, the other distribution is generated allowing for time variation to the parameters of equation (1). As in FF, the focus of this simulation is on the t -statistic associated with the estimate of α , which is referred to as $t(\alpha)$, rather than estimates of α . This allows control for differences in precision in which α is estimated.

6.1 Results

Table II reports results on the CDF of the actual $t(\alpha)$, the CDF of bootstrap simulations of a world of no skill and the %<Actuals. The results under the heading 'Without variation in α or β s' are those for the 4F-CAPM where the parameters are kept constant throughout the sample period and those under the heading 'With variation in α or β s' are those for the 4F-CAPM where the parameters are allowed to change if the testing procedure has detected a break.

First, equation (1) was estimated on each of the 5,785 funds' return series assuming no change in the parameters of the 4F-CAPM. This results in a cross-section of $t(\alpha)$ estimates were then ordered and percentiles for the distribution reported in the first of the two columns labelled 'Actual $t(\alpha)$ ', under the general heading 'Without variation in α or β s'. The second column labelled 'Actual $t(\alpha)$ ', under the general heading 'With variation in α or β s', reports the CDF after applying the test statistics developed in Section (5) that allow for one or more changes in the parameters. The results of these test statistics being used to decide which version of equation (1), detailed in equations (18) to (20), was estimated taking into account changes in parameters.

Table II also lists percentiles for the simulated distribution of $t(\alpha)$ associated with the simulated α under the two sub-headings 'Bootstrapped $t(\alpha)$ '. Our procedure for achieving this is that described in the last paragraph of Section 3. Again, this is carried out separately for the estimates with and without time variation in the estimated parameters. Lastly, Table II also reports fractions of the 10,000 simulated bootstrap α that produces lower values of the percentiles of the CDF from the actual $t(\alpha)$ estimates in the columns with the headings '%<Actual'.

Note that we only analyse gross fund returns, *i.e.* costs associated with trading were not deducted, whereas FF report results based on gross returns (gross returns for FF include trading costs but exclude management expenses) and net returns (net returns deduct both trading costs and management expenses).

To develop some perspective of the results produced by our simulation, it is useful first to make qualitative comparisons of the centiles of the cross-section of $t(\alpha)$ estimates from actual fund returns to average values of the percentiles from the simulations. This is carried out by comparing the values of $t(\alpha)$ at selected centiles of the CDF of the $t(\alpha)$ estimates from actual fund returns to percentiles formed by averaging across the 10,000 simulation runs of $t(\alpha)$ estimates. In the world of the simulation, true α is set to zero. As such, the CDF formed by averaging centiles across the simulations correspond to a world where the true α is zero and comparisons of the actual CDF to the corresponding bootstrapped CDF are informative on whether managers have skill to generate returns that beat returns on the passive benchmark.

We then move onto more detailed analysis of these results via comparison of the likelihoods. Comparing likelihoods allows us to judge whether the tails of the cross-section of $t(\alpha)$ estimates for actual fund returns are extreme relative the simulated results where the true α is zero. This

Table II: Distribution of t -statistics on α estimates in 4F-CAPM for US domestic funds

Centile	Without time variation in α or β s			With time variation in α or β s		
	Actual	Bootstrap	%<	Actual	Bootstrap	%<
	$t(\alpha)$	$t(\alpha)$	Actual	$t(\alpha)$	$t(\alpha)$	Actual
1	-3.90	-27.54	24.70	-4.15	-25.73	24.61
2	-3.41	-21.45	26.78	-3.67	-20.50	26.86
3	-3.15	-18.07	27.94	-3.32	-17.51	28.62
4	-2.99	-15.76	28.73	-3.11	-15.53	29.70
5	-2.85	-13.86	29.41	-2.95	-13.79	30.54
10	-2.34	-9.47	32.11	-2.39	-9.43	33.75
20	-1.74	-5.25	35.58	-1.78	-5.35	37.74
30	-1.37	-2.74	37.99	-1.39	-3.05	40.62
40	-1.05	-1.08	40.23	-1.04	-1.47	43.31
50	-0.71	0.24	42.66	-0.70	-0.22	45.96
60	-0.40	1.60	45.10	-0.36	1.03	48.83
70	-0.08	3.25	47.53	-0.03	2.48	51.59
80	0.31	5.67	50.48	0.40	4.37	55.05
90	0.90	10.23	54.90	1.17	8.21	61.10
95	1.39	15.11	58.47	1.94	12.85	66.58
96	1.52	17.05	59.44	2.21	14.26	68.31
97	1.70	19.51	60.67	2.57	16.87	70.54
98	2.00	22.76	62.58	3.12	20.30	73.98
99	2.30	27.58	64.47	4.21	26.02	79.35

provides information on whether some managers lack sufficient skill to beat the benchmark portfolio and whether some managers have skill. For example if low fractions of simulated values of $t(\alpha)$ are less than centiles in the extreme lower tail of actual $t(\alpha)$ estimates, we conclude that most managers lack skill to beat the benchmark portfolio. There may be some inferior funds: funds where the true α is negative but on average most funds produce returns on par with the benchmark. Similarly, we conclude few managers have sufficient skill to beat returns on the passive benchmark if large fractions of the simulation runs produce $t(\alpha)$ values below those centiles in the extreme right tail of actual $t(\alpha)$ estimates. If this fraction is sufficiently large, it may suggest the existence of a few superior performing funds: a small number of funds may produce large returns.

6.2 Comparing the CDFs of actual $t(\alpha)$ estimates

Here, we compare CDFs of actual $t(\alpha)$ estimates with CDFs from our simulated bootstrapped $t(\alpha)$ estimates reported in Table II. We first compare centiles produced for actual $t(\alpha)$ estimates listed in both sections. When time variation is allowed for, the 95th centile in the upper tail

(large centiles) is 1.94 but only 1.39 when no time variation is allowed. The disparity between corresponding centiles of the CDFs increases: the 99th with-time-variation centile is 4.21 which is substantially larger than 99th without time-variation centile of 2.30. Next, comparing centiles on the lower tails (small centiles), we do not see a similar effect of time variation on these centiles as we did for the upper tail. The 1st centile of the CDF with time variation is -4.15 which is slightly smaller than the corresponding 1st centile of -3.90 without changes. Comparing centiles from the 10th to 80th, we note that they are similar in magnitude. It is clear from Table II and the results contained therein allowing for time variation has had much larger effect on the upper tails of the actual $t(\alpha)$ estimates than on the lower tail. Allowing for time variation in the parameters of the 4F-CAPM has had a significant impact on CDFs of actual $t(\alpha)$ estimates. This can have significant effects on conclusions made regarding the existence of superior funds.

Now we compare centiles from the CDF of actual $t(\alpha)$ estimates to centiles of simulated bootstrapped $t(\alpha)$ s. Centiles in the lower tail of simulated $t(\alpha)$ s CDF are much smaller in magnitude when compared to the centiles of actual $t(\alpha)$ estimates. For example, the 3rd, 4th and 5th centiles of the simulated CDF (recorded in the Bootstrapped t column located in the 'Without time variation in α or β s' section of the Table II) are -18.07, -15.76 and -13.86, while the corresponding centiles of the CDF of actual $t(\alpha)$ estimates are -3.15, -2.99 and -2.85. We can see that simulated centiles are much smaller. Comparing centiles in the upper tail of simulated and actual CDFs, we find simulated centiles are much larger than actual centiles detailed in the 'With time variation in α or β s' section of Table II. This indicates that the bootstrap simulated t -distribution is much more dispersed than the actual t -distribution, both with and without time varying parameters.

Another interesting consequence of time variation in parameters can be observed when comparing simulated CDFs (see information contained in Bootstrap t columns of Table II) with and without time varying parameters: simulated CDF that includes time variation produces larger centiles in the lower tail (e.g. 1st centiles of -25.73 is less than -27.54) and smaller centiles in the upper tail (e.g. 99th centile of 26.02 compared to 27.58). This result again confirms the significant influence time variation can have on simulated $t(\alpha)$ estimates.

The simulation suggests that, after adjusting for time varying parameters, relatively few managers generate returns above passive benchmark portfolios. Nevertheless, there appear to be more occurrences of skilled managers after accounting for time variation than when not accounting for it. Analysis via likelihoods will enable us to say much more on the presence of skilled and unskilled managers.

6.3 Likelihoods (i.e. %<Actual)

As mentioned in Section 6.2, comparing CDFs based on centiles of $t(\alpha)$ estimates of actual fund returns with centiles from the simulation CDFs can be only suggestive of how manager skill affects expected returns. FF rely more on results from the analysis of %<Actual (FF call them likelihoods) than comparing centiles of actual and simulated $t(\alpha)$ estimates. Table II also records the likelihoods in the column header titled %<Actual. This column records the proportion of the 10,000 simulation runs that produce values of $t(\alpha)$ s lower than the centiles from the CDF of actual $t(\alpha)$ estimates. The likelihoods allow one to assess more formally whether the tails of the cross-section of $t(\alpha)$ estimates for actual fund returns are extreme relative to what is observed from a CDF where the true α is set to zero. We follow FF and infer that some managers lack skill

sufficient to cover costs if low fractions of the simulation runs produce left tail centiles of $t(\alpha)$ estimates below those from actual net fund returns. Likewise, FF infer some managers produce benchmark-adjusted expected returns that cover costs if large fractions of the simulation runs produce right tail centiles of $t(\alpha)$ estimates below those from actual fund returns.

Turning now to interpreting results in the %<Actual column. Comparing the two columns of %<Actual, we see that the percentages are similar for both CDFs at percentiles in the lower tail of this distribution but these percentages begin to differ after the 50th centile. At the 60th centile 48.83% of the simulations produced centiles of $t(\alpha)$ estimates smaller than this actual $t(\alpha)$ estimates centile (*cf.* with the 'With time variation' section of Table II) which is larger than 45.1%, the corresponding percent when no time variation is allowed. At the 90th centile, the percentage of simulated $t(\alpha)$ centiles less than the corresponding centiles for the actual $t(\alpha)$ estimates differ substantially (61.10% when time variation is allowed and only 54.90% without variation). By the 99th centiles, the percentage of simulated $t(\alpha)$ centiles less than the centiles from the actual $t(\alpha)$ estimates differ by 15% (79.35% when time variation is allowed for and 64.47% when no time variation is allowed). Time variation in parameters of the 4F-CAPM has had a dramatic effect. The fraction of simulated $t(\alpha)$ centiles that are less than centiles of actual $t(\alpha)$ estimates in the upper tail increase significantly relative to the lower tail. This result is not surprising. Time variation has served to increase the magnitude of the centiles in the upper tail of the CDF of actual $t(\alpha)$ estimates which reduces the frequency at which simulated $t(\alpha)$ s exceed them. This result also indicates that allowing for time-varying parameters in our bootstrap simulation $t(\alpha)$ estimates does not increase the variability of our $t(\alpha)$ estimates. Indeed, it has reduced their variability.

What does this imply about the presence of skilled managers? For centiles in the lower tail of the CDF of actual $t(\alpha)$ estimates, the fraction of simulated $t(\alpha)$ centiles that are less than the actual $t(\alpha)$ centiles is large. From this result, we find that managers generate returns that match the passive benchmark portfolio and conclude managers do not systematically under perform the benchmark portfolio. For centiles in the upper tail of actual $t(\alpha)$ estimates, the conclusions we draw from the %<Actual column is that there are a few managers who generate returns that beat the passive benchmark. Even though the percentage of the simulated $t(\alpha)$ centiles that are now higher when time variation is allowed than when it is not, the percentage is not sufficiently large to conclude there are few managers with sufficient skill to outperform the benchmark portfolio.

7 Conclusions

Much of the literature on mutual performance ignores the issue of time variation in performance. For example, Fama and French (2010) evaluate the performance of actively managed US mutual funds and find, when returns are net of expense and trading costs, that fund returns do not cover these costs. However when measured gross of management expenses, they find the presence of some inferior and some superior performing funds. Baras *et al.* (2010) and Cuthbertson *et al.* (2008) use the conditional 4F-CAPM of Ferson and Schadt (2010) that allows for time-variation in the slope parameters of this model. Baras *et al.* (2010) find using conditional versions of the 4F-CAPM has little effect on their simulations and produces results, and hence conclusions, similar to those produced using unconditional version of the 4F-CAPM.

Fama and French (2010) report that time variation is a thorny problem for their analysis

and leave the issue to future research. Using tests developed by Olmo and Pouliot (2014), we adapt their method to allow for any type of time-variation in parameters of the 4F-CAPM and adjust their bootstrap simulation to accommodate time varying parameters. Using gross returns on actively managed US mutual funds over the period from January 1984 to October 2014, we find significant evidence of time-variation in the parameters of the 4F-CAPM and that this time-variation has substantial effects on bootstrap simulations of $t(\alpha)$ estimates, the measure used to evaluate fund performance. Even though there is much time variation and it has substantial effects on the bootstrap simulations of $t(\alpha)$, we find actively managed US mutual funds generally produce returns that are similar to passive benchmark portfolios and conclude from this that there is little evidence of inferior or superior mutual fund performance.

Appendix

Table III: Distribution table for structural break tests

$G(x) = \mathbb{P} \left\{ \sup_{0 < \tau < 1} \frac{ \mathcal{B}(\tau) }{q(\tau, \nu = \frac{15}{128}) < x} \right\}$							
x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$
0.53	0.01	0.83	0.26	1.01	0.51	1.23	0.76
0.57	0.02	0.83	0.27	1.01	0.52	1.24	0.77
0.60	0.03	0.84	0.28	1.02	0.53	1.25	0.78
0.62	0.04	0.85	0.29	1.03	0.54	1.26	0.79
0.63	0.05	0.85	0.30	1.04	0.55	1.28	0.80
0.65	0.06	0.86	0.31	1.05	0.56	1.29	0.81
0.66	0.07	0.87	0.32	1.05	0.57	1.30	0.82
0.67	0.08	0.88	0.33	1.06	0.58	1.32	0.83
0.68	0.09	0.88	0.34	1.07	0.59	1.33	0.84
0.70	0.10	0.89	0.35	1.08	0.60	1.35	0.85
0.71	0.11	0.90	0.36	1.09	0.61	1.37	0.86
0.72	0.12	0.90	0.37	1.10	0.62	1.39	0.87
0.73	0.13	0.91	0.38	1.10	0.63	1.40	0.88
0.74	0.14	0.92	0.39	1.11	0.64	1.42	0.89
0.75	0.15	0.93	0.40	1.12	0.65	1.45	0.90
0.75	0.16	0.93	0.41	1.13	0.66	1.47	0.91
0.76	0.17	0.94	0.42	1.14	0.67	1.50	0.92
0.77	0.18	0.95	0.43	1.14	0.68	1.53	0.93
0.78	0.19	0.96	0.44	1.15	0.69	1.56	0.94
0.79	0.20	0.97	0.45	1.16	0.70	1.61	0.95
0.79	0.21	0.97	0.46	1.18	0.71	1.67	0.96
0.80	0.22	0.98	0.47	1.19	0.72	1.72	0.97
0.81	0.23	0.99	0.48	1.20	0.73	1.80	0.98
0.81	0.24	0.99	0.49	1.21	0.74	1.92	0.99
0.82	0.25	1.00	0.50	1.22	0.75	2.04	1.00

References

- [1] Admati, A. and Ross, S. (1985). Measuring investment performance in a rational expectations equilibrium, *Journal of Business* 58, 1-26.
- [2] Altissimo, F., and Corradi, V. (2003). Strong Rules for Detecting the Number of Breaks in a Time Series, *Journal of Econometrics*, 117, pp. 207-244.
- [3] Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, 61 pp. 821-859.
- [4] Andrews, D. W. K. and Ploberger, W. (1994). Optimal Tests when a Nuisance Parameter is Present only Under the Alternative. *Econometrica*, 62, No. 6 (November), pp. 1383-1414.
- [5] Antoch, J., Hušková, M. and Veraverbeke, N. (1995). Change-Point Problem and Bootstrap. *Journal of Non-parametric Statistics*, 5, pp. 123-144.
- [6] Bai, J., and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66, pp. 1, 47-78.
- [7] Barras, L., Scaillet, O., and Wermers, R. (2010). False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas. *Journal of Finance*, LXV, pp. 179-216.
- [8] Berndt, E (1990). The Practice of Econometrics, Classical and Contemporary. Addison-Wesley Publishing company.
- [9] Billingsley, P. (1968). Convergence of Probability Measures, John Wiley & Sons, New York.
- [10] Black, F., Jensen, M. and Scholes, M. (1972). The Capital Asset Pricing Model: Some Empirical Tests, in Jensen, M. ed. *Studies in the theory of Capital Markets*, New York: Praeger, pp. 79-121.
- [11] Brown, S., Goetzmann, W., Ibbotson, R. and Ross, S. (1992) Survivorship Bias in Performance Studies. *Review of Financial Studies* 5, pp 553-580.
- [12] Brown, R., Durbin, J. and Evans, J. (1975). Techniques for Testing the Constancy of Regression Coefficients over time, *Journal of The Royal Statistical Society Series B (Methodology)*, 37, pp. 149-192.
- [13] Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. (1986) Weighted Empirical and Quantile Processes. *Annals of Probability* 14, pp. 32-85.
- [14] Carhart, M. (1997). On the Persistence in mutual fund performance, *Journal of Finance*, 52, 57-82.
- [15] Cuthbertson, Nitzsche, D. and O'Sullivan, Niall (2008). UK Mutual Fund Performance: Skill or Luck?, *Journal of Empirical Finance*, 15, 613-634.
- [16] Dybvig, P. and Ross, S. (1985). The Analytics of Performance Measurement Using a Security Market Line, *Journal of Finance*, XL, 401 - 415.

- [17] Efron, B. and Tibshirani, R. (1998). An Introduction to the Bootstrap, Chapman & Hall/CRC press.
- [18] Elton, E., Gruber, S. D., Hlavka, M. (1992). Efficiency with costly information: A reinterpretation of evidence form managed portfolios, *Review of Financial Studies*, vol. 6, 1-22.
- [19] Fama, E. and French, K. (2010). Luck versus Skill in the Cross-Section of Mutual Fund Returns, *Journal of Finance* LXV, 1915-1947.
- [20] ————— (1993). Common Risk Factors in the Returns on stocks and bonds, *Journal of Financial Economics* 33, pp. 3-56.
- [21] Ferson, W.E. and Schadt, R.W. (1996) Measuring Fund Strategy and Performance in Changing Economic Conditions Source: *The Journal of Finance*, Vol. 51, No. 2, June, pp. 425-461.
- [22] Gombay E. and Horváth L. and Hušková, M. (1996). Estimators and Tests for Change in Variances. *Statistics & Decisions*, 14, pp. 145 - 159.
- [23] Jensen, M. (1968). The Performance of Mutual Funds in the Period 1945-1964. *Journal of Finance* 23, pp. 389-416.
- [24] Kuan, C. M. and Hornick, K. (1995) The Generalized Fluctuation test: A unifying view. *Econometrics Reviews* 14, pp. 135-161.
- [25] Kristensen, D. (2012). Non-parametric detection and estimation of structural change. *Econometrics Journal* 15, pp. 420-461.
- [26] Kapar, B. and Pouliot, W. (2013). Multiple Change-point Detection in Linear Regression Models via U -statistic type Processes. Department of Economics Discussion Paper, University of Birmingham.
- [27] Olmo, J. and Pouliot, W. (2011) Early Detection for Market Risk Failure, *Studies in Non-linear Dynamics & Econometrics* 15, pp. 1-20.
- [28] Olmo, J. and Pouliot, W. (2014) Tests to Disentangle Breaks in Intercept from Slope in Linear Regression Models with Application to Management Performance in the Mutual Fund Industry. Department of Economics Discussion Papers No 14-01, University of Birmingham.
- [29] Olmo, J., Pilbeam, K. and Pouliot, W. (2011). Detecting Insider Trading via Structural Break Tests, *Journal of Banking and Finance* 35, pp. 2820-2828.
- [30] Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, pp. 1361-1401.
- [31] Sharpe, W. and Cooper, M. (1972). Risk-Return Classes of New York Stock Exchange Common Stocks, 1931-1967 *Financial Analysts Journal* 28, 46-54.