Power outages and firm performance

Elliott, Robert J.R.; Nguyen-Tien, Viet; Strobl, Eric A.

DOI:
10.1016/j.eneco.2021.105571

Power Outages and Firm Performance: A Hydro-IV Approach for a Single Electricity Grid

Robert J. R. Elliott[1], Viet Nguyen-Tien[2], Eric A. Strobl[3]

[1] University of Birmingham, UK
[2] London School of Economics, UK
[3] University of Bern, Switzerland

JEL: O13, O44, Q54, R11, N55

Keywords: Power outages, Firm Performance, Vietnam, Growth, Instrumental Variables.

Power Outages and Firm Performance: A Hydro-IV Approach for a Single Electricity Grid

August 11, 2021

1 Introduction

Access to a reliable power is arguably essential if a country is to industrialize and continue to grow. However, in many developing countries access to electricity remains limited (Lipscomb *et al.*, 2013), while in others, although there may have been great progress made in overall electrification rates, consumption of electricity per head remains low. At the same time, an unreliable power supply can hinder firm performance which can have a wider impact on economic growth. Despite the obvious economic importance, power reliability in developing countries is generally given far less attention than power accessibility concerns (Meles, 2020).

A primary challenge in quantifying the impact of power reliability on firm performance is that there are a number of potential endogeneity concerns, including measurement error (Allcott *et al.*, 2016), selection bias (Alam, 2013), and simultaneity. The recent approach in the literature to address these concerns has been to focus on hydropower generation under the realisation that this energy source can largely be explained by exogenous shocks from weather variability, and that this variability can be used as an instrument for power outages. Such an identification strategy has recently been employed in different contexts, such as India (Allcott *et al.*, 2016) and Sub-Saharan Africa (Mensah, 2016; Cole *et al.*, 2018), where it has been shown that the potential endogeneity bias is non-negligible and tends to lead to substantial under-estimates of the impact of outages on firm performance. Nonetheless, the current state-of-the-art is restricted to investigating multiple-grid cases (different countries or states within a large country), where access to the different grids provides the variation in firm level power provision. Such an approach is thus not applicable to single grid contexts, i.e., where all firms and utilities are connected to the same centrally managed grid. In this paper we extend the current approach to a single grid network by developing a hydro-instrumental variable strategy which integrates a river flow model with a hydropower generation model and an electricity-grid-based distance interpolation technique, using the case study of Vietnam.

To investigate the impact of outages on firm performance in a single grid context we match firms with (hydro)power plants that are connected to a single grid. Given that we, as in most cases, have limited information on the electricity distribution rules (if any) within this single grid system, and that these decision rules might themselves be endogenous, we are presented with several challenges. First, because of the interconnected system it is difficult to match the power provision of a firm to particular large hydropower plant(s) as the impact of reduced electricity production could be nationwide. Second, a number of fossil fuel power plants are connected to the grid and can be used to increase the power supply if production from hydropower plants is reduced. Our solution is to take an interdisciplinary approach and use a rainfall-runoff model based on Soil and Water Assessment Tool (SWAT) to simulate river flows to the 40 largest hydropower dams across Vietnam. This allows us to take into account a variety of terrain conditions and variation in the weather. The simulated series from the SWAT model are then used to predict hydropower generation, thus eliminating the component of electricity production driven by changes in demand. The estimated series from the electricity generation model are subsequently used to construct an index for the weighted hydro-plant factors that are explained by the exogenous variation in hydrological conditions, using a grid-based distance penalty parameter calibrated by the 'reduced-form equation' of the IV estimation. Finally, the hydro-index is employed as a single instrument for power outages measured across multiple dimensions in the 'structural equation' to address the endogeneity concerns.

The response of firms to an unreliable electricity supply can vary. At its simplest it means additional costs, if, for example, the firm needs to purchase and operate a backup generator (that also comes with a higher unit cost of electricity). One implication is that firms with limited financial resources will find it harder to access electricity-intensive sectors (Reinikka and Svensson, 2002; Adenikinju, 2003; Alby and Dethier, 2013). A further consequence of increased costs is the impact on productivity (Mensah, 2016) and reduced firm size (Allcott *et al.*, 2016; Grainger and Zhang, 2017), and hence lower profitability (Doe and Asamoah, 2014). As a result, firms may mitigate the impact of an unreliable energy supply by switching to less energy-intensive technologies (Alam, 2013), or may substitute electricity for other fuel types (Allcott *et al.*, 2016), or materials (Fisher-Vanden *et al.*, 2015) (where firms outsource the production of energy intensive intermediates instead of making them in house). In addition, these factors will lead the firm to have a revenue gap that is wider than the productivity gap, as the former is driven by the combination of the latter and the reduction in input usage Allcott *et al.* (2016). We use our single grid Hydro-IV approach to test these implications for the case of Vietnamese firms, using the 2005 and 2015 World Bank Enterprise surveys for Vietnam.

Vietnam arguably represents an ideal country to study the changing dynamics in the power sector as it has enjoyed strong economic growth alongside a rapid expansion of electrification across the country. As a result, it has seen electrification rates increase from under 2.5% in 1975 to around 96% in 2009 (Min and Gaba, 2014). Between 1990 and 2013, Vietnam's growth rate has averaged round 6.8% a year (ADB, 2015), the $1.90-a-day poverty rate reduced to under 3% (WB and MPI, 2016), and in 2008 the economy entered the group of lower-middle income countries. During this time annual power consumption has grown at double digit rates while at the same time Vietnam has almost completed the process of universal electrification with the capacity of the power system reaching 15 GW of installed capacity in 2010 (Vagliasindi and Besant-Jones, 2013).[1] However, despite

[1]Since the 1986 social-economic reforms (Doi moi), and the post-1995 energy sector reforms, almost the entire population

considerable investment in electricity infrastructure, the reliability of the power network remains patchy, driven in part of hydrological uncertainty.[2] Despite such seemingly impressive progress, the quality of the installed electricity was ranked only 113th out of 144 countries by the Global Competitiveness Index 2012-14 (Cattelaens *et al.*, 2015). The result is that the average customer experiences between 18 and 40 power interruptions per year (equivalent to approximately 3,000 to 8,000 minutes per year) (EVN, 2017). Our ability to estimate the impact of outages on firm performance for two separate years allows us to compare their importance both at the early and late stage of electricity infrastructure development in the country, which is in contrast to most existing cross-sectional and panel studies; see Mensah (2016); Cole *et al.* (2018); Allcott *et al.* (2016); Alam (2013); Fisher-Vanden *et al.* (2015); Grainger and Zhang (2017).

To briefly summarize our results, we find that the impact of power outages on firm performance is relatively small but became more important in the later period as firms increased their dependency on electricity. In 2015 we find a significant reduction in revenue of between 0.73% to 1.81% in response to a 1% increase in power outages. In other results we show that frequent outages are found to cause larger losses than a small number of long-lasting outages

The remainder of the paper is organised as follows. Section 2 describes our dataset. Section 3 presents our empirical strategy, with the results shown in Section 4. Section 5 concludes.

of nearly 100 million people have been connected to the electricity grid. Just under 2% of households remain off the grid and in 2014 just under 3% reported that their electricity needs were not being met (Ha-Minh and Nguyen, 2017).

[2] Vietnam's great strides in electrification have been driven in part by its considerable investment in hydropower production that accounts for between 37.6% to 40.2% of installed capacity (ADB, 2011; EVN, 2015*a*,*b*)

2 Data

2.1 Firm-level Data

Our firm level data comes from the World Bank Enterprise Surveys (WBES) that includes a range of questions on infrastructure and firm performance.[3] The WBES data are collected from interviews with owners/ top managers of registered companies in the manufacturing, and services sectors, with at least five employees. For Vietnam, three surveys were undertaken in 2005, 2009, and 2015. Each WBES survey records firm-level data for the previous fiscal year. Hence, the 2005, 2009, and 2015 surveys capture the business environment and firm performance indicators for the years 2004, 2008, and 2014, respectively. In this paper we use the 2005 and 2015 surveys.[4] These surveys samples five regions out of eight most important regions in terms of economic activity (See Appendix A.1 for details).

Firm characteristics. As our study includes a spatial dimension it is important to have firm location information. We choose the province as the spatial unit of analysis based on an administrative GIS map for 2005. To account for the expansion of the capital city (Hanoi) after 2008, all centroid-based variables for firms in Hanoi in Survey 2015 were constructed using weights for the old Hanoi area, and the former Ha Tay area computed from their industrial values just before the expansion.[5] We redefine sector, and size variables, using variables collected from the face-to-face interview phase of the

[3] The WBES covers 127,000 firms across 139 countries and also asks questions on finance, corruption, crime, competition, labor, obstacles to growth. The WBES has been operating since the 1990's, and has been run from the Enterprise Analysis Unit (EAU) since 2005-06.

[4] We exclude the 2009 survey as the 2007-2008 financial crisis means that the 2008 firm level characteristics are less reliable.

[5] In 2007 industrial gross output at current prices for Hanoi and Ha Tay are 116,096.4, and 20,173.5 (billion VND), respectively. The weight applied is 1:0.1737. We ignore the rural areas that used to belong to Hoa Binh, and Vinh Phuc provinces, and latter appended to Hanoi due to their limited economic importance.

WBES surveys.[6] A series of variables were created to control for firm heterogeneity, including sector dummies (based on the first two digits of the main product ISIC (International Standard Industrial Classification) code), four size dummies, a firm age variable, four dummies for state and foreign ownership at the 10%, and 50% thresholds, a share-holding dummy, a publicly quoted company dummy, a dummy for access to credit, and an exporter dummy. For more details on the original variables, treatments for missing values, and the data cleaning process see Appendix A.1.

Power supply provision. We use two variables to measure the frequency and intensity of power outages: (1) the average number of power outages per month, and (2) the typical duration of an outage. To analyze the impact of both factors we create a variable that proxies power outage volume (hours per month), calculated as the product of power outage frequency (occurrences per month), and power outage intensity (hours per occurrence).[7] We construct a variable for generator usage as a percentage of electricity used, and assign a zero value for those firms that do not own or share a generator. See Appendix A.2 for details.

Firm performance. Monetary variables (revenues, and input variables, including the replacement value of machinery, vehicles, and equipment (proxied for capital stock); materials cost; labor cost; fuel cost; electricity cost, and energy cost) are deflated using the World Bank deflators and then logged (base year 2010=100). The energy cost variable is only available for the 2005 Survey, and the electricity, and fuel cost variables are only available for the 2015 Survey. To reduce the impact of outliers we apply the 'three sigma rule' to account for extreme values in the revenue and factor input variables.[8] We also estimate a two total factor productivity revenue (TFPR) variable based on

[6] The surveys were designed as a two-stage procedure. The variables in the first stage (screening by phone) include sector and size are and considered less reliable than those collected in the second stage (face-to-face interview with firm owners/managers).

[7] The three proxies for power quality were log transformed after being added to 1 (to address the log of zero problem).

[8] We calculate the mean, and the standard error of the log of firm revenue for each year in the original database then define 'extreme values' as those that are more than three standard errors deviations from the mean. A similar process is applied for

YKL and YKLM models using the method suggested by EAU-WB (2017). We pool the data of two surveys to obtain a decent sample and estimate the log of revenues as a function of the logs of input values separately for different sector groups, including province and year fixed effects to control for unobserved spatially variant and temporally variant factors. TFPR is computed as the component of fitted revenues that are not explained by input factors (details of how we measure the efficiency of input usage are given in Appendix A.3).

2.2 Province-level Data

It is possible that certain characteristics of a province may simultaneously affect firm performance and power provision. We construct two variables to control for economic conditions at the province level based on data from the Statistical Yearbooks of Vietnam by Vietnam's General Statistics Office (GSO).[9] Province industrial product (IP) share is calculated as the ratio of the gross industrial output of each province to national gross industrial output for each year and is a proxy for agglomeration economies and also captures the importance of each province to the national economy. We argue that this variable may affect both firm performance and the priority that may be given to power distribution when there are shortages (for example, priority is given to areas of strategic economic importance). Second, we calculate a Province IP index to capture IP growth at the province level that could both improve a firm's performance but worsen power reliability if power supply significantly lags demand.[10]

We also control for topography differences. In general, we hypothesize that less elevated provinces offer better conditions for business (i.e. better transportation and access to the sea ports). However,

factor input variables.

[9] Data is accessible at http://www.gso.gov.vn/Default_en.aspx?tabid=515.

[10] See Appendix B1 for details

they may also be more susceptible to electricity interruptions caused by an overloaded system. Similarly, in highly-elevated provinces it may be harder to maintain the distribution and transmission pf electricity. To control for elevation we use the void-filled DEM from the HydroSHEDS database to calculate the mean elevation of each province where the boundaries are defined by the GAUL dataset.[11]

In addition to elevation, we also include controls for temperature that can both affect power provision (hot weather increases electricity demand for air-conditioners, and hence is more likely to cause a system overload) and firm performance (that could, for example, negatively affect labor productivity). Following the literature, we control for cooling degree (Allcott *et al.*, 2016) which is derived from the forecast variable of air temperature at 2m extracted from a gridded temperature data set (NCEP-DOE Reanalysis 2 provided by the NOAA/OAR/ESRL PSD). See Appendix B.2 for more detail of the spatial interpolation and computation methods used in the paper.

Previous studies suggest that rainfall shocks can affect the economy through multiple channels other than through its effect on hydropower generation. For example, more rain may increase agriculture-related activities, raise electricity demand in rural areas due to increased income for farmers, and together with storms, may affect power transmission, and the electricity distribution network (Alam, 2013). In addition, floods frequently hamper transportation networks. To measure the impact of rainfall shocks, we construct a Standardized Precipitation Index (SPI) for each province derived from 'Terrestrial Air Temperature, and Precipitation: Monthly Climatologies' (version 4.1) by the Daleware University (Matsuura and Willmott, 2009). Positive values of SPI indicate a positive rainfall shock while a negative SPI indicates a negative rainfall shock. Near-zero values of SPI indicate normal rainfall conditions while large values indicate extreme weather conditions.[12]

[11] HydroSHEDS DEM was derived from Space Shuttle flight for NASA's Shuttle Radar Topography Mission (SRTM) at three arc-second resolution.

[12] As the measure fits a rainfall series into a gamma series to account for the skewness of rainfall, it has a number of benefits

2.3 Hydrological and Hydropower Data

Our interdisciplinary approach to tackle endogeneity is based on a variety of data. In terms of topographic and hydrological data, we rely on HydroSHEDS (Hydrological data, and maps based on SHuttle Elevation Derivatives at multiple Scales) (Lehner *et al.*, 2008), and its subset HydroBASINS (Lehner and Grill, 2013).[13] Soil and land cover information is extracted from Soil Map of the World (DSMW) (version 3.6) (FAO, 2007) and the University of Maryland Department of Geography (UMD) Land Cover classification collection at the 1km pixel resolution (Hansen *et al.*, 1998, 2000). Daily weather data for the watershed (the maximum, and minimum temperature, precipitation, wind speed, relative humidity, and solar radiation) were supplied by 2,755 gridded stations from the Climate Forecast System Reanalysis (CFSR) which is part of the US's National Centers for Environmental Prediction (NCEP) (Saha *et al.*, 2010, 2014). Our hydropower operation data (installed capacity and electricity generation) at the plant level was obtained from Electricity of Vietnam (EVN, 2015b).

2.4 Summary Statistics

Table 1 provides a series of summary statistics for our regression sample. Compared with the 2005 sample, the sample in 2015 is characterized by a higher number of smaller firms; and domestic private firms and fewer exporters; publicly quoted firms; and firms that have access to credit. In terms of power provision, firms in the 2015 Survey on average have access to a more reliable power

not shared by those used in the literature (Duflo and Pande, 2007; Kaur, 2014; Sarsons, 2015), which construct rainfall shocks from the long-term mean, and for certain degrees assume normality of the rainfall series. See Appendix B.3 for more details about the spatial interpolation, and computation methods used in this paper.

[13] HydroSHEDS is a derivative of the digital elevation model (DEM) at a three arc-second resolution of the Shuttle Radar Topography Mission (SRTM). The elevation data was void-filled, hydrologically processed, and corrected to produce a consistent, and comprehensive suite of geo-referenced data that enables the analysis of upstream, and downstream connectivity of watersheds. Among the subsets of the HydroSHEDS database, the polygon layers that depict watershed boundaries, and sub-basin delineations at a global scale critical for hydrological analysis are termed HydroBASINS.

supply (regardless of how we measure power outages) and the variation across firms is also lower. An average firm in the 2005 Survey experiences 0.6 outages per month, which typically last 2.51 hours each while an average firm in the 2015 Survey experiences 0.37 outages per month, which typically last 1.46 hours each. The average outage volume is 4.61 hours/month in the 2005 survey and 3.12 hours/month in the 2015 survey. This suggests an improvement in the overall reliability of the power system. The share of firms that own a generator is comparable between the two surveys (34% in 2005 and 35% in 2015).

In terms of our firm performance variables, because the 2015 survey includes a larger share of small and medium firms and a smaller share of large and very large firms, the mean (deflated) values for revenues and inputs are lower than the 2005 survey.[14] An average firm in the 2005 survey pays out 97.4 billion in materials costs and 10.7 billion in labor costs to generate 129 billion in revenues. Meanwhile, an average firm in the 2015 survey pays out 63.7 billion in materials cost and 5.75 billion in labor costs to generate 72.8 billion in revenues. However, productivity is higher for firms in the 2015 survey. TFPR estimated by YKLM model averages 1.89 in 2015 compared to 1.43 in 2005. The time fixed effects in TFPR estimation (see Table A4) confirm a significant increase in the average productivity in nine of eleven manufacturing sector groups except for Paper, printing publishing and the Garment sector.

[14] The change in firm size distribution reflects a more better business environment for small- and medium-sized enterprises (SMEs) to grow and survive and is due in part to a range of policies that help support SMEs such as the privatization of large state-owned enterprises, pro-investment and pro-innovation policies, and greater financial support for SMEs (Pham, Phan and Takayama, 2020).

3 Econometric Stragey

3.1 Empirical Specification

To evaluate the impact of power reliability on the operation of firms, we estimate the following regression:

$$y_{ijpt} = \alpha + \beta Outage_{ijpt} + \Pi' FIRM_{ijpt} + \Sigma' PROVINCE_{pt} + \theta_{jt} + \varepsilon_{ijpt} \tag{1}$$

where i, j, p, and t are subscripts for firm, sector, province, and year, respectively. The dependent variable y_{ijpt} is a measure of the performance or factor inputs (in log form) of firm i in sector j located in province p, and surveyed in year t. Finally, $Outage_{ijpt}$ is a measure of the quality of power supplied to that firm, which is one of three variables: power outage frequency, intensity and volume (in log form). The intercept is α, and β is our parameter of interest. $FIRM_{ijpt}$ is a vector of firm characteristics including size, age, ownership, legal status, foreign trade activities, a measure of financial constraints, and corresponds to the vector of coefficients Π. $PROVINCE_{pt}$ is a vector of province specific variables (in the given year) including the average elevation, cooling degree, rainfall shocks, province industrial product (IP) share, and province IP growth and corresponds to the vector of coefficients Σ. Sector-year dummies are given by θ_{jt}, and ε_{ijpt} is the idiosyncratic error term. Equation 1 is estimated for each individual survey (t = 2005, 2015).

3.2 Instrumental Variables Approach

A major concern with taking an OLS approach in estimating equation 1 is the endogeneity concerns previously discussed. To overcome these challenges we employ an instrumental strategy that integrates information about weather, river flow, hydropower generation, and the electricity grid. More specifically, we take advantage of the fact that the power supply of Vietnam is heavily dependent on hydropower sources, which in turn is largely determined by the weather-induced variations in river flows to hydro electric dams. This kind of exogenous shock is unlikely to affect the performance of firms while it will drive the shift in hydropower supply, a component of electricity supply, once we already control for factors that may give cause for concern like local cooling degree, and rainfall shocks.

3.2.1 River Flow Simulation

At a national scale it is difficult, and prohibitively costly, to obtain discharge data for a large number of stations over a long period. Hence, we use the Soil and Water Assessment Tool (SWAT) (Arnold *et al.*, 1998), one of the most widely used river basin-scale models (Gassman *et al.*, 2014), to simulate river flow. The simulation process follows the approach taken by Nguyen-Tien *et al.* (2018). First, we deploy high-resolution topographic and predefined river network data to delineate the watershed, which is shown in left panel of Figure 1. It is a combination of three large basins with a total area of 977,964 km^2 as defined by the 'FAO Rivers in South, and East Asia' that was selected to take into account the interconnection of Vietnam's rivers with those in upstream countries.[15] The watershed is divided into 7,887 sub-basins at level 12 of HydroBASINS, then broken further into 53,024 terrain units called Hydrologic Response Units (HRU) that are heterogeneous in terms of soil, land cover conditions, and slope. Daily weather data for the watershed (the maximum, and minimum temperature, precipitation, wind speed, relative humidity, and solar radiation) were then added to simulate monthly river flow for the whole watershed for the period from January 1995 to July 2014. The simulation period was chosen to best fit the available performance data of hydropower plants,

[15] Three basins include Red River (165,007 km^2), Vietnam Coast (186,187 km^2), and a part of Mekong River (similar to Lower Mekong River with an area of 626,771 km^2).

subject to the availability of weather data.

3.2.2 Hydropower Generation Model

We use regression analysis to predict hydropower supply at each of the 40 large hydropower dams in Vietnam between 1995 and mid 2014. The combined installed capacity accounts for 75-85% of all hydropower sources, which in turn accounts for 35%-53% of energy generation across Vietnam. This model utilizes installed capacity, a quadratic function of SWAT simulated flow, upstream combined installed capacity, and dam fixed effects as regressors to determine the level of electricity generation.[16] Our model explains 87.8% of the variation in electricity generation at the dam level. From this energy regression we predict the average daily production for each dam by month ($\widehat{Gen}_{ist}$), then calculate the Hydrologically Predicted Plant Factor (HPPF) given by:

$$HPPF_{ist} = \frac{\widehat{Gen}_{ist}\ (MWh/day)}{CAP_{ist}\ (MW) \times 24\ (hours/day)} \times 100\% \tag{2}$$

where i, s, t are indices for dam, month, and year, respectively. $\widehat{Gen}_{ist}$ is the predicted value of dam generation, which excludes the error term that could be correlated with power demand fluctuation. The $HPPF$ is the ratio between predicted generation, and the generation under full utilization at the designed discharge. $HPPF$ of a dam at a given time point is mainly determined by the hydrological conditions at that dam (water availability, and flow extreme degree), which in turn is driven by weather conditions and not affected by human activities or firm performance.

We calculate the HPPF for each year as the average of HPPF for each month in that year:

$$HPPF_{it} = \frac{\sum_{s=1}^{T} HPPF_{ist}}{T} \tag{3}$$

[16] See Appendix C for more details

Where T is the number of months in a year: $T = 12$ for $t = 2004$, and $T = 7$ for $t = 2014$, as we were not able to simulate river-flow series for the last five months of 2014 due to the unavailability of weather data.[17] It should be noted that for each survey we used a different number of hydropower plants to calculate $HPPF$, taking into account the dynamics of dam hydropower construction: 11 for the 2005 survey, and 40 for the 2015 survey (for details see Table Appendix C.1).

3.2.3 Linking Provinces and Hydropower Plants

After estimating hydropower plant factors based on weather shocks, we need to link the performance of power utility companies with the power provided to firms. One solution is to define a cutoff radius (100km, 200km or 400km) to determine the hydopower plant that supplies a certain province before constructing an instrumental variable (Cole *et al.*, 2018). However, this approach is not appropriate for a country like Vietnam that has a single grid system that connects all provinces and power sources. More specifically, since 1994, Vietnam has constructed, and operates a 500kV North - South line that has enabled the interconnection and exchange of electricity across regions. For example, a power shortage in a large industrial centre in the South could be filled by the transfer of a surplus of supply of a power plant located in the North. Hence, in our single grid context, it is important to incorporate the performance of all hydropower plants for each province centre. We consider two factors: the size, and the distance of the utility to our sample of firms. The size of a utility can be proxied by its installed capacity. However, we need to construct a measure of distance. Right panel of Figure 1 shows the provinces in the WBES, the hydropower plants used in this study and the electric grid network that connects them.

[17] The incompleteness of the series however does not affect the IV for 2014 as the first seven months already cover the entire dry season when the variation in river flows to hydropower dams is most likely to affect power reliability. As the last five months are within the rainy season, outages due to hydropower are less of a concern.

A simple solution is to estimate the geodesic distance, the minimum length of a curve that links a province centroid to a hydropower dam along the surface of the earth.[18] A more advanced measure is distance measured through the electricity transmission network based on the grid GIS dataset by WB (2017).[19] For the 2005 survey, the grid-based distance between province centroids and large dams varies between 34 and 1,822 km with a mean of 800 km. For the 2015 survey the distance ranges between 3 and 2,085 km with a mean of 842 km.

To calculate our instrumental variable we incorporate the $HPPF$ from listed hydropower sources that we call the Hydropower Availability Index (HAI) for each province that captures the power supply availability in that province where the HAI is given by:

$$HAI_{jt}(\rho) = \frac{\sum_{i=1}^{N} HPPF_{it} \times w_{ijt}}{\sum_{i=1}^{N} w_{ijt}}, \text{ where } w_{ijt} = \frac{CAP_{it}}{d_{ij}^{\rho}} \tag{4}$$

where the Hydropower Availability Index (HAI) of province j in year t is the weighted average of $HPPF$ of all (N) hydropower plants (indexed by i) within our sample in that year. The weight w_{ijt} is proportionate to the installed capacity of plant i but inversely related to the distance raised to power ρ between plant i, and province j (d_{ijt}^{ρ}).

[18] More specifically, we compute this distance by the *geodist* package by Picard (2017), which uses the coordinates in the WGS 1984 datum, and the equations in Vincenty (1975).

[19] See Appendix B.4 for details).

4 Results

4.1 OLS Estimation

Table 2 presents our estimates of the impact of outages on firm revenues using OLS controlling for firms characteristics, province characteristics, and sector dummies. The coefficients on the variable of interest are negative but insignificant. This suggests that power outages have no effect on firm performance in either 2004 or 2014. Turning to the other dependent variables, in Table 3 we present our OLS estimates of the impact of power outages on productivity (Panel A), firms' use of energy inputs (Panel B), and firms' use of other flexible inputs (Panel C). We include the same controls as Table 2 including sector dummies but they are not reported for reasons of space. Panel A shows that for 2004, power outages resulted in a significant reduction in TFPR estimated by the YKL model. More precisely, a one percent increase in power outage intensity is associated with a reduction in TFPR of 0.06%, while the equivalent reduction for power outage frequency is 0.15%, everything else equal. The estimates are significant at the 10%, and 1% level, respectively. The coefficients for 2014 are also negative but insignificant. When material inputs are taken into account in the TFPR estimation (YKLM model), the impact of power disruption is insignificant in both years.

Panel B shows the results for the use of a series of different factor inputs, where the top panel demonstrates a significant increase in generator use in both years. A 1% increase in the volume of power outages increases the share of self-generated power by 0.09% in 2004, and 0.05% in 2014. Due to the difference in the availability of energy input variables in the two waves of WBES, we report results for two different dependent variable sets in Panel B: energy cost only for the 2005 Survey, and electricity cost and fuel cost for the 2015 Survey.[20] Our OLS results show no significant change

[20]The WBES 2005 survey has limited support documentation other than the questionnaire itself which unfortunately does not include a precise explanation of the 'energy cost' variable. Correspondences with World Bank staff suggests that energy cost in

in energy use. Finally, in Panel C we investigate the relationship between power reliability, and non-power flexible factor utilization. Again, in both years power outages are not found to affect the amount of material inputs or the number of workers employed within a firm.

4.2 IV First Stage

4.2.1 Distance Penalty Parameter Selection

One important aspect in the calculation of our IV is that it may be sensitive to the selection of the distance penalty parameter (ρ). Hence, we need to make a decision on the value of ρ for each survey. To this end we estimated HAI using ten different penalty parameters. Increasing ρ tends to give us a lower minimum and higher maximum values and a larger variation (measured by the standard error). A higher penalty for distance means it is harder to smooth power production across the country, and a more 'localised' measure of our hydrological predictor of power shortages and can take on more extreme values. It is also useful to consider each cross section individually. A firm located in a province with a higher *HAI* is less likely to be constrained by poor power supply conditions everything else equal. However, because in a given year, there is only one transmission and distribution system, there should be only one value of ρ that is the most suitable to describe the system at that time. As we have limited information on the power transmission and distribution system we rely on our WBES sample to choose a suitable value of ρ for each survey.

2005 is likely to include both electricity and fuel costs rather than just the cost of electricity alone.

4.2.2 Instrument Validity

For our IV to be a valid instrument it must be correlated with our outage measures (the relevance condition) but should not drive firm performance via any channel other than through the deficiencies in the power supply (the exogeneity condition). As we intend to use a single hydro-IV for our endogenous variable (power outages), it is not possible to formally test (partial) exogeneity of our instrument. However, as our IV is generated from an arguably exogenous source of variation (the weather), it is difficult to imagine any mechanism through which the hydro-IV could influence firm performance other than through power outage once factors that may give cause for concern in the literature such as cooling degree, and rainfall shocks (Allcott *et al.*, 2016; Alam, 2013) are controlled for. More specifically, our IV generation process ensures that our hydro instruments are not dependent on weather local to the firm but the weather conditions upstream from the dams (and in some cases even beyond Vietnam's borders) that generate electricity supplied to the local region where the firm is located.

In order to investigate the relevance of our instrument we, besides the usual t-tests which are derived from the heteroscedasticity-robust standard errors clustered at the province level, also undertook an underidentification test to check whether the 'reduced form' equation is identified, namely the Sanderson-Windmeijer (SW) first-stage χ^2 test (Sanderson and Windmeijer, 2016).[21] We also consider the problem of weak instruments, which may bias the 2SLS estimates, and make the size of associated tests less reliable (Stock and Yogo, 2005). The widely-used rule of thumb by Staiger and Stock (1997) suggests that an IV should not be used if the first stage F-statistic is less than 10. In our case, to report a first stage F-statistic we rely on the Kleibergen-Paap rank Wald F statistic, which is generalised for the non-i.i.d errors (Kleibergen and Paap, 2006; Kleibergen and Schaffer,

[21] The SW test builds on the procedure by Angrist and Pischke (2008) [p217-218], and implemented using the *ivreg2* command (Baum *et al.*, 2010).

2007). A remedy for a weak IV (if any) is to use a test that is robust to weak instruments for the significance of the potentially endogenous variable. Hence, we use the Anderson-Rubin (AR) test (Anderson and Rubin, 1949) where the null hypothesis is that the coefficients on the endogenous variable in the structural equation (power outage) are not statistically different from zero, and the overidentifying restriction is valid.

Our approach is to run the first stage with the same set of control variables as the OLS estimations, using $\boldsymbol{HAI}(\rho)$ with varying values of ρ as an instrument for the measurement of outages. Hence, to be considered as a possible instrument the variable must be relevant (i.e., correlated with the outage measurements conditional on the control variables) and the correlations should be negative (i.e., less hydro-availability is associated with a greater risk of a power supply shortage). Among the qualified candidates for the choice of IV, for a given set of control variables, we favour those that return the highest values in the first stage Kleibergen-Paap rank Wald F-statistic. For the 'calibration' process we use the values of ρ from 1 to 10 with one unit intervals with revenues (in logs) as the dependent variable (of the second stage).

Figure 2 illustrates the performance of the first stage analysis for different choices of $\boldsymbol{HAI}(\rho)$. In all of the graphs the coefficients of $\boldsymbol{HAI}(\rho)$ in the first stage regression, represented by the red lines, lie below the blue horizontal lines (constant value at 0) and the province-clustered robust confidence interval bands (grey areas) do not touch the blue lines in any of the graphs presented. These indicate that the IVs that we generate in the first stage are within the chosen range of ρ and are negative and relevant. This gives us a certain flexibility when it comes to the selection of IVs for the main regressions results.

Hence, for the 2005 survey we choose $\rho = 1$ for the baseline regressions where the IV is at its most powerful (highest blue bars) regardless of which measure of power outage is used. For the

2015 Survey the IV is strongest when $\rho = 9$ for outage frequency and $\rho = 10$ for outage intensity and volume. As a result, for the firms in the 2015 survey, we choose $\rho = 10$ to generate the IV for our baseline 2SLS regressions. The first stage has allowed us to decide on an optimal value for the distance penalty parameter for the models that use different measures of outages and gives us confidence in the selection process given our prior belief that in each year there should be only one value of ρ that best describes the current system of power transmission and distribution. One possible explanation for the higher value of ρ in 2015 is that more hydropower plants were used in the construction of the IVs (40 vs 11). When the network of power sources became more intensive, the impact of each plant tends to be more localised (i.e., a local utility is more important to local supply) and hence one might expect a higher distance penalty parameter (ρ).

4.2.3 First Stage Estimates

The first panel of Table 4 presents the results of our baseline first stage regressions (where revenue is the dependent variable in the second stage) using our chosen IVs for each survey. The IVs are negatively correlated with our outage measures and the correlations are significant at the 1% level. The exception is the correlation between $\boldsymbol{HAI}(1)$ and outage frequency in the 2005 Survey (Column 1), which is significant only at the 5% level. The first stage (Kleibergen-Paap rk Wald) F-statistic for this column is lower than the rule of thumb (6.08), while those of the five others exceed 10, ranging between 13.29 (Column 4), and 20.12 (Column 2). Hence, our IVs can be considered reasonably powerful. Everything else equal, if the (weighted) hydropower sources that supply a particular firm experience more advantageous hydrological conditions, the firm is less likely to face power constraints across all three measures. Our results for the 2005 Survey show that a one percent increase in $\boldsymbol{HAI}(1)$ corresponds to a 4.34% reduction in outage frequency, a 9.39% decrease in outage intensity, and a

11.65% decrease in outage volume. For the 2015 Survey we find that a one percent increase in $\boldsymbol{HAI}(10)$ corresponds to a 0.71% decrease in outage frequency, a 1.56% decrease in outage intensity, and a 1.76% decrease in outage volume. Under-identification is rejected at the 1% level for each regressions according to the SW χ^2 test, with the exception of Column 1, which is rejected at the 5% level.

Table 4 also provides the first stage results for our other second stage dependent variables (TFPR, and our energy input factors). Due to missing values for a number of these dependent variables, the sample sizes are smaller than those for total revenues. In general, the first stage appears to be consistent across the different samples: The IV coefficients are negative and of a similar magnitude to those shown for revenues and are generally significant at the 1% level with some only at the 5% level. The exceptions are the estimates for Frequency, and when the second stage dependent variable is a measure of TFPR in the 2015 survey. However, the under-identification problem of these first stage regressions are rejected at the 10% level implying the IVs are still valid. For the other first stage results, IV irrelevance is rejected at least at the 5%, and for the majority it is rejected at the 1% level.

4.3 IV Second Stage Results

Table 5 provides a comparison of the impact of outages on revenues. In both years, the 2SLS coefficients are negative, and larger than those of the OLS estimates. The smaller coefficients in the OLS results are in line with the hypothesized attenuation biases caused by measurement error, and biases caused by pro-business environment factors and match what was found for India (Allcott *et al.*, 2016).[22] Our findings for 2004 show that, on average, a 1% increase in power outage frequency,

[22] Omitted environmental factors that support business may increase power demand and put further stress on power supplies, causing more power disruption. This kind of bias pushes the negative impact of power deficiency toward zero.

intensity, and volume is associated with a loss of 0.74%, 0.34%, and 0.27% in revenue although the coefficients remain insignificant, confirmed by the weak instrument test. The coefficients for revenue losses are much larger in 2014, at 1.81%, 0.82%, and 0.73%, respectively and the coefficients are significant at the 1% level for all three outage measures.

The insignificance of an outage impact is also rejected by the weak-instrument test at the 5% level. Overall, the 2SLS estimates support the view of outages in 2004 had very little impact on firm revenues, but, in contrast to the OLS results, they do provide strong evidence that power unreliability in 2014 had a significant impact.

Panel A of Table 5 also shows the 2SLS estimates for the impact of power outages on productivity. Again, the 2SLS estimated coefficients are negative and their magnitudes are greater than those estimated by OLS. In line with the testable hypotheses, the size of the productivity losses are smaller than the revenue losses. In 2004, a 1% increase in the power outages decreases the efficiency of machinery, and labor usage by (YKL model) 0.25 - 0.77%, and the efficiency of machinery, labor, and material input usage (YKLM model) by 0.13 - 0.38%. For 2014 the productivity losses for the same two groups range from 0.55 - 1.36%, and 0.53 - 1.30%, respectively. The significance level of these coefficients varies. The loss of TFPR measured by YKLM in the 2005 Survey is significant at the 10% level. The loss of TFPR measured by the YKL model in the 2015 Survey is significant at the 1% level. The loss of TFPR measured by the YKLM model in Survey 2015 is insignificant according to the t-test but significant at the 5% level according to the weak-instrument test (AR χ^2).

Panel B of Table 5 shows how power deficiency affects a firm's use of energy in the production process. A common adaptation that firms appear to make, and that we find evidence for in both years, is to use self-generated electricity with the coefficients being significant at at least the 5% level. An increase in the degree of power unreliability explains 0.35% - 1.02%, and 0.31% - 0.76% of the increase in the share of self-generated electricity in the 2005 and 2015 Surveys, respectivel, with the magnitude depending on the measure of outages. Of the three measures, firms are more sensitive to the frequency rather than the intensity or volume of outages.

Overall, the higher the level of unreliability in the power supply, the less energy firms use. For the 2005 Survey the magnitude of energy reduction in terms of reduced energy costs, ranges from 0.76 - 2.03, and is significant at the 5% level (t-test) and the 1% level (AR χ^2 test). The same significance levels are found for the use of electricity in the 2015 Survey with the size of the coefficient ranging between -1.03, and -2.46. The impacts of outages on fuel costs in 2015 is found to be negative but not statistically different from zero.

Panel C of Table 5 shows the impact on other flexible inputs (material inputs and labor). For both inputs, there is no impact on firms in the 2005 Survey. Those in the 2015 Survey are only significant at the 10% level according to the AR χ^2 test. Hence, in 2014 a one percent increase in power outages is estimated to decrease the use of material inputs, and labor by 0.62 - 1.39%, and 0.64 - 1.53%, respectively.

The lack of solid results may suggest that firms are not able to adjust factor inputs that easily in the face of power outages, or that the impact is not homogenous across all firms. One mechanism may be the flexibility in the labor which means that if a power cut is unanticipated, firms cannot reduce employment that quickly to save costs when machines and production lines are not operating. When the theoretical model à la Allcott *et al.* (2016) simply predicts a reduction in materials used during machine down times, some firms may adapt by outsourcing energy-intensive materials as was found to be the case for Chinese firms (Fisher-Vanden *et al.*, 2015), which meant that other inputs were effectively substituted by greater use of materials. As Vietnam's supply chains are less developed and complete than China, the outsourcing option is unlikely to be as accessible (or affordable) for a large majority of Vietnamese firms. The mixed responses are one possible reason for the large standard errors attached to the negative coefficients of outage impact on material use.

4.4 Distance Penalty Parameter Uncertainty Analysis

As part of our sensitivity analysis we investigate whether our 2SLS results are sensitive to the selection of the distance penalty parameter (ρ). Figure 3 illustrates how the estimates for the responses of revenue to power unreliability change across different values of ρ. In both years the 2SLS estimates (red dots) are negative and considerably below the OLS estimates no matter which ρ is chosen to generate the IV. The magnitude of the 2SLS coefficients for the 2005 Survey are largest for $\rho = 2$, and gradually decreases as ρ becomes larger. However, the change is negligible in comparison with the standard error of our baseline estimates. As the 95% confidence intervals (grey bands) always cross the green line (zero coefficient) we can conclude that the impact of power outages on revenues in 2004 is insignificant regardless of the value of ρ.

However, when we turn to the 2015 Survey we find our results are sensitive to the choice of ρ. From the baseline values, the coefficients increase when we put a lower penalty (ρ) on distance (the size of hydropower plants is more important), and exceeds one deviation from the baseline values when ρ reaches 1. This implies the role of distance causes a range of coefficient uncertainty. Nevertheless, regardless of the value of ρ, the coefficients for the 2015 Survey are always negative and significant (at the 5% level or better), and their magnitudes are always larger than those for the 2005 Survey. In Appendix D we provide additional robustness checks to show that our findings are robust to the use of alternative IVs or when regressions are run on subsets of the data.

4.5 Differences across Years

Our robustness checks reinforce our result that power unreliability was more harmful (both in terms of size and significance) for firm performance in the 2015 Survey rather than the 2005 Survey. In this subsection we investigate the mechanisms that may be driving this difference. One possible explanation is that the extent to which firms require electricity to operate has intensified. If electricity is mainly used for general purposes, such as lighting, or makes only a minimal contribution to the value-added processes, there is no reason to expect a significant impact of a power outage on firm performance. However, if firms have become more technologically advanced and now operate complex production processes then it is easy to imagine how power outages could become more costly.

To investigate whether electricity intensity changes after power outages, we pool all manufacturing firms (that report a value of machine use) and test whether there were changes in factor intensities across the two surveys. We employ a median regressor to estimate the three log-linear versions of the Cobb-Douglas function: YKL, YKLM, and YKLMN, where YKLMN is an extension of YKLM, adding energy (N) to the production function. We assume that the energy cost reported in the 2005 Survey is the sum of electricity costs and fuel costs. All regressions include the factor costs of production (in log form), their interaction terms with an indicator for year 2015, and a year dummy. For robust checks, we sequentially include province and sector dummies. The results are presented in Table 6.

Of primary interest are the interaction terms that indicate whether factor intensities change across the two surveys. All twelve regressions confirm that there was a significant rise in capital intensity, and the final four columns show that there has also been a significant increase in energy intensity. The magnitude of the shift is relatively large. The results in the final column show that in the

2005 Survey, the intensity of capital is just 2.3% whereas it almost five-folds to 11.5% in the 2015 Survey. Over the same period the average energy intensity almost doubles from 5.8% to 10.2%. In contrast, material input intensity significantly decreases by 13%. Labour intensity also falls, however the significance of the reduction is not robust, especially when taking material inputs and energy into account.

Assuming the firms in the WBES are representative, our results suggest that over the period 2004 to 2014 there was substantial development in the Vietnam manufacturing sector. Firms engaged in more comprehensive and complicated manufacturing processes that use fewer intermediate material inputs, generate more value from existing capital and energy inputs through the more intensive use of machinery, and hence are less reliant on simple product assembly. It is also likely, although we cannot test it directly, that firms became more electricity intensive. This would explain Vietnam's increasing sensitivity to changes in the quality of electricity provision.

The increase in energy intensity of manufacturing firms estimated in a Vietnam context is in line with previous studies that confirm positive links between energy consumption and industrialisation (Sadorsky, 2013; Samouilidis and Mitropoulos, 1984), trade liberalisation (Cole, 2006) and economic complexity (Liu *et al.*, 2020). During the period between the two WBESs, Vietnam continued its policy of pursuing export-oriented industrialisation while maintaining a subsidised electricity tariff for industry with a price that is relatively low compared to neighbouring countries (ADB, 2015). Vietnam joined the World Trade Organisation (WTO) in 2007 and concluded or negotiated trade deals with Japan, Korea, the EU and important trading and investment partners within the Trans-Pacific Partnership (TPP, now known as CPTPP).[23]). The development of a strong industrial sector benefited from a higher level of economic openness and Vietnam's deeper integration into the global

[23] The Comprehensive and Progressive Agreement for Trans-Pacific Partnership.

economy (Nguyen *et al.*, 2016). As a result, Vietnam emerged as an Asian manufacturing powerhouse (Bloomberg, 2015). Thanks to trade liberalisation, following accession to the WTO, Vietnam's export structure became more sophisticated and closer to that of Thailand rather than Indonesia and the Philippines to which Vietnam used to compare itself (Nguyen, 2016). Between 2004 and 2014, Vietnam moved from 85th to 54th place in the global economic complexity index ranking (The Growth Lab at Harvard University, 2020). Such a move up the ranking suggests that Vietnam's managed to improve its competitiveness as it diversified its export basket by producing and exporting more sophisticated products. More specifically, exports have been modernised with a transition away from primary commodities such as crude oil and rice to manufactured goods that require low and medium technological inputs such as apparel and footwear as well as high technology goods such as electronics, mobile phones, electronic tablets, and incorporated circuits (WB and MPI, 2016; Pham, Hollweg, Mtonya, Winkler, Nguyen and Nguyen, 2020).

5 Conclusions

In this paper we investigated the impact of power outages on firm performance using an innovative instrumental variable strategy. More specifically, we contribute to the hydro-IV literature by offering an alternative approach for a single-grid country. More specifically, we integrate a rainfall-runoff model with a hydropower generation model and employ an interpolation technique using grid-based distances to generate an appropriate instrument to deal with the endogeneity of power outages in their impact on firm performance. We apply our methodological approach to firm-level data for Vietnam for two waves of the World Bank Enterprise Surveys spanning ten years, allowing us to identify and compare the causal effect in the context of early and late stage development of electricity infrastructure. Our methodological contribution can be applied to study the impact of power outages on firm performance or household welfare in other contexts with hydropower dependency.

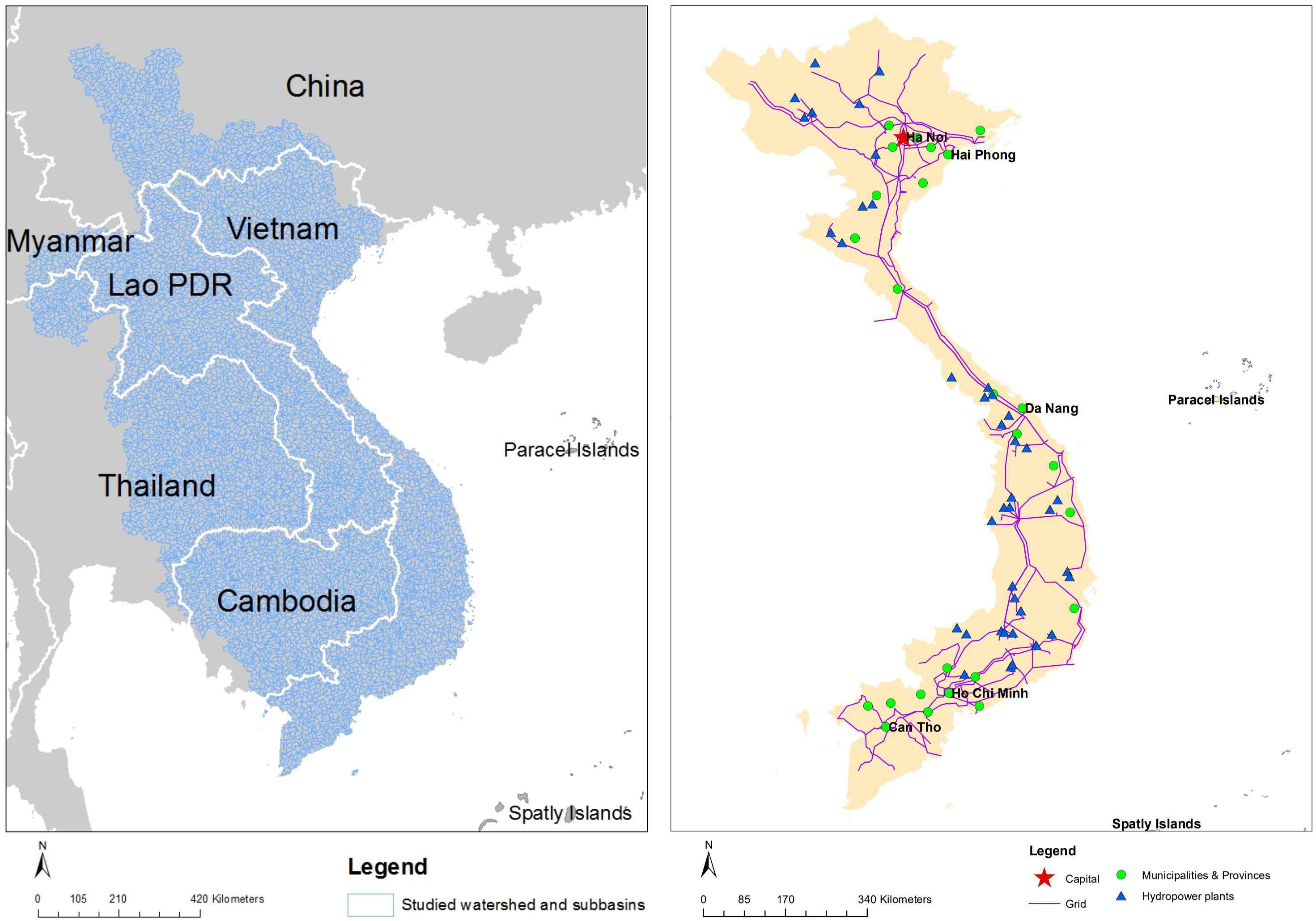
Our results suggest that, despite an overall improvement in access to power and better reliability, in a rapidly growing economy like Vietnam firms have become increasingly reliant on electricity and, hence, any power disruption incurs greater financial costs. As expected, we also find that firms that face frequent or long lasting power disruptions tend to generate less revenue than firms that do not experience outages. Moreover, our results show that firms that suffer power outages have lower productivity and use less flexible inputs, such as material inputs and labor. Power outages additionally encourage firms to use backup generators to supply more of their electricity.

Of the different types of outages, increased frequency appears to be more damaging than the intensity of individual events. Comparing the impact of power outages in 2004 and 2014, our findings indicate that firm performance appears to be less responsive to an unreliable power supply in 2004 and perhaps not surprisingly became more sensitive as firms became more dependent on electricity. This finding supports a progressive approach to power development policy making. Our results show that an increase in the production complexity of the manufacturing process for the average firm helps to explain the differences in our empirical results compared to previous studies, where the reduction in revenues due to power outages are found to be significant in some (Allcott *et al.*, 2016) but insignificant in others (Mensah, 2016).

In terms of policy prescriptions, our findings imply that an improvement in power reliability could substantially enhance economic growth in Vietnam. According to our 2SLS estimates, in 2014 a small reduction in power disruption by 1% would have increased revenues by 0.73%. Given that total revenues of registered firms in 2014 was 13,516 trillion dong (639 billion USD), and assuming that firms in the survey we used are representative, such an improvement in power quality could have added 4.66 billion USD to the performance of Vietnam's firms.[24] As Vietnam has been attempting to exploit its untapped potentials for low-carbon solar and wind power (The Diplomat, 2020), it is critical for the country to develop energy storage technology and smart grid network to address the intermittency of these new power sources and ensure the stability of the power system.

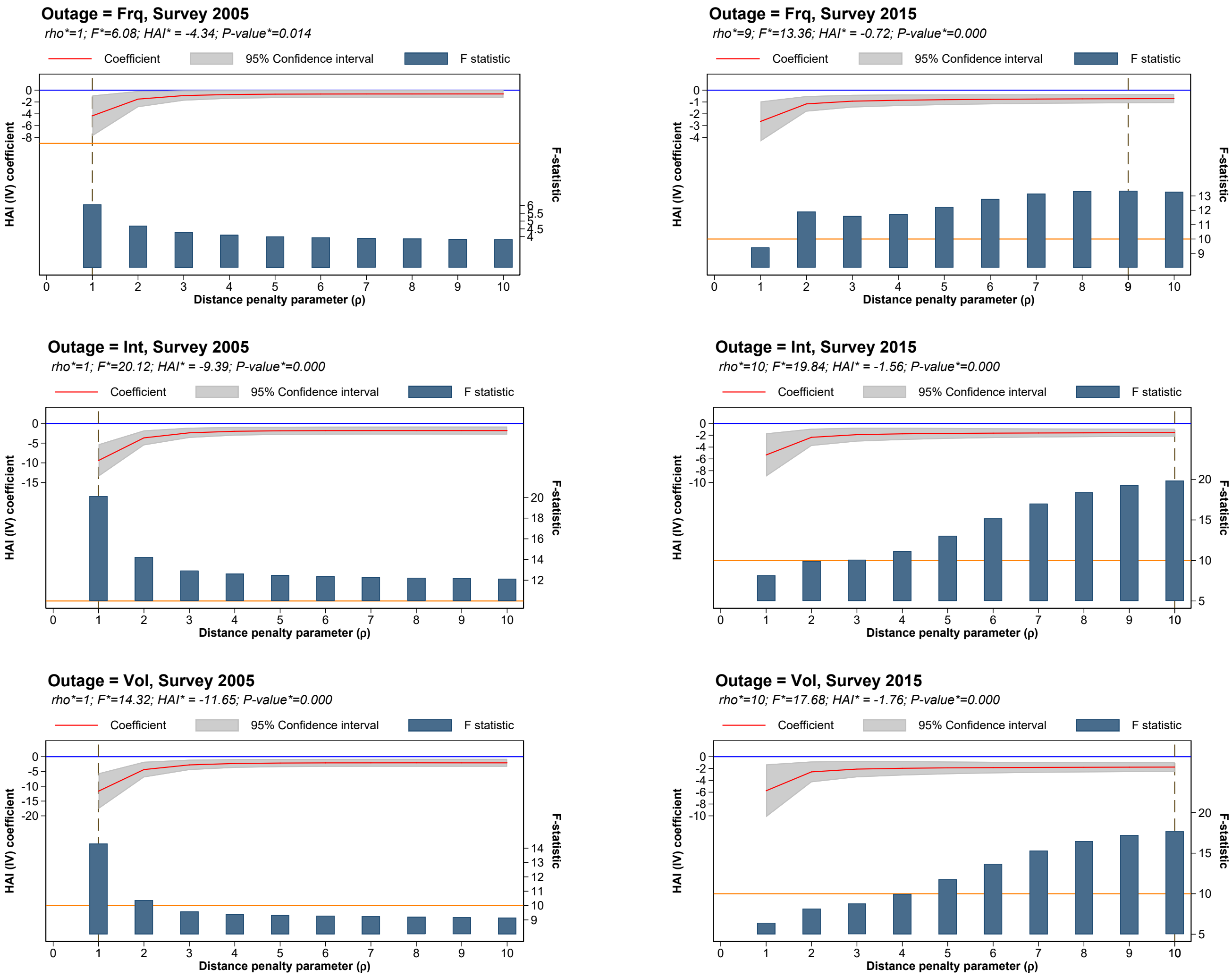
[24]Total revenues of registered firms is from Vietnam General Statistics Office (GSO). The average official exchange rate for 2014 (1USD= 21,148 VND) is from International Monetary Fund, International Financial Statistics.

Figure 1: River Basins, Provinces, Hydropower plants and Electric Grids in the study



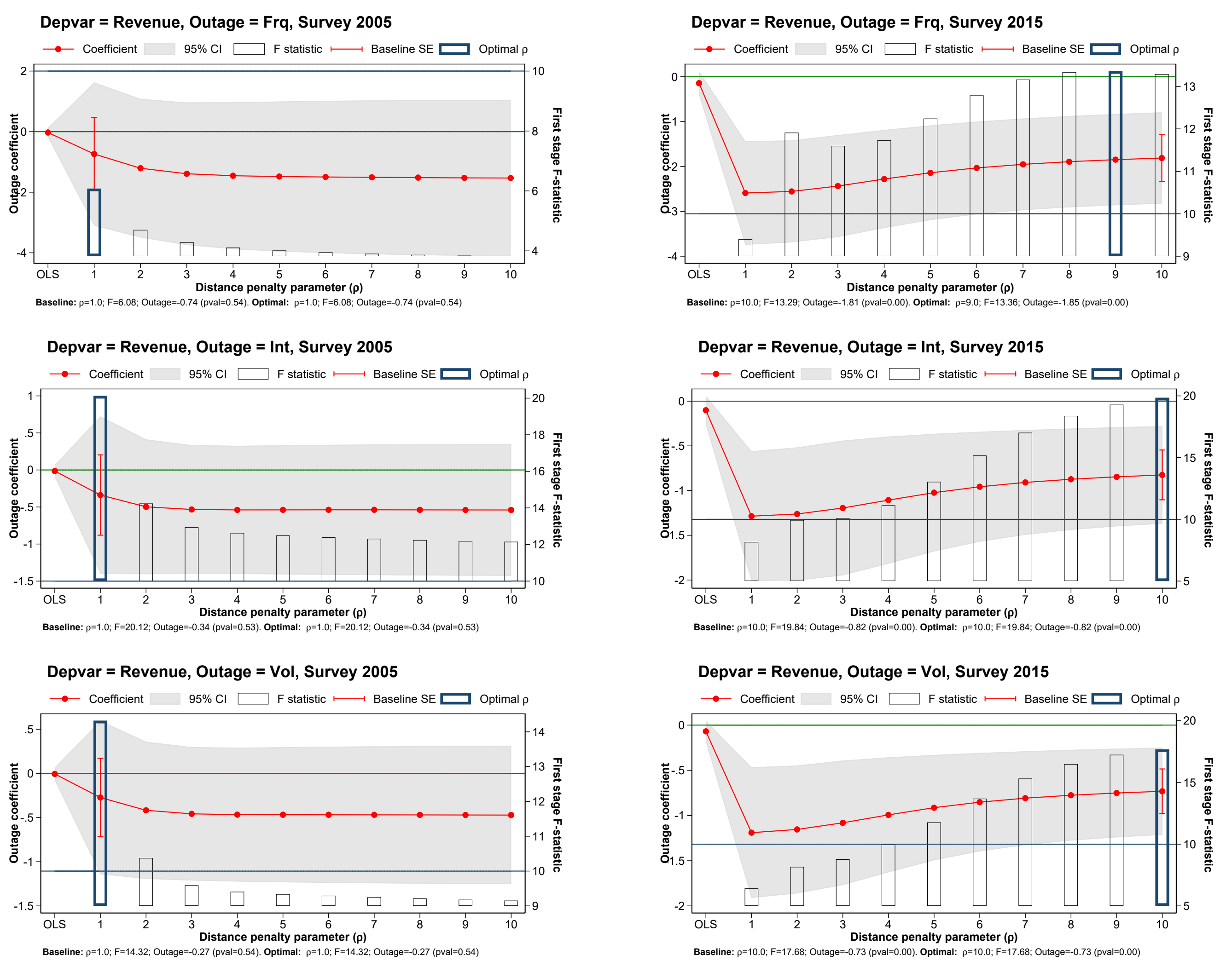
Source: The watershed used to generate river was compiled from HydroSHEDS, HYdroBASINS, and "FAO Rivers in South and East Asia" MONRE (2012). It is shared by Vietnam, China, Myanmar, Lao PDR, Thailand and Cambodia. The administrative map is derived from GAUL dataset (FAO, 2015). The electric line network is from (WB, 2017).

Figure 2: First stage performance using *HAI*(ρ) as the IV and varying the distance penalty parameter (ρ)



Note: (a) Each graph draws the estimate of coefficient of *HAI* (line, left y-axis), its 95% confidence interval (area, left y-axis) and the F-statistic (bar, right y-axis) of the first stage regression for each value of ρ within the stated range. (b) The vertically dashed line is drawn at the optimal value of ρ, the horizontal line at 0 (blue, left y-axis) is drawn with reference to the confidence intervals to identify the significance of the coefficients, while the horizontal line at 10 (orange, right y-axis) is sketched with reference to the F-statistic to identify the "weak" instrument issue, based on the rough rule of thumb (F-statistic <10). (c) F statistic reported is for the heteroskedasticity Kleibergen-Paap weak instrument test.

Figure 3: Distance penalty parameter uncertainty analysis



Note: (a) Each graph draws the estimate of coefficient power outage impact on revenue (red line, left y-axis), its 95% confidence interval (area, left y-axis) and the F-statistic (bar, right y-axis) of the first stage regression for each value of ρ within the stated range. (b) The vertical spikes indicate the standard errors of the baseline IV estimates at the optimal values of ρ, the horizontal line at 0 (green, left y-axis) is drawn with reference to the confidence intervals to identify the significance of the coefficients, while the horizontal line at 10 (blue, right y-axis) is sketched with reference to the F-statistic to identify the "weak" instrument issue, based on the rough rule of thumb (F-statistic <10). (c) F statistic reported is for the heteroskedasticity Kleibergen-Paap weak instrument test.

Table 1: Summary statistics

	Survey 2005					Survey 2015				
VARIABLES	N	mean	sd	min	max	N	mean	sd	min	max
FIRM LEVEL DATA										
Power provision										
Power outages; frequency (occurences per month)	1,150	0.60	1.20	0	20	960	0.37	1.19	0	20
Power outages; intensity (hours per occurence)	1,149	2.51	8.56	0	96	960	1.46	5.82	0	96
Power outages; volume (hours per month)	1,149	4.61	20.7	0	384	960	3.12	18.7	0	480
Generator ownership or share; binary	1,131	0.34	0.47	0	1	684	0.35	0.48	0	1
Firm characteristics										
Firm age	1,147	12.5	13.2	1	115	989	13.3	10.3	1	113
Small size	1,149	0.10	0.30	0	1	989	0.38	0.49	0	1
Medium size	1,149	0.37	0.48	0	1	989	0.35	0.48	0	1
Large size	1,149	0.26	0.44	0	1	989	0.17	0.38	0	1
Very large size	1,149	0.27	0.45	0	1	989	0.098	0.30	0	1
State ownership 10%-50%	1,145	0.14	0.34	0	1	994	0.020	0.14	0	1
State ownership > 50%	1,145	0.20	0.40	0	1	994	0.015	0.12	0	1
Foreign ownership 10%-50%	1,145	0.017	0.13	0	1	993	0.013	0.11	0	1
Foreign ownership > 50%	1,145	0.100	0.30	0	1	993	0.065	0.25	0	1
Share-holding	1,149	0.37	0.48	0	1	989	0.22	0.41	0	1
Share-traded	1,149	0.0035	0.059	0	1	989	0.039	0.19	0	1
Exporter	1,150	0.46	0.50	0	1	994	0.27	0.45	0	1
Access to credit	1,150	0.70	0.46	0	1	971	0.49	0.50	0	1
Firm performance										
Revenues (bil. VND)	1,141	129	333	0.16	5,271	975	72.8	228	0.034	4,108
TFPR YKL model; log	986	2.34	1.23	-1.76	6.57	432	2.34	1.37	-1.35	7.57
TFPR YKLM model; log	975	1.43	0.84	-1.52	3.92	401	1.89	1.10	-0.44	5.85
Energy use										
Generator use; % electricity	1,045	1.37	3.51	0	25	637	0.37	0.80	0	7
Energy cost (mil. VND)	1,146	2,854	9,659	1.14	128,264					
Fuel cost (mil. VND)						476	4,271	40,972	0.68	753,043
Electricity cost (mil. VND)						892	681	2,387	0.68	34,229
Other factor use										
Material cost (bil. VND)	1,134	97.4	299	0.036	5,808	552	63.7	447	0.0034	9,844
Labor cost (bil. VND)	1,139	10.7	23.0	0.023	324	899	5.75	16.2	0.010	222
PROVINCE LEVEL DATA										
Province characteristics										
Elevation (m)	24	128	154	1.94	464	19	119	142	1.94	397
Rainfall shocks (SPI)	24	0.080	0.63	-1.35	0.76	19	-0.35	0.38	-0.95	0.46
Cooling degree (F)	24	11.9	2.95	7.20	15.8	19	11.6	2.60	7.90	15.4
Province IP share	24	0.033	0.058	0.0016	0.25	19	0.042	0.049	0.0051	0.18
Province IP Index	24	1.18	0.053	1.08	1.32	19	1.07	0.057	0.88	1.15

Note: The mean of volume measure does not necessarily equal the product of the means of frequency and intensity as a sum of products differ from a product of sums.

Table 2: The impact of outages on firm revenue (OLS estimates)

	Survey 2005			Survey 2015		
VARIABLES	*Frq*	*Int*	*Vol*	*Frq*	*Int*	*Vol*
	(1)	(2)	(3)	(4)	(5)	(6)
Power outage; log	-0.03	-0.01	-0.01	-0.14	-0.10	-0.07
	(0.07)	(0.04)	(0.04)	(0.13)	(0.08)	(0.06)
Age; log	0.09**	0.09**	0.09**	0.05	0.05	0.05
	(0.04)	(0.04)	(0.04)	(0.08)	(0.08)	(0.08)
Medium size	0.79***	0.78***	0.78***	1.50***	1.50***	1.50***
	(0.13)	(0.13)	(0.13)	(0.10)	(0.10)	(0.10)
Large size	1.97***	1.96***	1.97***	2.70***	2.71***	2.70***
	(0.18)	(0.18)	(0.18)	(0.20)	(0.20)	(0.20)
Very large size	3.13***	3.13***	3.13***	3.89***	3.90***	3.89***
	(0.22)	(0.22)	(0.22)	(0.20)	(0.19)	(0.20)
State ownership 10%-50%	0.28**	0.28**	0.29**	-0.09	-0.09	-0.09
	(0.11)	(0.11)	(0.11)	(0.32)	(0.32)	(0.32)
State ownership > 50%	0.66***	0.66***	0.66***	0.50*	0.52*	0.51*
	(0.15)	(0.15)	(0.15)	(0.25)	(0.25)	(0.25)
Foreign ownership 10%-50%	0.70**	0.70**	0.70**	0.23	0.22	0.23
	(0.25)	(0.25)	(0.25)	(0.33)	(0.33)	(0.33)
Foreign ownership > 50%	0.73***	0.73***	0.73***	0.49***	0.49***	0.49***
	(0.23)	(0.23)	(0.23)	(0.12)	(0.12)	(0.12)
Share-holding	0.20**	0.20**	0.20**	0.39**	0.38**	0.38**
	(0.09)	(0.09)	(0.09)	(0.16)	(0.16)	(0.16)
Share-traded	0.71**	0.71**	0.72**	0.45*	0.44*	0.44*
	(0.28)	(0.28)	(0.28)	(0.23)	(0.22)	(0.22)
Exporter	0.20***	0.20***	0.20***	-0.14*	-0.14*	-0.14*
	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.08)
Access to credit	0.55***	0.55***	0.55***	0.06	0.05	0.05
	(0.07)	(0.08)	(0.08)	(0.16)	(0.16)	(0.16)
IP share	0.84	0.85	0.85	7.03***	7.04***	7.03***
	(0.72)	(0.73)	(0.73)	(1.26)	(1.26)	(1.27)
IP index	-0.21	-0.20	-0.20	1.94	1.93	1.94
	(0.89)	(0.90)	(0.90)	(1.32)	(1.28)	(1.29)
Cooling degree; log	0.39	0.39	0.39	0.03	0.03	0.03
	(0.32)	(0.32)	(0.32)	(0.14)	(0.13)	(0.14)
Elevation (km); log	-0.10**	-0.10**	-0.09**	0.04	0.05	0.04
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Rainfall shocks (SPI)	0.04	0.04	0.04	0.03	0.02	0.02
	(0.09)	(0.09)	(0.09)	(0.13)	(0.12)	(0.12)
R-squared	0.60	0.60	0.60	0.48	0.48	0.48
Observations	1,133	1,132	1,132	915	915	915

Notes: (a) Regression includes dummies for sectors. (b) Robust standard errors clustered at province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. (c) Three measures of power outages (in log form) indicated in each column name: *Frq* - Frequency (average number of outage per month); *Int* - Intensity (average duration (hour) per outage); and *Vol* - Volume (average outage hour per month = *Frq*× *Int*).

Table 3: Impact of power outages on TFP and other inputs (OLS estimates)

	Survey 2005			Survey 2015		
VARIABLES	Frq	Int	Vol	Frq	Int	Vol
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: TFPR			
TFP YKL model; log						
Power outage; log	-0.15***	-0.06*	-0.05**	-0.12	-0.04	-0.03
	(0.05)	(0.03)	(0.02)	(0.16)	(0.09)	(0.07)
R-squared	0.07	0.07	0.07	0.10	0.10	0.10
Observations	978	977	977	421	421	421
TFP YKLM model; log						
Power outage; log	-0.03	-0.01	-0.01	-0.01	0.01	0.00
	(0.02)	(0.01)	(0.01)	(0.09)	(0.05)	(0.04)
R-squared	0.02	0.02	0.02	0.06	0.06	0.06
Observations	967	966	966	392	392	392
			Panel B: Energy inputs			
Generator use; log						
Power outage; log	0.17***	0.08***	0.09***	0.11***	0.08***	0.05***
	(0.06)	(0.02)	(0.02)	(0.04)	(0.02)	(0.01)
R-squared	0.10	0.10	0.11	0.16	0.18	0.17
Observations	1,036	1,035	1,035	605	605	605
Energy cost; log						
Power outage; log	0.17	0.00	0.01			
	(0.12)	(0.05)	(0.04)			
R-squared	0.37	0.36	0.36			
Observations	1,137	1,136	1,136			
Electric cost; log						
Power outage; log				0.11	0.06	0.06
				(0.10)	(0.04)	(0.03)
R-squared				0.40	0.40	0.40
Observations				841	841	841
Fuel cost; log						
Power outage; log				-0.26	-0.08	-0.09
				(0.21)	(0.12)	(0.09)
R-squared				0.31	0.31	0.31
Observations				458	458	458
			Panel C: Other inputs			
Material cost; log						
Power outage; log	-0.03	-0.02	-0.01	-0.19	-0.14	-0.09
	(0.10)	(0.05)	(0.04)	(0.19)	(0.13)	(0.09)
R-squared	0.50	0.50	0.50	0.40	0.40	0.40

(cont.)

	Survey 2005			Survey 2015		
VARIABLES	Frq	Int	Vol	Frq	Int	Vol
	(1)	(2)	(3)	(4)	(5)	(6)
Observations	1,125	1,124	1,124	529	529	529
Labor cost; log						
Power outage; log	0.04	0.01	0.01	0.02	-0.00	-0.00
	(0.05)	(0.02)	(0.02)	(0.07)	(0.03)	(0.03)
R-squared	0.76	0.76	0.76	0.57	0.57	0.57
Observations	1,130	1,129	1,129	845	845	845

Notes: (a) Regressions include control variables as those in Table 2 (including sector dummies).(b) Dependent variables are in bold. Three measures of power outages (in log form) indicated in each column name are Frq - Frequency (average number of outage per month); Int - Intensity (average duration (hour) per outage); and Vol - Volume (average outage hour per month = Frq× Int). Robust standard errors clustered at province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: First stage

VARIABLES	Survey 2005, ρ = 1			Survey 2015, ρ = 10		
	Frq	Int	Vol	Frq	Int	Vol
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Performance						
Revenues; log						
HAI(ρ)	-4.34**	-9.39***	-11.65***	-0.71***	-1.56***	-1.76***
	(1.76)	(2.09)	(3.08)	(0.19)	(0.35)	(0.42)
Observations	1,133	1,132	1,132	915	915	915
Clusters	24	24	24	19	19	19
First stage F-statistic	6.08	20.12	14.32	13.29	19.84	17.68
SW χ^2 test	6.56**	21.73***	15.46***	14.69***	21.93***	19.54***
TFP YKL model; log						
HAI(ρ); IV	-3.83**	-9.56***	-11.53***	-0.62*	-1.46**	-1.53*
	(1.78)	(2.67)	(3.72)	(0.35)	(0.61)	(0.80)
Observations	978	977	977	421	421	421
Clusters	24	24	24	19	19	19
F-statistic	4.61	12.85	9.59	3.15	5.68	3.62
SW χ^2 test	4.98**	13.88***	10.36***	3.61*	6.50**	4.15**
TFP YKLM model; log						
HAI(ρ); IV	-3.87**	-9.52***	-11.54***	-0.61	-1.44**	-1.50
	(1.82)	(2.78)	(3.84)	(0.40)	(0.68)	(0.91)
Observations	967	966	966	392	392	392
Clusters	24	24	24	19	19	19
F-statistic	4.54	11.69	9.01	2.25	4.45	2.71
SW χ^2 test	4.90**	12.63***	9.73***	2.60	5.13**	3.13*
Panel B: Energy inputs						
Generator use; log						
HAI(ρ); IV	-4.49**	-10.61***	-12.88***	-0.75***	-1.71***	-1.86***
	(1.84)	(2.75)	(3.82)	(0.26)	(0.46)	(0.58)
Observations	1,036	1,035	1,035	605	605	605
Clusters	24	24	24	19	19	19
F-statistic	5.95	14.85	11.38	8.20	13.67	10.14
SW χ^2 test	6.42**	16.02***	12.28***	9.17***	15.28***	11.34***
Energy cost; log						
HAI(ρ); IV	-4.53***	-10.04***	-12.19***			
	(1.73)	(2.00)	(3.00)			
Observations	1,137	1,136	1,136			
Clusters	24	24	24			
F-statistic	6.82	25.16	16.47			
SW χ^2 test	7.36***	27.16***	17.78***			
Electric cost; log						
HAI(ρ); IV				-0.78***	-1.62***	-1.85***
				(0.19)	(0.35)	(0.42)

(cont.)

VARIABLES	Survey 2005, ρ = 1			Survey 2015, ρ = 10		
	Frq	Int	Vol	Frq	Int	Vol
	(1)	(2)	(3)	(4)	(5)	(6)
Observations				841	841	841
Clusters				19	19	19
F-statistic				17.13	20.89	19.32
SW χ^2 test				19.01***	23.18***	21.44***
Fuel cost; log						
HAI(ρ); IV				-1.16***	-2.17***	-2.42***
				(0.26)	(0.52)	(0.66)
Observations				458	458	458
Clusters				19	19	19
F-statistic				20.15	17.52	13.59
SW χ^2 test				23.09***	20.07***	15.57***

				Panel C: Other inputs		
Material cost; log						
HAI(ρ); IV	-4.25**	-9.32***	-11.52***	-0.83***	-1.72***	-1.87***
	(1.72)	(2.15)	(3.12)	(0.27)	(0.48)	(0.63)
Observations	1,125	1,124	1,124	529	529	529
Clusters	24	24	24	19	19	19
F-statistic	6.12	18.83	13.63	9.29	13.03	8.85
SW χ^2 test	6.61**	20.34***	14.72***	10.53***	14.76***	10.03***
Labor cost; log						
HAI(ρ); IV	-4.20**	-9.28***	-11.37***	-0.68***	-1.43***	-1.62***
	(1.73)	(2.11)	(3.10)	(0.19)	(0.36)	(0.43)
Observations	1,130	1,129	1,129	845	845	845
Clusters	24	24	24	19	19	19
F-statistic	5.90	19.32	13.45	12.41	16.07	14.10
SW χ^2 test	6.37**	20.86***	14.52***	13.77***	17.82***	15.64***

Notes:(a) Table reports results for first stage of 2SLS with control variables as in Table 2 and dummies for sectors. (b) Dependent variable in the second stage is in italics and are: Frq - Frequency (average number of outage per month); Int - Intensity (average duration (hour) per outage); or Vol - Volume (average outage hour per month = Frq× Int). (c) First stage F-statistic is Kleibergen-Paap rk Wald F-statistic robust to heteroskedasticity. (d) SW χ^2 test is Sanderson-Windmeijer first stage χ^2 test for underidentification. (e) Robust standard errors clustered at province level in parentheses. (f) *** p<0.01, ** p<0.05, * p<0.1.

Table 5: 2SLS estimates

VARIABLES	Survey 2005, ρ = 1 Frq (1)	 Int (2)	 Vol (3)	Survey 2015, ρ = 10 Frq (4)	 Int (5)	 Vol (6)
			Panel A: Outputs			
Revenues; log						
Outage; 2SLS	-0.74	-0.34	-0.27	-1.81***	-0.82***	-0.73***
	(1.21)	(0.54)	(0.44)	(0.52)	(0.28)	(0.25)
Outage; OLS	-0.03	-0.01	-0.01	-0.14	-0.10	-0.07
	(0.07)	(0.04)	(0.04)	(0.13)	(0.08)	(0.06)
Observations	1,133	1,132	1,132	915	915	915
First stage F-statistic	6.08	20.12	14.32	13.29	19.84	17.68
SW χ^2 test	6.56**	21.73***	15.46***	14.69***	21.93***	19.54***
AR χ^2 test	0.46	0.45	0.45	5.67**	5.67**	5.67**
TFP YKL model; log						
Outage; 2SLS	-0.77	-0.31	-0.25	-1.36***	-0.57**	-0.55**
	(0.67)	(0.25)	(0.22)	(0.47)	(0.25)	(0.23)
Outage; OLS	-0.15***	-0.06*	-0.05**	-0.12	-0.04	-0.03
	(0.05)	(0.03)	(0.02)	(0.16)	(0.09)	(0.07)
Observations	978	977	977	421	421	421
First stage F-statistic	4.61	12.85	9.59	3.15	5.68	3.62
SW χ^2 test	4.98**	13.88***	10.36***	3.61*	6.50**	4.15**
AR χ^2 test	2.59	2.50	2.50	3.97**	3.97**	3.97**
TFP YKLM model; log						
Outage; 2SLS	-0.38	-0.15*	-0.13*	-1.30	-0.55	-0.53
	(0.25)	(0.08)	(0.07)	(1.03)	(0.38)	(0.41)
Outage; OLS	-0.03	-0.01	-0.01	-0.01	0.01	0
	(0.02)	(0.01)	(0.01)	(0.09)	(0.05)	(0.04)
Observations	967	966	966	392	392	392
First stage F-statistic	4.54	11.69	9.01	2.25	4.45	2.71
SW χ^2 test	4.90**	12.63***	9.73***	2.60	5.13**	3.13*
AR χ^2 test	3.65*	3.66*	3.66*	4.66**	4.66**	4.66**
			Panel B: Energy inputs			
Generator use; log						
Outage; 2SLS	1.02***	0.43***	0.35***	0.76***	0.33***	0.31***
	(0.33)	(0.14)	(0.11)	(0.21)	(0.10)	(0.10)
Outage; OLS	0.17***	0.08***	0.09***	0.11***	0.08***	0.05***
	(0.06)	(0.02)	(0.02)	(0.04)	(0.02)	(0.01)
Observations	1,036	1,035	1,035	605	605	605
First stage F-statistic	5.95	14.85	11.38	8.20	13.67	10.14
SW χ^2 test	6.42**	16.02***	12.28***	9.17***	15.28***	11.34***
AR χ^2 test	8.16***	8.26***	8.26***	6.49**	6.49**	6.49**
Energy cost; log						
Outage; 2SLS	-2.03*	-0.92**	-0.76**			
	(1.04)	(0.38)	(0.33)			

(cont.)

VARIABLES	Survey 2005, $\rho = 1$ Frq (1)	Survey 2005, $\rho = 1$ Int (2)	Survey 2005, $\rho = 1$ Vol (3)	Survey 2015, $\rho = 10$ Frq (4)	Survey 2015, $\rho = 10$ Int (5)	Survey 2015, $\rho = 10$ Vol (6)
Outage; OLS	0.17	0	0.01			
	(0.12)	(0.05)	(0.04)			
Observations	1,137	1,136	1,136			
First stage F-statistic	6.82	25.16	16.47			
SW χ^2 test	7.36***	27.16***	17.78***			
AR χ^2 test	7.40***	7.63***	7.63***			
Electric cost; log						
Outage; 2SLS				-2.46**	-1.18**	-1.03**
				(1.14)	(0.49)	(0.43)
Outage; OLS				0.11	0.06	0.06
				(0.10)	(0.04)	(0.03)
Observations				841	841	841
First stage F-statistic				17.13	20.89	19.32
SW χ^2 test				19.01***	23.18***	21.44***
AR χ^2 test				9.31***	9.31***	9.31***
Fuel cost; log						
Outage; 2SLS				-1.31	-0.70	-0.62
				(1.16)	(0.59)	(0.55)
Outage; OLS				-0.26	-0.08	-0.09
				(0.21)	(0.12)	(0.09)
Observations				458	458	458
First stage F-statistic				20.15	17.52	13.59
SW χ^2 test				23.09***	20.07***	15.57***
AR χ^2 test				1.78	1.78	1.78
Panel C: Other inputs						
Material cost; log						
Outage; 2SLS	-0.75	-0.34	-0.27	-1.39	-0.67	-0.62
	(1.89)	(0.85)	(0.70)	(0.88)	(0.42)	(0.40)
Outage; OLS	-0.03	-0.02	-0.01	-0.19	-0.14	-0.09
	(0.10)	(0.05)	(0.04)	(0.19)	(0.13)	(0.09)
Observations	1,125	1,124	1,124	529	529	529
First stage F-statistic	6.12	18.83	13.63	9.29	13.03	8.85
SW χ^2 test	6.61**	20.34***	14.72***	10.53***	14.76***	10.03***
AR χ^2 test	0.18	0.17	0.17	3.34*	3.34*	3.34*
Labor cost; log						
Outage; 2SLS	0.10	0.05	0.04	-1.53	-0.73	-0.64
	(0.50)	(0.23)	(0.18)	(1.28)	(0.58)	(0.52)
Outage; OLS	0.04	0.01	0.01	0.02	0	0
	(0.05)	(0.02)	(0.02)	(0.07)	(0.03)	(0.03)
Observations	1,130	1,129	1,129	845	845	845
First stage F-statistic	5.90	19.32	13.45	12.41	16.07	14.10
SW χ^2 test	6.37**	20.86***	14.52***	13.77***	17.82***	15.64***
AR χ^2 test	0.04	0.04	0.04	2.75*	2.75*	2.75*

Notes: (a) Regression includes control variables as in Table 2 and dummies for sectors. (b) Endogenous variable is measure of power outages: Frq - Frequency (average number of outage per month); Int - Intensity (average duration (hour) per outage); or Vol - Volume (average outage hour per month = Frq× Int). (c) Excluded IV is *HAI*(ρ) with predetermined values of ρ for each survey. (d) OLS estimate included for comparison. (e) First stage F-statistic: Kleibergen-Paap rk Wald F-statistic that is robust to heteroskedasticity, Underidentification test: SW χ^2 test (Sanderson-Windmeijer first stage χ^2 test), Weak-instrument robust tests for significant endogenous variables: AR χ^2 test (Anderson-Rubin Wald χ^2 test). (f) Robust standard errors clustered at province level in parentheses. (g) *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Factor intensity: Survey 2005 vs Survey 2015, median regressor

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES						Revenues; log						
α_K	0.20***	0.21***	0.18***	0.16***	0.035***	0.041***	0.037***	0.033***	0.028***	0.030***	0.029***	0.023***
	(0.025)	(0.026)	(0.024)	(0.024)	(0.0096)	(0.0099)	(0.010)	(0.010)	(0.0088)	(0.0092)	(0.0096)	(0.0088)
α_K * Year 2015	0.18***	0.18***	0.16***	0.18***	0.12***	0.12***	0.12***	0.11***	0.098***	0.10***	0.095***	0.092***
	(0.040)	(0.042)	(0.037)	(0.037)	(0.016)	(0.016)	(0.016)	(0.017)	(0.015)	(0.016)	(0.016)	(0.015)
α_L	0.81***	0.79***	0.84***	0.82***	0.24***	0.23***	0.24***	0.22***	0.23***	0.23***	0.21***	0.19***
	(0.031)	(0.032)	(0.029)	(0.041)	(0.014)	(0.014)	(0.015)	(0.020)	(0.013)	(0.013)	(0.014)	(0.017)
α_L * Year 2015	-0.25***	-0.21***	-0.21***	-0.23***	-0.0018	-0.0098	-0.0028	-0.0029	-0.039*	-0.040*	-0.035	-0.033
	(0.050)	(0.051)	(0.046)	(0.047)	(0.021)	(0.022)	(0.022)	(0.023)	(0.022)	(0.023)	(0.024)	(0.022)
α_M					0.73***	0.73***	0.73***	0.72***	0.70***	0.70***	0.71***	0.71***
					(0.010)	(0.011)	(0.011)	(0.012)	(0.0096)	(0.010)	(0.011)	(0.010)
α_M * Year 2015					-0.15***	-0.14***	-0.14***	-0.14***	-0.14***	-0.13***	-0.13***	-0.13***
					(0.015)	(0.016)	(0.016)	(0.017)	(0.015)	(0.015)	(0.016)	(0.015)
α_E									0.053***	0.056***	0.056***	0.058***
									(0.0086)	(0.0090)	(0.0095)	(0.0087)
α_E * Year 2015									0.050***	0.044**	0.047**	0.044**
									(0.019)	(0.020)	(0.020)	(0.019)
Observations	1,440	1,440	1,440	1,439	1,398	1,398	1,398	1,398	1,343	1,343	1,343	1,343
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province dummies	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Sector dummies	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Size dummies	N	N	N	Y	N	N	N	Y	N	N	N	Y

Notes: (a) Standard errors in parentheses. (b) *** p<0.01, ** p<0.05, * p<0.1.(c) Input factors included in regression: Machine values (K), labor cost (L), Material cost (M) and Energy cost (N)

Bibliography

ADB (2011), Energy for all: Viet Nam's success in increasing access to energy through rural electrification, Report, Asian Development Bank.

ADB (2015), Assessment of Power Sector Reforms in Viet Nam - Country Report, Report, Asian Developement Bank.

Adenikinju, A. F. (2003), 'Electric infrastructure failures in Nigeria: a survey-based analysis of the costs and adjustment responses', Energy Policy **31**(14), 1519–1530.

Alam, M. (2013), 'Coping with blackouts: Power outages and firm choices'.

Alby, P. and Dethier, J.-J. (2013), 'Firms Operating under Electricity Constraints in Developing Countries', World Bank Economic Review **27**(1), 109–132.

Allcott, H., Collard-Wexler, A. and O'Connell, S. D. (2016), 'How Do Electricity Shortages Affect Industry? Evidence from India', American Economic Review **106**(3), 587–624.

Anderson, T. W. and Rubin, H. (1949), 'Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations', The Annals of Mathematical Statistics **20**(1), 46–63.

Angrist, J. D. and Pischke, J.-S. (2008), Mostly harmless econometrics: An empiricist's companion, Princeton university press.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R. (1998), 'Large Area hydrologic modeling and assessement part 1: Model development', JAWRA Journal of the American Water Resources Association **34**(1), 73–89.

Baum, C., Schaffer, M. and Stillman, S. (2010), ivreg2: stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression.

Bloomberg (2015), 'Meet asia's new manufacturing powerhouse: Vietnam'.

Cattelaens, P., Limbacher, E.-L., Reinke, F., Stegmueller, F. F. and Brohm, R. (2015), Overview of the Vietnamese Power Market – A Renewable Energy Perspective, Report, GIZ Viet Nam Energy Support Programme.

Cole, M. A. (2006), 'Does trade liberalization increase national energy use?', Economics Letters **92**(1), 108–112.

Cole, M. A., Elliott, R. J., Occhiali, G. and Strobl, E. (2018), 'Power outages and firm performance in Sub-Saharan Africa', Journal of Development Economics **134**, 150 – 159.

Doe, F. and Asamoah, E. S. (2014), 'The Effect of Electric Power Fluctuations on the Profitability and Competitiveness of SMEs: A Study of SMEs within the Accra Business District of Ghana', Journal of Competitiveness **6**(3), 32–48.

Duflo, E. and Pande, R. (2007), 'Dams*', The Quarterly Journal of Economics **122**(2), 601–646.

EAU-WB (2017), Firm Level Productivity Estimates: Methodological Note, Report, Enterprise Analysis Unit - World Bank.

EVN (2015a), EVN Profile 2014-2015, Report, Electricity of Vietnam.

EVN (2015b), Report on National Power System Operation 2014, Report, Electricity of Vietnam.

EVN (2017), Vietnam Electricity Annual Report 2016, Report, Electricity of Vietnam.

FAO (2007), 'Digital Soil Map of the World'.

FAO (2015), 'GAUL 2015 Data License'.

Fisher-Vanden, K., Mansur, E. T. and Wang, Q. (2015), 'Electricity shortages and firm productivity: Evidence from China's industrial firms', Journal of Development Economics **114**, 172–188.

Gassman, P. W., Sadeghi, A. M. and Srinivasan, R. (2014), 'Applications of the SWAT Model Special Section: Overview and Insights', J Environ Qual **43**(1), 1–8.

Grainger, C. A. and Zhang, F. (2017), The impact of electricity shortages on firm productivity : evidence from Pakistan, Report, World Bank Group.

Ha-Minh, D. and Nguyen, H. S. (2017), 'Is electricity affordable and reliable for all in Vietnam'.

Hansen, M., DeFries, R., Townshend, J. and Sohlberg, R. (1998), 'UMD Global Land Cover Classification, 1 Kilometer'.

Hansen, M., DeFries, R., Townshend, J. and Sohlberg, R. (2000), 'Global land cover classification at 1km resolution using a decision tree classifier', International Journal of Remote Sensing **21**, 1331–1365.

Kaur, S. (2014), 'Nominal Wage Rigidity in Village Labor Markets', National Bureau of Economic Research Working Paper Series **No. 20770**.

Kleibergen, F. and Paap, R. (2006), 'Generalized reduced rank tests using the singular value decomposition', Journal of Econometrics **133**(1), 97–126.

Kleibergen, F. and Schaffer, M. (2007), RANKTEST: Stata module to test the rank of a matrix using the Kleibergen-Paap rk statistic.

Lehner, B. and Grill, G. (2013), 'Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems', Hydrological Processes **27**(15), 2171–2186.

Lehner, B., Verdin, K. and Jarvis, A. (2008), '"New Global Hydrography Derived From Spaceborne Elevation Data"', Eos, Transactions American Geophysical Union **89**(10), 93–94.

Lipscomb, M., Mobarak, A. M. and Barilam, T. (2013), 'Development effects of electrification: Evidence from the topographic placement of hydropower plants in brazil', American Economic Journal: Applied Economics **5**(2), 200–231.

Liu, H., Liang, S. and Cui, Q. (2020), 'The nexus between economic complexity and energy consumption under the context of sustainable environment: Evidence from the lmc countries', International Journal of Environmental Research and Public Health **18**(1), 124.

Matsuura, K. and Willmott, C. J. (2009), 'README: Terrestrial Precipitation: 1900-2014 Gridded Monthly Time Series Version 4.01 '.

Meles, T. H. (2020), 'Impact of power outages on households in developing countries: Evidence from ethiopia', Energy Economics **91**, 104882.

Mensah, J. T. (2016), Bring Back Our Light: Power Outages and Industrial Performance in Sub-Saharan Africa, Report.

Min, B. and Gaba, K. (2014), 'Tracking Electrification in Vietnam Using Nighttime Lights', Remote Sensing **6**(10), 9511.

MONRE (2012), Report on National Environment 2012, Report, Ministry of Natural Resources and Environment.

Nguyen, A., Luu, D. and Trinh, C. (2016), The Evolution of Vietnamese Industry, Oxford University Press.

Nguyen, D. X. (2016), 'Trade liberalization and export sophistication in Vietnam', The Journal of International Trade & Economic Development **25**(8), 1071–1089.

Nguyen-Tien, V., Elliott, R. J. R. and Strobl, E. A. (2018), 'Hydropower generation, flood control and dam cascades: A national assessment for Vietnam', Journal of Hydrology **560**, 109–126.

Pham, D. M., Hollweg, C. H., Mtonya, B. G., Winkler, D. E., Nguyen, T. and Nguyen, T. X. (2020), Vietnam : Connecting Value Chains for Trade Competitiveness, World Bank Group.

Pham, H. T., Phan, N. B. and Takayama, S. (2020), Productivity, Efficiency and Firm Size Distribution: Evidence from Vietnam, Discussion Papers Series 617, School of Economics, University of Queensland, Australia.

Picard, R. (2017), GEODIST: Stata module to compute geodetic distances.

Reinikka, R. and Svensson, J. (2002), 'Coping with poor public capital', Journal of Development Economics **69**(1), 51–69.

Sadorsky, P. (2013), 'Do urbanization and industrialization affect energy intensity in developing countries?', Energy Economics **37**(C), 52–59.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Van Den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G. and Goldberg, M. (2010), '"The NCEP Climate Forecast System Reanalysis"', Bulletin of the American Meteorological Society **91**(8), 1015–1057.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M. and Becker, E. (2014), 'The NCEP Climate Forecast System Version 2', Journal of Climate **27**(6), 2185–2208.

Samouilidis, J. E. and Mitropoulos, C. S. (1984), 'Energy and economic growth in industrializing countries : The case of Greece', Energy Economics **6**(3), 191–201.

Sanderson, E. and Windmeijer, F. (2016), 'A weak instrument F-test in linear IV models with multiple endogenous variables', Journal of Econometrics **190**(2), 212–221.

Sarsons, H. (2015), 'Rainfall and conflict: A cautionary tale', Journal of Development Economics **115**, 62–72.

Staiger, D. and Stock, J. H. (1997), 'Instrumental Variables Regression with Weak Instruments', Econometrica **65**(3), 557–586.

Stock, J. H. and Yogo, M. (2005), Testing for Weak Instruments in Linear IV Regression, Cambridge University Press, Cambridge, pp. 80–108.

The Diplomat (2020), 'Developing Vietnam' s Renewable Energy Industry'.

The Growth Lab at Harvard University (2020), 'The Atlas of Economic Complexity'.

Vagliasindi, M. and Besant-Jones, J. (2013), Power Market Structure Revisiting Policy Options, Directions in Development - Energy and Mining, The World Bank, Washington, D.C.

Vincenty, T. (1975), 'Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations', Survey Review **23**(176), 88–93.

WB (2017), 'Vietnam electricity transmission network'.

WB and MPI (2016), Vietnam 2035 : Toward Prosperity, Creativity, Equity, and Democracy, Report, World Bank (WB) and Vietnam Misnistry of Planning and Investment (MPI).

Power Outages and Firm Performance:
A Hydro-IV Approach for a Single Electricity Grid

Online Appendices

November 24, 2021

Contents

A Appendix A: Firm-level Data Supplements

This appendix provides supplemental information about the raw data and our data cleaning process. For variables not mentioned below there was no assumptions made on missing or their extreme values. WBES apply the stratified random sampling method with three 3 levels of stratification for Vietnam: industry, firm size, and region. In the 2015 Survey, firms were stratified into 7 industry groups (5 manufacturing, and 2 services industries); 3 size groups including small (5 to 19 employees), medium (20 to 99 employees), and large (100 or more employees); and 4 region groups (Red River Delta, North Central Area, and Central Coastal Area, South East, and Mekong River Delta). There is no accessible documentation on the sampling method in the 2005 Survey.[1] However, there are similarities in the distribution of firm size, region, and manufacturing sector (see below), which suggests that the three levels of stratification were also applied to the 2005 Survey. As the surveys were conducted in a two-stages, it is recommended that information from the first stage (screening by phone) should be replaced by more reliable information from the second stage (face-to-face interview with the mangers/owners) whenever possible.

[1] Nevertheless, 1.13% of firms recorded that their major product belongs to the service sector.

A.1 Firm Characteristics

A.1.1 Location

Vietnam's administration is organised in a 3-tier system: province level, district level, and commune level. The highest tier (province level) is comprised of 63 units including 5 municipalities (centrally managed cities), namely Ha Noi, Ho Chi Minh City, Hai Phong, Da Nang, and Can Tho, and 58 provinces. The regional classification, although it has no official administrative impact, can be helpful for spatial analysis as it groups municipalities, and provinces that share common features in location, topography, terrain, climate, socio-economy. Accordingly, Vietnam can be divided into three parts (the North, the Central, and the South) or 8 regions (Northwest, Northeast, Red River Delta, North Central Coast, South Central Coast, Central Highlands, Southeast, and Mekong River Delta). Within the context of the WBES Surveys cities are referred to as municipalities. We use 'provinces' to indicate all province level administrative units (municipalities, and provinces).

The 2005 and 2015 surveys contain different variables for location. Survey 2005 has two variables for region, and province. The 2015 Survey only has a variable for region but no variable for province. However, they have an optional variable for location (a3x), which provides information on province, district and/or commune where a firm is located although the details vary across observation. The majority of firms provide information on province, and only a few firms provide details on district and/or commune. Therefore, we choose province as the spatial unit for analysis in our study. Since we can not find information on the location of 2 firms at the province level (a3x variable) in the 2015 Survey, we treat their location as 'unidentified', and exclude them from our analysis. The distribution of firms' locations is illustrated in Table A1. Out of the eight regions across the country only 5 were sampled in the surveys, namely, the Red River Delta, the North Central, the Coastal Central, the South East, and the Mekong River Delta. Among them, the Red River Delta together with the South East account for the majority of the enterprises in the surveys, reflecting their importance to the economy. More than 30% of firms are located in each of these two regions.

There were some changes (splits and mergers) in administration between two surveys. The most prominent is the expansion of the capital city (Hanoi) in 2008.[2] As a large number of variables (see below) are constructed based on the centroids of provinces, one implication to our spatial analysis is that the centroid of Hanoi changes over two surveys. To generate centroid-based variables for firms in Hanoi in Survey 2015, we used a GIS map before the merge (in 2005), construct separate variables for Hanoi, and Ha Tay as if there were no merge, then weighted them by the industrial values just before the merge.[3] Another more simple method is just to calculate the centroid of Hanoi based on a new GIS map after the merge. However, this is less reliable as the appended parts are very large in area but less important in terms of economic activity.

A.1.2 Sector

The main difference between our surveys is that the 2005 Survey looks only at manufacturing firms, which are classified into 17 sector groups.[4] The 2015 Survey includes both manufacturing,

[2]In 2008, the whole Ha Tay province, four communes in Hoa Binh province (Yen Trung, Yen Binh, Tien Xuan, and Dong Xuan of Luong Son district), and one district (Me Linh) of Vinh Phuc province were merged into the capital Hanoi.

[3]Gross output of industry at current prices by province in 2007 of Hanoi, and Ha Tay are 116,096.4, and 20,173.5 (billion VND). The weight applied is 1:0.1737. We ignore the rural areas that used to belong to Hoa Binh, and Vinh Phuc provinces, and latter appended to Hanoi.

[4](1) Food & Beverage; (2) Textiles; (3) Apparel; (4) Leather products; (5) Wood & wood product, including furniture; (6) Paper; (7) Chemical & Chemical products; (8) Rubber & plastic products; (9) Non-metallic mineral products; (10) Basic metals; (11) Metal products; (12) Machinery, and equipment; (13) Electrical machinery; (14) Electronics; (18) Construction materials; (19) Vehicles, and other transport equipment, and (20) Other.

and service firms following the sector classification of ISIC revision 3.1.[5] We need to reclassify sectors for several reasons. First, the sector classification is inconsistent across surveys. Second, in the 2005 Survey, there are 95 firms (8.26 %) coded 20 ('Other') which is much larger than many other sector groups in the same year useful. The 2015 Survey assigned a firm to a sector as a result of screener questionnaires (based on the first stage of the survey on the phone) rather than a face-to-face interview with the Manager/Owner /Director of each establishment (in the second stage of the survey). Hence, it is possible that they contain inaccurate information on the main activity of the firm. We define a sector of firm based on the the two first digits of the ISIC code of the main product in terms of revenue share (provided by the variable d1a2 in the panel database), which were consistently collected in the face-to-face interview phase of all surveys. For observations whose variable was missing or out of the standard value range, we filled in the ISIC code using the information provided by the name of the main product (d1a1x). If no information on the main product was available we used the screener sector to assign the 2-digit ISIC code. This manual filling process was applied for 46 observations (35 in 2005 and 1 in 2015). When all sector codes were assigned, we grouped 2-digit ISIC code into sectors as shown in Table A2.

A.1.3 Firm Size

There is no variable that explicitly classifies firm size in the 2005 Survey. In the 2015 Survey firms are labelled as small (5-19 employees), medium (20-99 employees) or large (100, and above employees). However, this is based on the screener questionnaires, which may contain inaccurate

[5] The codes for manufacturing firms (Section D) are: (2) Other manufacturing; (15) Food; (16) Tobacco; (17) Textiles; (18) Garments; (19) Leather; (20) Wood; (21) Paper; (22) Publishing, printing, and Recorded media; Refined petroleum product; (23) Refined petroleum product; (24) Chemicals; (25) Plastics & rubber; (26) Non metallic mineral products; (27) Basic metals); (28) Fabricated metal products; (29) Machinery, and equipment; (31) Electronics; (33) Precision instruments; Transport machines (34); Furniture (36), and Recycling (37). The service sectors included in the Surveys are: (45) Construction (section F); (50) Services of motor vehicles; (51) Wholesale; (52) Retail; (55) Hotel, and restaurant (section H);, and (60) Transport (section I); (72) IT;, and other .

information. Hence, we followed the manuals and reclassify firm size using the employment data gathered in the face-to-face interviews. Ideally, we would have relied on a combination of data on the number of permanent workers (l1), number of temporary workers (l6), and the average length of temporary employments (l8) to define the dimensions of firms. However, a large number of firms (especially in the 2005 Survey) did not provide their temporary employment data. To avoid the serious loss of information, we use the permanent employment numbers to put firms into four categories: small (<20), medium (20-99), large (100-299), and very large (300, and above). The small number of firms that did not provide this information are excluded from our analysis.

A.1.4 Age

We calculate firm age based on a question when firms started their business. No assumption was made on missing values

A.1.5 Ownership

We are interested in the share of the state and foreign owned firms. Two dummies were generated corresponding to an ownership that exceeds two thresholds 50% or 10%, which reflect different levels of state/ foreign influence on firm performance. The former threshold captures a majority ownership. According to Vietnam' Enterprise Law 2005 and 2014, the holder of 10% shares have some special rights including the right to nominate candidates to the Management Board and the Control Board.

A.1.6 Legal Status

Based on the legal status variable, we generated a dummy for publicly traded companies. We also include a dummy for companies with traded shares among share-holding companies. We drop a negligible number of firms whose legal status is missing.

A.1.7 Credit

We use questions on whether a firm has access to loan/credit from a financial institution to generate a binary variable for access to credit. No assumptions were made on missing values.

A.1.8 Exporters

All surveys contain information about the exporting activities of firms. We assume that firms with missing values in the variables for the percentage of sales made up by indirect or/and direct exports (d3b, and d3c) are not involved in either/both activities (zero percent). Hence, we define a dummy variable for exporters as those that have a combination of both direct, and indirect exports value that account for at least 5% of their revenues.

A.2 Power Supply Provision

In each survey there are a number of variables that relate to the quality of power supply. The first is a binary variable (c6), which reveals whether firms suffered from any power outages over the last fiscal year. If the answer is yes, questions are then asked about the average number of power outages per month (c7), the typical length of an outage (c8a), estimated losses as a percentage of total sales due to power interruption (c9a), and the extent on a five level scale in which electricity is an obstacle to the operation of firms (c30a). Power outage duration was lower bounded at 1 hour in the 2005 Survey and 1 minute in the 2015 Survey.[6] As some firms report unreasonably high values for their power outage measurements, we put a limit of 20 (occurrences/month) on the outage numbers, and 96 (hours/occurrence). Any observation that exceeds these thresholds is replaced by these thresholds. Only 5/1,149 observations in Survey 2005, and 2/960 observations in Survey 2015 were subject to the replacement. The most extreme values of self-reported outage number for Survey 2005, and 2015 are 60, and 20 hours/occurrence respectively. The most extreme values of self-reported outage duration for Survey 2005, and 2015 are 400, and 480 hours/occurrence respectively. If we just use the raw data to calculate the volume of outages, there are three values that exceed 960 hours/month, roughly equivalent to 32 hours/day. We use the power outage frequency variable (the average number of power outages per month - c7), and the power outage intensity variable (typical duration of an outage - c8a) as proxies for power (un)reliability.[7] The outage volume variable was computed as the product of these variable (after limits were imposed).

A typical mitigation measure of firms against the unreliability of power supply is the use of generators. There are two variables that proved useful: a binary variable for the equipment or share of generator(s) (variable c10), and the percentages of electricity generated from the generator(s) owned/shared by firms (variable c11). These variables are available for all firms in the 2005 Survey but not every firm in the 2015 Survey. Only firms that were interviewed by the Manufacturing Questionnaire in this year were asked these questions while other firms (in service sectors) interviewed by the Core Questionnaire or Retail Questionnaire were not.

[6]There is a variable for minutes to add to the average power outage duration in hours (*_2015_c8*), but this exists in the 2015 Survey only.

[7]For the special case of the 2015 Survey, we converted the variable *_2015_c8* from minutes to hours before added to *c8a* to compute the outage intensity variable)

A.3 TFPR Estimate

Since the concept of 'technical change' proposed by Solow (1957), productivity analysis has become an important tool to examine the economic performance in the sense of how efficiently resources are used by an aggregated economy or its agents. Specifically, firm data are used to estimate a production function, and its residual, measured total factor productivity (TFP), can be used to study policy impacts on firm performance (Van Beveren, 2012). OLS estimates of TFP, are known to be biased due to the simultaneity issues (Marschak and Andrews, 1944). The unobserved productivity shocks that firms already take into account when deciding the level of inputs, such as managerial ability, expected machine breakdown, expected defection rate, and expected rainfall pose identification challenges for OLS estimation (Ackerberg *et al.*, 2015). Consequently, positive shocks may lead to overestimation of input usage (De Loecker, 2011). The estimates of labour are downward biased in a two-factor setup where quasi-fixed capital, and flexible labour are positively correlated (Levinsohn and Petrin, 2003). Another source of endogeneity associated with panel data, where firms exit, and entry are dependent on their productivity, is attrition or selection bias. (Ackerberg *et al.*, 2007). This kind of endogeneity underestimates capital input, and overestimates TFP (Van Beveren, 2012).

There have been numerous methods designed to address the endogeneity issues. Fixed effects estimator are widely used to account for time-invariant firm specific characteristics that could correct both simultaneous bias, and selection bias (Pavcnik, 2002; Levinsohn and Petrin, 2003). However the validity of this estimator requires the strict exogeneity of the inputs conditional on firm characteristics (Wooldridge, 2009). Another solution is seeking a valid instrument, and relying on either an IV or GMM estimator. Some sources of instruments are price shocks of inputs, and outputs if the market is perfectly competitive, weather shocks, exogenous shocks from input markets, and dynamics of inputs (Ackerberg *et al.*, 2007; Van Beveren, 2012). Yet the validity of IVs are subject to debate. Olley and Pakes (1996) propose a semi-parametric estimator that uses investment decision as a proxy for the unobserved productivity shocks. The survival problem is also incorporated to account for selection bias that may arise in an unbalanced panel dataset. Alternatively, Levinsohn and Petrin (2003) use intermediate inputs rather than investment, which is susceptible to efficiency losses due to observations with zero value, as a proxy. However, such an estimate does not address selection bias. Some later estimation improvements include those proposed by Ackerberg *et al.* (2007); Wooldridge (2009); Katayama *et al.* (2009); De Loecker (2011).

In this study, as we work with cross-sectional data with a limited lags, there is a limited ability to apply the above methods. Hence, we decided to follow EAU-WB (2017), which uses two versions of the Cobb-Douglas production function to estimate TFP for each sector group, and addresses the endogeneity problem by adding dummies proxied for spatially-variant factors, and time-variant factors that is not observed . The first (the YKL model) assumes firms using two factors, namely capital (K), and labor (L), to generate output (Y). The second (the YKML model) adds materials (M) to the production function. The corresponding parameters $\alpha_K, \alpha_L, \alpha_M$ are factor intensities.

$$Y = A.K^{\alpha_K}.L^{\alpha_L} \text{ (YKL model)}$$

$$Y = A.K^{\alpha_K}.L^{\alpha_L}.M^{\alpha_M} \text{ (YKLM model)}$$

In our approach the estimation is based on a log-linear regression for each sector (indexed by g), including province fixed effects (β_p), and year fixed effects (γ_y), where p, and y are, respectively, subscripts for provinces in which firms are located, and the year of the survey. Manufacturing firms are allocated to 11 sector groups (to ensure a reasonable sample size), including Food, beverages & tobacco (ISIC code 15, 16); Textiles (17); Garments (18); Leather products (19); Wooden products & furniture (20, 36); Paper, printing & publishing (21, 22); Refined petroleum products, chemicals, plastic & rubber (23 - 25); Non-metallic mineral products (26); Basic metal & Fabricated metal products (27, 28); Machinery, office, electronics & precision instruments (29 - 33), and Transport machines (34). Estimations were applied for manufacturing firms within our sample, excluding sectors that had too few observations (i.e., Mining & quarrying, and Recycling).

$$ln(Y_{gpy}) = \alpha_0 + \alpha_K.ln(K_{gpy}) + \alpha_L.ln(L_{gpy}) + \beta_p + \gamma_y + u_{gpy} \text{ (YKL model)}$$

$$ln(Y_{gpy}) = \alpha_0 + \alpha_K.ln(K_{gpy}) + \alpha_L.ln(L_{gpy}) + \alpha_M.ln(M_{gpy}) + \beta_p + \gamma_y + u_{gpy} \text{ (YKLM model)}$$

All variables are observed in monetary terms rather than physical ones: revenue (Y); replacement value of machinery, vehicles, and equipment (K); labor cost (L);, and materials cost (M). Hence, we do not explicitly estimate TFP, but its revenue-based version, TFPR (Total factor product revenue) with the assumption of perfect market as follows:

$$ln(\widehat{TFPR}_{gpy}) = \hat{\alpha}_0 + \hat{\beta}_p + \hat{\gamma}_y + \hat{u}_{gpy}$$

From the two models (YKL, and YKLM), we obtain two sets of estimates of TFPR for manufacturing firms. The results of our production function are shown in Table A3, and A4. The YKLM model predicts total sales better than YKL (higher $\bar{R}^2$) but with a greater loss of observations due to missing values for material input costs. Our results also show heterogeneity in productivity across provinces, and across years.

B Appendix B: Province-level Data Supplements

B.1 Provincial Economic Conditions

Variables for economic conditions at the province level were derived from gross output value of industry, and measures for provincial year-to-year industrial product (IP) growth. All data are extracted from 'Statistical Yearbooks of Vietnam' by Vietnam's General Statistics Office (GSO).[8] Before 2010, the official measure for IP growth at the province level was the Index of industrial output value at constant 1994 prices. From 2010 on-wards, it was replaced by a more precise measure with a different methodology, namely the index of industrial production (IIP) with the base prices in 2010. As gross output value of industry by province is not available for the year 2014, we calculate an alternative, multiplying the gross output value of industry by province at current price in 2010 by a series of IIP from 2011 to 2014. This roughly reflects the gross output value of industry by province in 2014 at 2010 prices. We explicitly use two control variables. First, we compute the province IP share calculated as the ratio between the gross output industry at current prices or at 2010 prices of each province, and the whole country in each year. Second, we explicitly use the province IP index as a measure of IP growth at the province level.

B.2 Cooling Degrees

To compute cooling degree at the province level, we use gridded temperature data from the NCEP-DOE Renanlysis 2 provided by the NOAA/OAR/ESRL PSD. It is among a small number of datasets that provide global daily weather data for the entire year 2014 (and beyond). We down-

[8] Accessible at http://www.gso.gov.vn/Default_en.aspx?tabid=515. Gross output value of industry comprises of the value of industrial products, and industrial services presented at either current prices or constant prices.

loaded the daily forecast of Air temperature at 2m for a bounding box covering all of Vietnam from http://www.esrl.noaa.gov/psd/ in netCDF format then read, and cleaned the data using MatLAB.[9]

While covering a long period of updated, and consistent weather data, one disadvantage of the dataset is the low resolution (2.5^0x2.5), which is not ideal for capturing spatial variations in weather factors across provinces. Hence, we use a spatial interpolation technique to dis-aggregate the data. There are a wide range of techniques, but all fall under 3 general categories: non-geostatistical interpolators, geostatistical interpolators, and combined methods (see Li and Heap (2008) for a comprehensive comparison). We use a combined method, the 'Gradient plus Inverse Distance Squared' (GIDS) method (proposed by Nalder and Wein (1998)) as it is able to account for a wide range of factors in a large scale region with complex terrains (Stahl *et al.*, 2006). The method combines the advantages of multiple linear regression (MLR), and Inverse Distance Squared (IDS).

In the first step, we regress the temperature at daily time steps on a set of regressors including longitude (Lon), latitude (Lat), altitude (Alt), and distance to the sea (Sea), assuming that the temperature at each point (i) on day (t) is determined by its location, and topography characteristics.[10] To take into account of seasonality, regressions for two panels (Year 2004, and 2014) are run separately by month $m = 1, 2, ..., 12$.

$$Tit = \alpha_m + \beta_m Lon_{it} + \gamma_m Lat_{it} + \delta_m Alt_{it} + \phi_m Sea_{it} + \varepsilon_{it} \tag{B1}$$

[9] The bounding box is 6-26^{0}N, and 100-112^{0}E, containing 66 nodes (stations. NetCDF (Network Common Data Form), hosted by the Unidata program at the University Corporation for Atmospheric Research (UCAR), is a set of software libraries, and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data. See https://www.unidata.ucar.edu/software/netcdf/docs/faq.html#whatisit for more details.The original data have been packed into short integers for optimal storage. We unpacked the data by computing $unpacked\_value = add\_offset + packed\_value \times scale\_factor$ with $add\_offset = 0.01$, and $scale\_factor = 447.65$ as provided by the metadata.

[10] Lat, Lon, Alt are the variables most frequently used in GIDS. We include Sea to the model as Vietnam has a long coastal line hence its climate is affected by the sea. The impact of the sea on local climate is studied in Maureen and Jean (2000).

The results of the MLR are shown in Table B1. All regressors appear to be significant predictors of temperature. Overall, temperature is lower in the North (high latitude) across all years, while the sign of variation across longitude is subject to seasonality. Temperature falls in more elevated regions, with an estimated lapse rate of 1.07 - 1.99 $^0C/km$ in 2004, and 0.88-1.85 $^0C/km$ in 2014. These values are lower than those proposed in the literature ($6-6.5^0C$), probably due to the multi-collinearity with other regressors (for example, mountainous, and hilly terrains are concentrated in the Northern and Western parts of the countries). In addition, the sea also has a significant impact on climate patterns, lowering the temperature with a magnitude that varies across the month. The predictability of the regression is stronger between June- January (R^2 exceeding 80%). In other months, the R^2 is lower but never falls below 65%. The goodness-of-fit of the model is much lower (56.1% in 2004, and 49.7% in 2014) when we pool all observations across different months and captures the strong seasonality of the climate pattern. As a result we run separate regressions for each month.

After obtaining parameters of the MLR, we predict the temperature of province j using temperatures at the five nearest stations based on IDW method.

$$\begin{aligned} T_{jt} = &\Big(\sum_{k=1}^{5} \frac{1}{d_{jk}^2}\Big)^{-1} \\ &\times \sum_{k=1}^{5} \Big[T_{kt} + \hat{\beta}_m(Lon_{jt} - Lon_{kt}) + \hat{\gamma}_m(Lat_{jt} - Lat_{kt}) + \hat{\delta}_m(Alt_{jt} - Alt_{kt}) + \hat{\phi}_m(Sea_{jt} - Sea_{kt})\Big]\frac{1}{d_{jk}^2} \end{aligned} \tag{B2}$$

where d_{jk} is the geodesic distance between the centroid of province j, and station k. T_{kt} is the temperature observed in station k at time t, which is located at Lon_{kt}, Lat_{kt}, Alt_{kt}, and within a distance of Sea_{kt} to the sea. T_{jt} is the interpolated temperature of province j at time t, whose centroid is located at Lon_{jt}, Lat_{jt}, and within a distance of Sea_{jt} to the sea. Alt_{jt} is calculated as the average elevation of the target province. $\hat{\beta}_m$, $\hat{\gamma}_m$, $\hat{\delta}_m$, and $\hat{\phi}_m$ are parameters estimated in Equation B1. The method is essentially equivalent to assigning $T_{jt} = \hat{T}_{jt} + \bar{\hat{\varepsilon}}_{jt}$, where $\hat{T}_{jt}$ is the predicted temperature for province j at time t by regression B1, and $\bar{\hat{\varepsilon}}_{jt}$ is the weighted average of the residuals at 5 nearest stations.[11] Intuitively, $\hat{T}_{jt}$ accounts for predetermined factors in the model B1 while ε_{jt} accounts for other unobserved factors outside the model. The predicted values are in Celsius degrees.[12] We converted them into Fahrenheit, and then calculated the cooling degree $C_{it} = max\{0; T_{it} - 65\}$, which was then aggregated at the year level.

B.3 Rainfall Shocks

For rainfall data we use 'Terrestrial Air Temperature, and Precipitation: Monthly Climatologies' (version 4.1) by the Daleware University (Matsuura and Willmott, 2009). Previous studies have used this data to construct rainfall shock variables (Kaur, 2014; Sarsons, 2015; Allcott *et al.*, 2016). We obtained monthly data for 40 years (1975-2014) for 459 stations at 0.5× 0.5 degree resolution.[13] Although this resolution is finer than the temperature data but still not sufficient to capture variation across provinces.

To interpolate rainfall data, we use a similar GIDS method as described above, where temperature is replaced by rainfall. There are several differences. First, the rainfall variable is log transformed to reduce the skewness of the dependent variable, and to ensure that the predicted values of rainfall are non-negative. Such a transformation is necessary as the normality of all monthly rainfall series are rejected by either Shapiro –Wilk or Shapiro –Francia test at the 1%

[11] The mean(maximum) distance in km from province centroid to the 1st, 2nd, 3rd, 4th, and 5th nearest station are 75(124); 144(185); 191(213); 213(235); 250(294).

[12] The raw data were recorded in Kelvin.

[13] Restricted by the bounding box 8-24^{0}N, and $100 - 112^0$E.

level. We consider two regression specifications: double log regression, and semi log regression as illustrated in Table B2.

All predetermined factors (coordinates, altitude, and distance to the sea) appear with significant coefficients, however the sign, and magnitude of these variables change across months, reflecting the seasonal patterns. The predictability of the regressions are much lower than those for temperature in Table B1, so the IDW component is probably more important in the interpolation method. We decided to use the double log regression as it is easier to interpret the results, and the performance of the model across months is more stable.[14] After obtaining the parameters, we apply the IDW method, similar to Equation B2, using 7 nearest stations to predict the log of rainfall at the centroid of the province (assigned average province altitude). [15]

The Standardised Precipitation Index (SPI) for each province is constructed as follows. First we aggregate by 12 month moving windows the 40 years for each province. Second, we fit the data to a Gamma distribution whose cumulative probability is determined by 2 parameters $\alpha, \beta > 0$:

$$G(x) = \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} r^{\alpha-1} e^{-r/\beta} \tag{B3}$$

where $r > 0$ is the accumulated precipitation (rainfall in our case), and $\Gamma(\alpha)$ is the Gamma function defined by:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \tag{B4}$$

The cumulative distribution function $G(x)$ is then transformed into a normal distribution with

[14]For example, rainfall in January respectively increases by 2.06%, 9.66%, 0.62%, and 0.09% in response to 1% increase in latitude, longitude, altitude, and proximity to the sea. The R^2 varies between 0.033, and 0.209 in the double log specification, and between 0.020, and 0.227 in the semi log specification.

[15]The mean(maximum) distance in km from province centroid to the 1st, 2nd, 3rd, 4th, 5th, 6th, and 7th nearest station are 20(39); 41 (59); 52(73); 62(89); 74(98); 78(99) 87(114).

mean r^*, and standard deviation σ.[16] Finally, the SPI is computed as:

$$SPI = \frac{r - r^*}{\sigma} \tag{B5}$$

SPI for province j at time t is calculated as the SPI of province j in the window that contains all months of year t. The spatial distribution of cooling degree and SPI is shown in Figure B1.

B.4 Grid-based Distance to Hydropower Plants

The electricity lines dataset includes 116 polylines, of which 16 are 500-kV grid, and 100 are 220-kV grid. In terms of status, 48 are labelled 'existing', and 68 are lablled 'planned'. Supplementary documents suggest that the 'existing' lines are those on-line by January 2005, and the 'planned' are those proposed to be built by 2010 or 2020. Unfortunately, other details on the dynamic of the system are unavailable (the operation year of each line). For simplicity, we use all grids (both those labelled 'existing', and 'planned') for both WBES while allowing the number of hydropower sources to increase over time. A justification for the use of 'planned' transmission lines even in the earlier WBES is that although they certainly were not constructed in 2004, there could have been lower-voltage lines that had been operating in that location (i.e., 22 kV up to 110 kV), and were later upgraded to 220 kV or 500 kV as 'planned' lines in this dataset.

The construction of the grid-based distance between each pair of hydropower plants, and a province centre is as follows. First, we assigned the plant, and the province the nearest point on the grid, then compute the grid-based distance as the sum of the geodesic distances between the plant/ province, and their assigned nearest points, and the length of the shortest route that

[16] If r contain zero values, the cumulative prbability should be expressed as $H(x) = q + (1-q)G(x)$, where q is the probability to have zero rainfall. However, zero values do not present in our accumulate rainfall series at 12-month windows.

connects the two assigned points through the transmission network. Technically, we used a number of different ArcGIS tools to aid our computation. First, the assignment of the nearest points, and the calculation of their distances to either the power source or the electric consumption center was executed by the Near tool of the Analysis toolbox. The Integrate tool of the Data Management toolbox was applied to strengthen the connectivity of the poly-lines within the dataset, which was then transformed into a network dataset by Network Analyst extension. Finally, the distance between the assigned grid points of all pair of given hydropower plants, and WBES surveyed provinces were computed by the Origin-Destination cost matrix analysis tool, and visualised for validation by the Route analysis tool.

The visual validation detected a number of pairs that are linked by unreasonably long routes compared with their Geodesic distances. A reason is the absence of low-voltage lines (under 220 kV) in our grid dataset, which help connect those in close proximity. This requested for some adjustments. As a results, We computed the ratios between the grid-based distances, and Geodesic distances, and then define outliers as those that are more than one standard error deviation from the sample mean. We then adjust the grid-based distances of the outliers by the product of their geodesic distances, and the mean ratio of the sample. Figure B2 provides a comparison on the grid-based distances, and the Geodesic distances before, and after the adjustment.

C Appendix C: Hydropower Generation

Table C1 lists the name of large hydropower plants used to construct instrumental variables in this study. As mentioned in section 5.2.2, we use predicted values of hydropower generation after stripping an error term that could be corrected with demand variability to construct a metrics for

Hydrologically Predicted Plant Factor (HPPF). Our estimation is formally given by:

$$Gen_{ist} = \beta_0 + \beta_1 Flow_{ist} + \beta_2 Flow^2_{ist} + \beta_3 CAP_{ist} + \beta_4 UpCAP_{ist} + \mu_i + \varepsilon_{ist} \tag{C6}$$

where i, s, t are indices for dam, month, and year, respectively; Gen is average daily generation by month (MWh/day); $Flow$ is the simulated discharge from SWAT model; CAP is the installed capacity of the dam; $UpCAP$ is the combined installed capacity of upstream dams; μ_i is dam fixed effects, and ε_{ist} is the idiosyncratic error term. The installed capacity provides us with information on the optimized generation of a given dam at its designed level of discharge, while the function of simulated flow explains the deviation in the generation of electricity due to the deviations in the discharge from the value that the dam was designed for. The inclusion of a squared discharge term reflects differences in turbine efficiency: dams under extreme hydrological conditions (droughts) generate less electricity than predicted given water availability. The upstream installed capacity takes into account the impact of the upstream dams' operation on electricity generation. We include dam fixed effects to take into account other sources of heterogeneity across dams. Table C2 presents the results of the model used to generate our IV for the baseline regressions and explains 87.8% of the variation in electricity generation at the dam level.

D Appendix D: Robustness Checks

D.1 Alternative Instrumental Variables

Table D1 presents a number of robustness checks, where alternative IVs are used, while the other control variables are kept unchanged. The first six columns are for Survey 2005, and the others are for Survey 2015. For simplicity, we only use volume to measure the degree of power outages in these robustness checks.

In Columns (2), and (8) the IVs were constructed from geodesic distance rather than grid-based distance in baseline. In general there is just little difference in the estimates of coefficients in regressions using these IVs. If the grid map had been missing, regressions based on IVs derived from geodesic distances would have still been useful. The IVs that are constructed directly from the real generation of hydropower plants (Column (3), and (9)) appear to be much weaker than the baseline. The coefficients of revenue in these columns tend to be larger than those in the baseline, and even positive (Column 3), suggesting the problem with exogeneity of the IV constructed by actual generation. Due to the operation of large reservoirs, the real generation series incorporates a large part of power demand fluctuation, and loses a certain amount of exogenous shocks from weather factors, and just adds limited information to the reduced equations once demand variations are strictly controlled for (i.e., cooling degree, IP index, IP share). This highlights the importance of the use of the hydrologically-based generation model in this study.

In Columns (4), and (10) we construct the IVs from the generation model that excludes the installed capacity of upstream dams. Consequently, the IV for Survey 2015 become much weaker as cascades of hydropower play an important role in the power supply of the country. Taking the synergies of dams into account strengthens the IVs in our national-scaled study with a single-grid context. In column (5) we construct the IVs for Survey 2005 that strictly replicate those of Survey 2015, based on hydrologcially-predicted generation in the first 7 months rather than the entire year. The IVs seem to be a little bit weaker than the baseline, and tend to report a larger magnitude of the reduction in performance. However, such a change does not drive the reduction in output. In Columns (6), and (12) we include time dummies to the generation prediction equation prior to the construction of the IVs. These alternative IVs seem to be weaker than the baseline. The 2SLS estimates using different alternative IVs are consistent with our baseline results. The alternative IVs confirm the finding that power deficiency is more important in Survey 2015 rather than Survey 2005. The alternative IVs also find a reduction in TFPR estimated by the YKLM model for the 2005 Survey, and all of them report a significant reduction in the YKL model results for the 2015 Survey. Almost every set of IVs reports an increase in generator use, and a reduction in energy, in general, and grid electricity, in particular, in response to more severe power outages in both years. A t-test in column (9), and (10) report a significant reduction in materials, and labour inputs due to power disruptions for Survey 2015 while the ARχ^2 test report these results in the majority of columns.

D.2 Firm Heterogeneity

Table D2 shows the results of the baseline 2SLS regressions for different sub-samples: manufacturing firms, firms with generators, firms without generators, electricity-intensive firms, non-electricity-intensive firms, exporters, and non-exporters.[17] Again, we use outage volume to measure the unreliability of power supply. According to the first stage results, the IV is relevant for all sub-samples in both year (at the 5% level, and even the 1% level for many sub-samples). In panel A (Survey 2005) the signs for the impacts of power outages on sales are all negative, except for the subsample of generator-owners, and non-exporters. However, all of the impacts are insignificant as reported by the t-tests. The AR χ^2 tests reports a significant sale loss for the non-exporter sub-sample only where the associated coefficient is -0.86. For panel B (Survey 2015) all sub-samples excluding non-electricity-intensive show a significant reduction in revenues where outage volume

[17] Electricity-intensive firms are labelled as those in the industries whose electricity share in revenue exceeds the median of the sample.

is high. In terms of magnitude, the coefficients vary between -0.16 to -1.48. In both years the magnitude of sale loss due to power outages slightly increases when services firms are excluded.

D.3 Deviations in the Survey Designs

In this subsection we review some differences in the design, and implementation of two surveys, and test whether they drive the difference in the magnitude, and significance of power outage impacts. First, Survey 2015 reports the duration of power outages with higher accuracy as it includes a variable for minutes in addition to a variable of duration in hour (_2015_c8).[18] When we round up duration in Survey 2015 according to the style in Survey 2005, there are only discernible changes (at the second decimal place) in the estimates of the coefficient, and their standard errors, and no change in the significance of power outage impacts.

Second, while all firms in Survey 2005 reports the data for the entire 2004 fiscal year, a number of firms (293 firms ∼ 29.5%) do not report for the entire 2014 fiscal year.[19] In the baseline regressions we retain all observations, and assume that other control variables fully account for the difference between these firms, and those that report the entire 2014 fiscal year. For robustness we include a dummy to indicate the group of these special observations in Column (2) of Table D3. The coefficient of output loss is -0.55, a little bit smaller in magnitude than the baseline (-0.73), however still larger than the estimate for Survey 2005. A t-test can not reject the null that the newly added dummy is an insignificant predictor of the revenue given the rich control variable set. This may help justify our use of the baseline specification.

[18] That means firms in Survey 2015 report the duration up to a minute (for example 2 hour, and 18 minutes) while firms in Survey 2005 only report up to a half of hour (for example .5 hours).

[19] It could happen as some interviews occur too early for firms to prepare data for the whole fiscal year, and hence report the data of the latest 12 months instead.

Third, in columns (3) - (5), we take into account the weights, which are available for Survey 2015 but not 2005. Three weights are provided in the dataset, computed based on stratum, and three different assumptions on the eligible firms (strict, median, and weak).[20] As can be seen, the three weights provide quite similar coefficients for all dependent variables. These estimates confirm the evidence provided by the baseline that firms with a disadvantage in power provision generate less revenue, self-generate more, and use less electricity from the grid. As the stratification in the WBES is based on 3 conditioning variables (sector, size, province), we prefer the unweighted estimator. As long as the stratification is based on the independent variables, the unweighted estimator is shown to be consistent, and the validity of the usual asymptotic variance matrix estimator remains (Wooldridge, 2010, pg863). Furthermore, under generalised conditional information matrix equality, the weighted estimator is less efficient than the unweighted estimator (Wooldridge, 1999, 2001).

[20] Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility, and those that rejected the screener questionnaire or an answering machine or fax was the only response. Weak assumption: in addition to the establishments included in points a, and b, all establishments for which it was not possible to contact or that refused the screening questionnaire are assumed eligible.

Figure B1: Cooling degree and Standardised Precipitation Index (SPI)

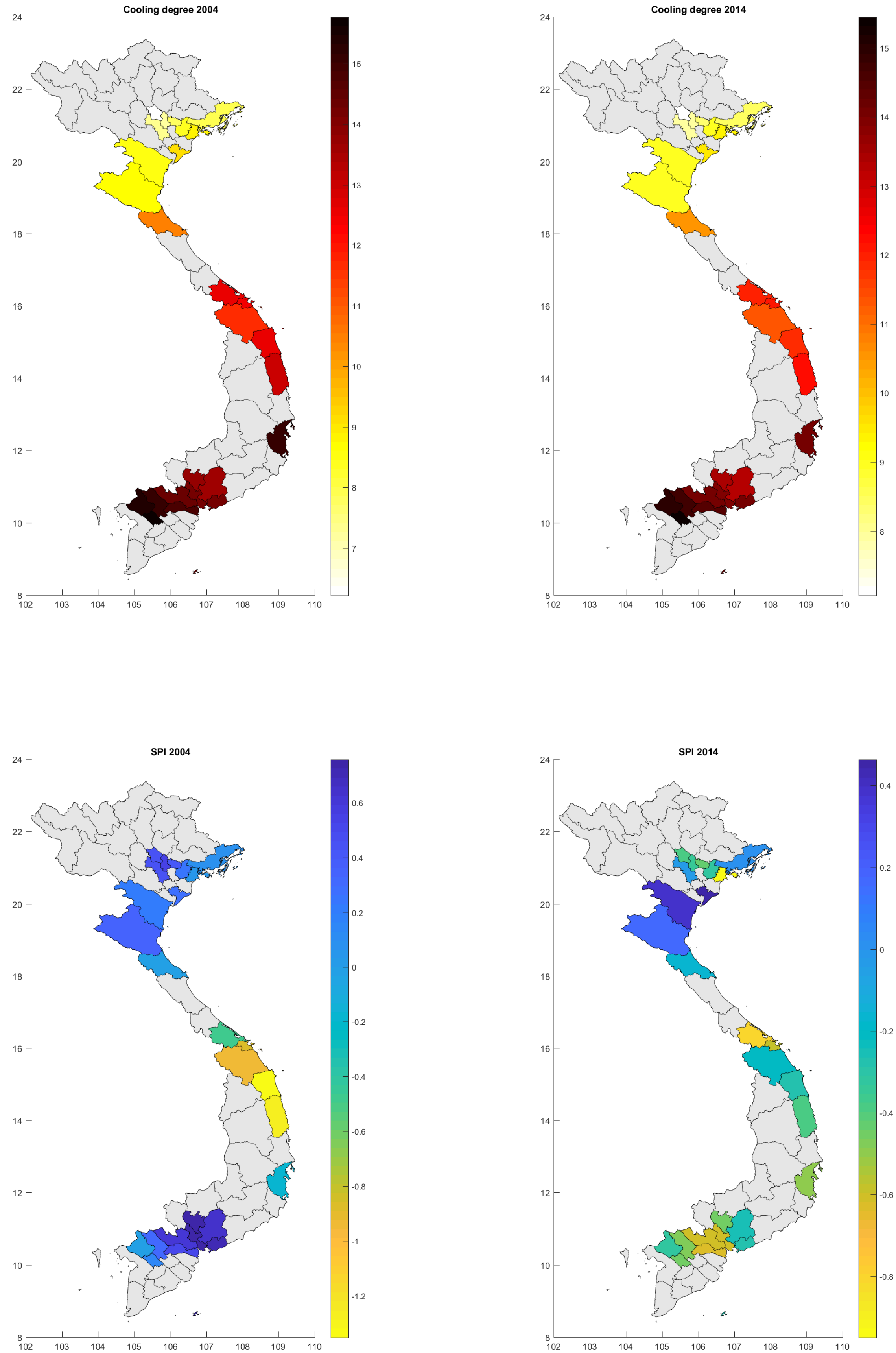


Table A1: Province, and region statistics by year

Province	Region	2005		2015	
		Frq	*Per*	*Frq*	*Per*
Ha Noi (mun. cap.)	Red River Delta	139	12.09%	174	17.47%
Hai Phong (mun.)	Red River Delta	80	6.96%	72	7.23%
Ha Tay (before merge)	Red River Delta	44	3.83%	0	0.00%
Bac Ninh	Red River Delta	22	1.91%	9	0.90%
Hai Duong	Red River Delta	19	1.65%	30	3.01%
Nam Dinh	Red River Delta	41	3.57%	11	1.10%
Quang Ninh	Red River Delta	0	0.00%	8	0.80%
Total region	***Red River Delta***	***345***	***30.00%***	***304***	***30.52%***
Thanh Hoa	North Central	64	5.57%	47	4.72%
Nghe An	North Central	39	3.39%	48	4.82%
Ha Tinh	North Central	27	2.35%	0	0.00%
Thua Thien Hue	North Central	20	1.74%	6	0.60%
Total region	***North Central***	***150***	***13.04%***	***101***	***10.14%***
Da Nang (mun.)	Coastal Central	46	4.00%	81	8.13%
Quang Nam	Coastal Central	16	1.39%	0	0.00%
Quang Ngai	Coastal Central	9	0.78%	0	0.00%
Binh Dinh	Coastal Central	39	3.39%	13	1.31%
Khanh Hoa	Coastal Central	38	3.30%	44	4.42%
Total region	***Coastal Central***	***148***	***12.87%***	***138***	***13.86%***
Ho Chi Minh (mun.)	South East	242	21.04%	191	19.18%
Binh Duong	South East	77	6.70%	30	3.01%
Dong Nai	South East	61	5.30%	83	8.33%
Ba Ria - Vung Tau	South East	15	1.30%	3	0.30%
Total region	***South East***	***395***	***34.35%***	***307***	***30.82%***
Long An	Mekong River Delta	36	3.13%	60	6.02%
Dong Thap	Mekong River Delta	7	0.61%	0	0.00%
An Giang	Mekong River Delta	16	1.39%	0	0.00%
Tien Giang	Mekong River Delta	16	1.39%	32	3.21%
Can Tho (mun.)	Mekong River Delta	37	3.22%	52	5.22%
Total region	***Mekong River Delta***	***112***	***9.74%***	***144***	***14.46%***
	Unidentified	***0***	***0.00%***	***2***	***0.20%***
	Total sample	**1150**	**100.00%**	**996**	**100.00%**

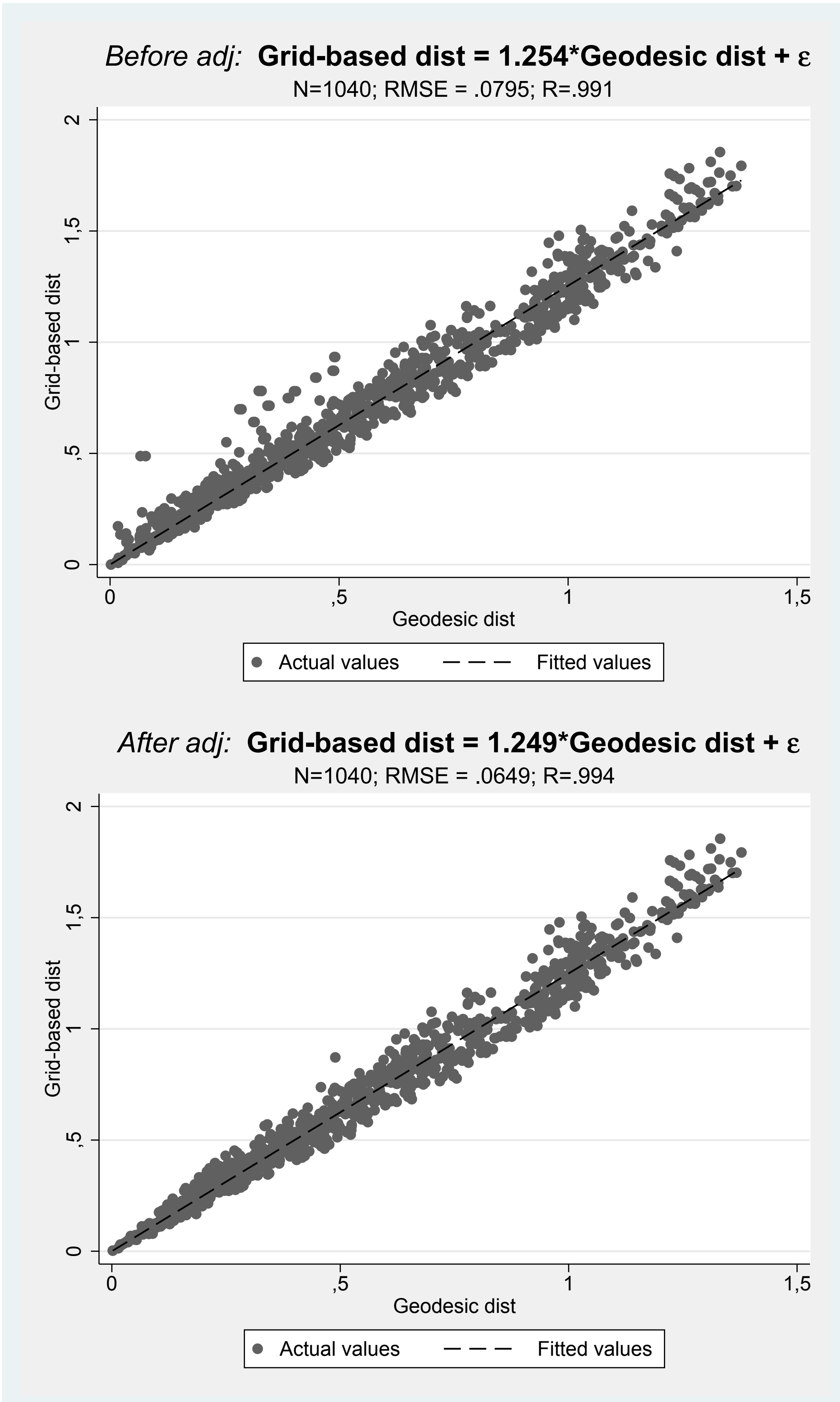
Source: Author's extracted from the clean database. *Note:* mun. = municipalities; cap. = the capital

Table A2: Sector reclassification

Sector	ISIC code	2005		2015	
		Freq	*Percent*	*Freq*	*Percent*
Mining & quarrying	10 - 14	7	0.61%	0	0.00%
Food, beverages & tobacco	15, 16	197	17.13%	130	13.05%
Textiles	17	46	4.00%	24	2.41%
Garments	18	110	9.57%	140	14.06%
Leather	19	41	3.57%	14	1.41%
Wood	20	88	7.65%	21	2.11%
Paper	21	70	6.09%	8	0.80%
Publishing, printing & recorded media	22	2	0.17%	12	1.20%
Refined petroleum product & chemicals	23, 24	72	6.26%	19	1.91%
Plastics & rubber	25	58	5.04%	20	2.01%
Non-metallic mineral products	26	126	10.96%	131	13.15%
Basic metals	27	36	3.13%	14	1.41%
Fabricated metal products	28	63	5.48%	98	9.84%
Machinery	29	66	5.74%	35	3.51%
Office, electronics & precision instruments	30-33	36	3.13%	10	1.00%
Transport machines	34, 35	57	4.96%	3	0.30%
Furniture	36	62	5.39%	12	1.20%
Recycling	37	0	0.00%	3	0.30%
Construction (section F)	45	9	0.78%	75	7.53%
Motor vehicles services	50	1	0.09%	13	1.31%
Wholesale	51	0	0.00%	106	10.64%
Retail	52	0	0.00%	61	6.12%
Hotels and restaurants (section H)	55	0	0.00%	15	1.51%
Transport (section I)	60 - 64	2	0.17%	28	2.81%
IT	72	0	0.00%	4	0.40%
Other services	85	1	0.09%	0	0.00%
Unidentified	99	0	0.00%	0	0.00%
Total		**1150**	**100.00%**	**996**	**100.00%**

Source: Authors redefined sector of firms based on the 2 first digits of the ISIC code of the main product in terms of revenue share.

Figure B2: Grid-based distances vs Geodetic distances comparison



Note: The graph compares the Grid-based distances and Geodetic distances between Vietnam's provinces surveyed by the WBES 2005/2015 and hydropower dams used to construct the IV. See text for more definitions and methodology of distance construction. Adjustments were applied on the outliers, whose ratio between the grid-based distance and geodetic distance more than one standard error deviates from the sample mean. The grid-based distances of the outliers were replaced by the product of the corresponding geodetic distances and the sample mean of the ratio.

Table A3: TFPR estimate using YKL model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11
Factor											
Capital stock; log	0.247***	0.166	0.477***	0.367*	0.288***	0.119**	0.123*	0.209***	0.260***	0.274***	0.123*
	(0.0524)	(0.107)	(0.0576)	(0.197)	(0.0617)	(0.0460)	(0.0657)	(0.0521)	(0.0730)	(0.0526)	(0.0700)
Labor cost; log	0.762***	0.803***	0.457***	0.514*	0.653***	0.853***	0.855***	0.801***	0.724***	0.793***	0.935***
	(0.0665)	(0.141)	(0.0741)	(0.275)	(0.104)	(0.0896)	(0.0810)	(0.0774)	(0.0802)	(0.0730)	(0.140)
Province											
Hai Phong (mun.)	0.0865		-0.298	-0.573	1.178*	-1.502***	0.0349	0.654	0.451	-0.140	-0.0268
	(0.358)		(0.387)	(1.252)	(0.605)	(0.363)	(0.285)	(0.623)	(0.408)	(0.244)	(0.477)
Ha Tay (before merge)	-0.556*	-1.253	-0.123		0.269	-0.808**		-0.290	0.834	-1.350***	0.0388
	(0.323)	(0.855)	(0.302)		(0.358)	(0.315)		(0.302)	(0.554)	(0.375)	(0.430)
Bac Ninh			0.285	-0.339	0.453	-0.559*		-0.000906	0.589**	-0.227	
			(0.946)	(0.914)	(0.484)	(0.279)		(0.488)	(0.258)	(0.208)	
Hai Duong	0.170	-1.846***	0.144	0.155	0.301	-0.297	-1.426***	0.353	0.499	-0.599**	
	(0.262)	(0.338)	(0.317)	(1.159)	(0.457)	(0.457)	(0.442)	(0.680)	(1.022)	(0.253)	
Nam Dinh	-0.194	-0.932	0.164	0.361	0.432	-0.271	0.235	-0.360		-0.388	-0.631
	(0.569)	(0.698)	(0.626)	(0.781)	(0.365)	(0.239)	(0.552)	(0.314)		(0.332)	(0.695)
Quang Ninh					1.748***			-0.177	0.0320		
					(0.283)			(0.288)	(0.266)		
Thanh Hoa	0.0613	-0.0214	0.102		-0.126	-0.686	-0.827***	-0.444	0.0744	-0.156	-0.445
	(0.420)	(0.308)	(0.509)		(0.325)	(0.760)	(0.276)	(0.286)	(0.391)	(0.273)	(0.448)
Nghe An	-0.527*		-0.216	0.216	0.258	-0.483*	0.665	-0.250	-0.205	0.467*	
	(0.299)		(0.400)	(0.771)	(0.265)	(0.287)	(0.591)	(0.291)	(0.697)	(0.279)	
Ha Tinh	0.00359	-0.722	-0.527		0.341	0.881***		-0.120	-0.402	-0.573***	-0.534
	(0.540)	(0.490)	(0.326)		(0.451)	(0.233)		(0.338)	(0.276)	(0.202)	(0.386)
Thua Thien Hue	-0.831***	-1.545**	-0.610*	1.110	0.347	1.418***	-0.0818		-0.349	-0.671	-0.00179
	(0.282)	(0.589)	(0.314)	(1.295)	(0.490)	(0.372)	(0.357)		(0.326)	(0.462)	(0.362)

cont.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11
Da Nang (mun.)	0.171	-0.229	0.0333	0.603	-0.150	-0.942***	0.589	-0.451	0.275	0.524	-0.182
	(0.438)	(0.752)	(0.359)	(1.254)	(0.356)	(0.273)	(0.524)	(0.332)	(0.321)	(0.592)	(0.410)
Quang Nam	-0.405		-0.0910	-1.396	0.316	-0.723***	-0.579	-0.0691			0.794*
	(0.499)		(0.378)	(0.896)	(0.387)	(0.198)	(0.932)	(0.471)			(0.393)
Quang Ngai	-0.956***				-0.566*		0.124	-0.182			-0.385
	(0.241)				(0.288)		(0.360)	(0.557)			(0.469)
Binh Dinh	0.255	-3.013***	0.114		0.598**	-0.745***	-0.428	-0.0135	-0.296		
	(0.430)	(0.420)	(0.617)		(0.295)	(0.216)	(0.305)	(0.612)	(0.399)		
Khanh Hoa	0.221	0.574*	0.200		0.463	-0.905***	-0.152		-0.403	0.259	-0.362
	(0.308)	(0.318)	(0.544)		(0.541)	(0.237)	(0.328)		(0.249)	(0.230)	(0.400)
Ho Chi Minh (mun.)	0.320	-0.472	0.567*	0.601	0.0701	-0.490*	0.187	-0.146	0.309	-0.136	0.420
	(0.296)	(0.432)	(0.306)	(1.264)	(0.281)	(0.264)	(0.280)	(0.420)	(0.324)	(0.231)	(0.515)
Binh Duong	0.171	-0.459	-0.303	0.423	0.769**	-0.848	0.155	-0.539**	0.212	-0.0998	-0.0439
	(0.716)	(0.425)	(0.398)	(0.978)	(0.364)	(0.515)	(0.520)	(0.265)	(0.581)	(0.458)	(0.540)
Dong Nai	0.373	-0.0166	-0.154	-0.286	0.469	0.228	0.382	0.333	0.857*	-0.118	0.395
	(0.547)	(0.335)	(0.439)	(0.604)	(0.390)	(0.376)	(0.352)	(0.303)	(0.464)	(0.262)	(0.512)
Ba Ria - Vung Tau	0.109		-0.666	0.0309					0.398		
	(0.317)		(0.634)	(1.172)					(0.291)		
Long An	0.106	0.197	-0.112	1.186	0.101	-0.698**	-0.00645	-0.0830	0.485	-0.0357	
	(0.307)	(0.560)	(0.410)	(1.304)	(0.236)	(0.280)	(0.431)	(0.513)	(0.457)	(0.295)	
Dong Thap	1.143										-1.747***
	(1.353)										(0.429)
An Giang	0.852***		0.102	0.931					0.240		
	(0.327)		(0.332)	(1.361)					(0.268)		
Tien Giang	-0.362	-2.503***	-0.127				-0.967**	-0.892***	1.945***		
	(0.416)	(0.385)	(0.393)				(0.378)	(0.263)	(0.347)		
Can Tho (mun.)	1.232***		-0.151	-0.0339		-0.521	-0.196	0.263	-0.241	-0.705	0.762
	(0.323)		(0.311)	(1.103)		(0.378)	(0.558)	(0.798)	(0.409)	(0.776)	(0.629)
Survey											
2015	-0.369**	0.566	0.452**	0.0578	0.428*	-1.222***	-0.163	0.154	-0.0320	-0.256	-0.917

cont.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11
	(0.177)	(0.431)	(0.192)	(0.814)	(0.244)	(0.257)	(0.229)	(0.176)	(0.213)	(0.258)	(0.605)
Constant	2.103***	2.900	2.253*	3.256*	2.606**	3.399**	2.726***	1.627	2.205**	0.962	1.387
	(0.807)	(2.196)	(1.196)	(1.646)	(1.298)	(1.348)	(0.990)	(0.993)	(0.991)	(0.829)	(2.148)
Observations	243	58	191	44	161	75	135	195	155	107	54
R-squared	0.785	0.799	0.751	0.787	0.720	0.875	0.801	0.748	0.693	0.880	0.867

Note: Sector group definition: Group 1 = Food, beverages & tobacco (ISIC code 15,16); Group 2 = Textiles (17); Group 3 = Garments (18); Group 4 = Leather products (19); Group 5 = Wooden products & furniture (20, 36); Group 6 = Paper, printing & publishing (21 & 22); Group 7 = Refined petroleum products, chemicals, plastic & rubber (23 - 25); Group 8 = Non-metallic mineral products (26); Group 9 = Basic metal & Fabricated metal products (27 & 28); Group 10 = machinery, office, electronics & precision instruments (29 - 33); Group 11 = Transport machines (34).

The reference group for fixed effects is firms that are located in Hanoi (mun. cap.), and surveyed in 2005. The dependent variable is the log of revenues. Standard errors in parentheses. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A4: TFPR estimate using YKLM model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11
Factor											
Capital stock; log	0.0762***	0.0186	0.211***	0.0636	0.0437**	0.0446	0.103*	0.184***	0.0560	0.123***	-0.00914
	(0.0222)	(0.0655)	(0.0505)	(0.0624)	(0.0184)	(0.0297)	(0.0535)	(0.0480)	(0.0402)	(0.0303)	(0.0237)
Labor cost; log	0.269***	0.508**	0.265***	0.262***	0.160***	0.348***	0.316***	0.390***	0.302***	0.348***	0.138***
	(0.0389)	(0.196)	(0.0517)	(0.0750)	(0.0418)	(0.0786)	(0.0770)	(0.0793)	(0.0834)	(0.0757)	(0.0326)
Material cost; log	0.616***	0.447***	0.464***	0.684***	0.780***	0.590***	0.546***	0.453***	0.633***	0.575***	0.877***
	(0.0368)	(0.152)	(0.0364)	(0.0694)	(0.0368)	(0.0821)	(0.0950)	(0.0645)	(0.0808)	(0.0944)	(0.0324)
Province											
Hai Phong (mun.)	0.125		0.0481	0.784*	0.214	-0.661**	0.148	-0.159	0.166	-0.00594	-0.00619
	(0.126)		(0.221)	(0.379)	(0.142)	(0.321)	(0.186)	(0.303)	(0.195)	(0.189)	(0.0992)
Ha Tay (before merge)	-0.155	-0.533	0.501*		0.0640	-0.531**		-0.123	0.230	-0.639**	-0.0740
	(0.122)	(0.400)	(0.263)		(0.149)	(0.236)		(0.230)	(0.261)	(0.244)	(0.101)
Bac Ninh			0.557***	0.421	0.0102	-0.419*		0.0237	-0.197	0.0738	
			(0.198)	(0.356)	(0.142)	(0.228)		(0.339)	(0.132)	(0.168)	
Hai Duong	0.360	0.0419	0.110	0.0368	-0.229	-0.376	-0.228	-0.248	0.0619	-0.260*	
	(0.267)	(0.807)	(0.300)	(0.392)	(0.148)	(0.243)	(0.266)	(0.418)	(0.426)	(0.153)	
Nam Dinh	-0.0858	-0.221	-0.211	0.199	-0.0477	-0.429*	-0.0665	-0.0717		-0.244	0.0182
	(0.216)	(0.217)	(0.553)	(0.228)	(0.128)	(0.217)	(0.314)	(0.278)		(0.216)	(0.151)
Quang Ninh					0.0159			-0.252	-0.302**		
					(0.155)			(0.281)	(0.142)		
Thanh Hoa	-0.0116	-0.156	0.000263		0.0869	-0.498*	-0.122	-0.393	0.0134	-0.177	0.0866
	(0.176)	(0.218)	(0.266)		(0.208)	(0.296)	(0.233)	(0.254)	(0.197)	(0.163)	(0.116)
Nghe An	0.0398		0.163	0.411	0.0244	-0.409*	0.0924	0.292	-0.00140	-0.0408	
	(0.187)		(0.343)	(0.271)	(0.144)	(0.216)	(0.330)	(0.374)	(0.242)	(0.182)	
Ha Tinh	0.165	-0.424	0.986***		0.111	1.450***		0.0147	0.178	-0.155	0.0276
	(0.124)	(0.264)	(0.185)		(0.176)	(0.275)		(0.239)	(0.176)	(0.200)	(0.150)

cont.

VARIABLES	(1) Group 1	(2) Group 2	(3) Group 3	(4) Group 4	(5) Group 5	(6) Group 6	(7) Group 7	(8) Group 8	(9) Group 9	(10) Group 10	(11) Group 11
Thua Thien Hue	0.0706	-0.489	0.0147	0.368	0.0563	0.662***	0.167		-0.131	-0.159	-0.119
	(0.205)	(0.566)	(0.166)	(0.233)	(0.140)	(0.222)	(0.149)		(0.180)	(0.196)	(0.0958)
Da Nang (mun.)	-0.162	-0.300	-0.0521	0.199	-0.136	-0.609**	0.999*	-0.357	0.141	0.123	0.0589
	(0.149)	(0.387)	(0.272)	(0.215)	(0.142)	(0.255)	(0.578)	(0.298)	(0.187)	(0.348)	(0.128)
Quang Nam	0.0970		0.316*	2.511***	-0.0113	-0.535**	-0.214	-0.312			0.350***
	(0.107)		(0.184)	(0.561)	(0.119)	(0.222)	(0.275)	(0.237)			(0.118)
Quang Ngai	0.0835				-0.222		-0.00991	-0.240			0.0384
	(0.116)				(0.147)		(0.209)	(0.347)			(0.114)
Binh Dinh	0.126	-1.562***	0.132		-0.0308	-0.566**	-0.106	-0.0288	-0.0450		
	(0.161)	(0.354)	(0.186)		(0.124)	(0.227)	(0.174)	(0.412)	(0.197)		
Khanh Hoa	0.290**	0.557***	0.100		-0.0409	-0.577**	0.131		-0.276*	-0.00533	0.138
	(0.142)	(0.170)	(0.287)		(0.139)	(0.262)	(0.182)		(0.145)	(0.128)	(0.184)
Ho Chi Minh (mun.)	0.162	-0.121	0.243	0.439*	0.00556	-0.330	0.0786	-0.185	0.230	-0.145	0.108
	(0.130)	(0.279)	(0.190)	(0.221)	(0.122)	(0.235)	(0.162)	(0.252)	(0.172)	(0.128)	(0.108)
Binh Duong	-0.00684	0.552	0.374	0.368	0.0945	-0.471*	0.304	-0.330	0.611	0.0479	0.162
	(0.185)	(0.657)	(0.309)	(0.218)	(0.204)	(0.259)	(0.184)	(0.253)	(0.523)	(0.156)	(0.126)
Dong Nai	0.247	0.174	-0.287	-0.102	0.0547	-0.148	0.757	-0.0157	0.264	-0.120	0.308**
	(0.211)	(0.240)	(0.229)	(0.392)	(0.193)	(0.204)	(0.666)	(0.275)	(0.210)	(0.136)	(0.116)
Ba Ria - Vung Tau	0.339		0.0343	0.0844					0.517***		
	(0.281)		(0.276)	(0.249)					(0.177)		
Long An	-0.0168	0.0366	0.210	-0.0534	-0.176	-0.443*	0.0445	-0.196	0.0510	-0.163	
	(0.144)	(0.393)	(0.219)	(0.367)	(0.125)	(0.246)	(0.205)	(0.415)	(0.185)	(0.157)	
Dong Thap	0.482										-1.063***
	(0.391)										(0.108)
An Giang	0.246*		0.376**	0.214					0.172		
	(0.140)		(0.169)	(0.242)					(0.134)		
Tien Giang	0.269	-0.889	0.550*				-0.451**	-0.161	0.610**		
	(0.185)	(0.785)	(0.297)				(0.225)	(0.279)	(0.300)		
Can Tho (mun.)	0.342**		0.298	-0.0261		-0.378	-0.183	-0.172	-0.0515	-0.677	-0.0460
	(0.139)		(0.383)	(0.254)		(0.278)	(0.371)	(0.457)	(0.192)	(0.410)	(0.108)

cont.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
VARIABLES	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11
Survey											
2015	0.394***	0.476	0.497***	0.303	0.329**	-0.191	0.291*	0.324***	0.321***	0.407*	0.862***
	(0.0922)	(0.375)	(0.112)	(0.232)	(0.130)	(0.207)	(0.157)	(0.113)	(0.104)	(0.209)	(0.165)
Constant	1.731***	2.031*	2.057***	0.474	1.086**	1.923***	1.785***	0.881	1.097**	0.418	0.462
	(0.404)	(1.091)	(0.754)	(0.615)	(0.477)	(0.578)	(0.629)	(0.790)	(0.549)	(0.509)	(0.357)
Observations	237	55	181	40	159	75	134	188	149	104	54
R-squared	0.957	0.892	0.890	0.980	0.951	0.975	0.916	0.879	0.909	0.961	0.995

Note: Sector group definition: Group 1 = Food, beverages & tobacco (ISIC code 15,16); Group 2 = Textiles (17); Group 3 = Garments (18); Group 4 = Leather products (19); Group 5 = Wooden products & furniture (20, 36); Group 6 = Paper, printing & publishing (21 & 22); Group 7 = Refined petroleum products, chemicals, plastic & rubber (23 - 25); Group 8 = Non-metallic mineral products (26); Group 9 = Basic metal & Fabricated metal products (27 & 28); Group 10 = machinery, office, electronics & precision instruments (29 - 33); Group 11 = Transport machines (34).

The reference group for fixed effects is firms that are located in Hanoi (mun. cap.), and surveyed in 2005. The dependent variable is the log of revenues. Standard errors in parentheses. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B1: Temperature interpolation - MLR

	Temperature (^0C)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
ARIABLES	All	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Year 2004													
Latitude	-0.400***	-0.780***	-0.718***	-0.562***	-0.403***	-0.287***	-0.118***	-0.0853***	-0.0230***	-0.192***	-0.394***	-0.534***	-0.713***
	(0.00609)	(0.0195)	(0.0225)	(0.0154)	(0.0143)	(0.0125)	(0.00774)	(0.00746)	(0.00655)	(0.00794)	(0.00930)	(0.0149)	(0.0181)
Longitude	-0.0275***	-0.246***	-0.171***	-0.309***	-0.160***	0.0447***	0.159***	0.152***	0.154***	0.144***	0.0501***	-0.0169	-0.132***
	(0.00792)	(0.0236)	(0.0279)	(0.0235)	(0.0193)	(0.0160)	(0.0109)	(0.00955)	(0.00919)	(0.0113)	(0.0113)	(0.0185)	(0.0224)
Altitude (km)	-1.629***	-1.355***	-1.986***	-1.072***	-1.923***	-1.469***	-1.565***	-1.675***	-1.725***	-1.833***	-1.643***	-1.785***	-1.563***
	(0.0873)	(0.230)	(0.320)	(0.255)	(0.217)	(0.196)	(0.119)	(0.0977)	(0.0757)	(0.115)	(0.125)	(0.204)	(0.227)
Distance to sea (km)	-0.00659***	-0.00705***	-0.00283***	-0.00320***	-0.00441***	-0.00602***	-0.00736***	-0.00694***	-0.00804***	-0.00695***	-0.00829***	-0.00818***	-0.00959***
	(0.000264)	(0.000776)	(0.000963)	(0.000768)	(0.000626)	(0.000539)	(0.000346)	(0.000289)	(0.000251)	(0.000339)	(0.000359)	(0.000615)	(0.000742)
Constant	34.55***	60.31***	51.61***	66.06***	49.89***	26.98***	12.50***	12.71***	11.74***	14.52***	26.36***	34.61***	47.56***
	(0.844)	(2.523)	(2.943)	(2.496)	(2.054)	(1.697)	(1.155)	(0.998)	(0.958)	(1.208)	(1.211)	(1.967)	(2.396)
Observations	24,156	2,046	1,914	2,046	1,980	2,046	1,980	2,046	2,046	1,980	2,046	1,980	2,046
R-squared	0.561	0.796	0.688	0.666	0.698	0.697	0.808	0.825	0.849	0.847	0.883	0.817	0.822
Year 2014													
Latitude	-0.398***	-0.746***	-0.748***	-0.550***	-0.330***	-0.233***	-0.0969***	-0.0403***	-0.0600***	-0.122***	-0.310***	-0.551***	-1.005***
	(0.00663)	(0.0154)	(0.0235)	(0.0156)	(0.0105)	(0.0108)	(0.00662)	(0.00702)	(0.00701)	(0.00780)	(0.00874)	(0.0128)	(0.0160)
Longitude	-0.00549	-0.0804***	-0.262***	-0.233***	-0.121***	-0.00481	0.101***	0.154***	0.174***	0.140***	0.0995***	0.00682	-0.0608***
	(0.00859)	(0.0221)	(0.0313)	(0.0232)	(0.0152)	(0.0135)	(0.00979)	(0.0105)	(0.0101)	(0.0103)	(0.0110)	(0.0164)	(0.0207)
Altitude (km)	-1.253***	-0.880***	-0.912***	-1.239***	-0.961***	-0.953***	-1.451***	-1.683***	-1.845***	-1.805***	-1.728***	-1.294***	-0.262
	(0.0916)	(0.233)	(0.318)	(0.236)	(0.189)	(0.182)	(0.0905)	(0.0796)	(0.103)	(0.105)	(0.110)	(0.157)	(0.203)
Distance to sea (km)	-0.00614***	-0.00586***	-0.00279***	-0.00257***	-0.00491***	-0.00610***	-0.00815***	-0.00738***	-0.00703***	-0.00663***	-0.00742***	-0.00829***	-0.00636***
	(0.000285)	(0.000692)	(0.00103)	(0.000758)	(0.000500)	(0.000465)	(0.000283)	(0.000273)	(0.000312)	(0.000317)	(0.000346)	(0.000523)	(0.000635)
Constant	32.04***	40.55***	61.13***	57.69***	44.55***	31.92***	19.08***	11.95***	9.816***	14.17***	20.39***	32.52***	43.00***
	(0.914)	(2.353)	(3.358)	(2.478)	(1.632)	(1.442)	(1.028)	(1.095)	(1.061)	(1.082)	(1.172)	(1.766)	(2.202)
Observations	24,090	2,046	1,848	2,046	1,980	2,046	1,980	2,046	2,046	1,980	2,046	1,980	2,046
R-squared	0.497	0.803	0.651	0.670	0.691	0.674	0.845	0.820	0.819	0.828	0.874	0.854	0.861

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B2: Rainfall interpolation - MLR

	Monthly rainfall (mm); log												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
ARIABLES	All	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Double log regression													
Latitude; log	0.355***	2.058***	2.609***	1.172***	0.339***	0.220***	0.342***	0.241***	0.219***	-0.565***	-1.820***	-0.996***	0.441***
	(0.0178)	(0.0529)	(0.0447)	(0.0484)	(0.0318)	(0.0281)	(0.0357)	(0.0350)	(0.0329)	(0.0346)	(0.0390)	(0.0586)	(0.0563)
Longitude; log	4.012***	9.635***	4.169***	4.611***	5.638***	3.456***	3.613***	1.800***	0.399*	-2.497***	-4.698***	6.838***	15.17***
	(0.160)	(0.473)	(0.414)	(0.407)	(0.261)	(0.202)	(0.252)	(0.236)	(0.221)	(0.211)	(0.367)	(0.503)	(0.485)
Altitude (km); log	0.265***	0.620***	0.00366	-0.217***	-0.0401*	-0.204***	-0.0784***	0.317***	0.173***	-0.0885***	0.441***	1.167***	1.089***
	(0.0152)	(0.0460)	(0.0384)	(0.0401)	(0.0237)	(0.0200)	(0.0252)	(0.0230)	(0.0214)	(0.0227)	(0.0349)	(0.0517)	(0.0477)
Distance to sea (km); log	-0.0910***	-0.437***	-0.278***	-0.0678***	0.125***	0.135***	0.137***	0.100***	0.0784***	-0.0201***	-0.198***	-0.332***	-0.335***
	(0.00483)	(0.0147)	(0.0126)	(0.0129)	(0.00995)	(0.00823)	(0.00949)	(0.00937)	(0.00852)	(0.00761)	(0.0106)	(0.0151)	(0.0151)
Constant	-15.25***	-46.71***	-22.80***	-21.22***	-23.56***	-12.22***	-13.24***	-4.261***	2.532**	18.65***	32.46***	-24.41***	-68.22***
	(0.740)	(2.198)	(1.927)	(1.890)	(1.240)	(0.942)	(1.173)	(1.108)	(1.043)	(0.996)	(1.723)	(2.309)	(2.225)
Observations	217,440	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120
R-squared	0.011	0.198	0.209	0.055	0.049	0.040	0.041	0.047	0.033	0.045	0.175	0.090	0.157
Semi log regression													
Latitude	0.0232***	0.130***	0.174***	0.0794***	0.0278***	0.0174***	0.0226***	0.0175***	0.0120***	-0.0423***	-0.117***	-0.0649***	0.0230***
	(0.00111)	(0.00318)	(0.00261)	(0.00300)	(0.00188)	(0.00162)	(0.00206)	(0.00198)	(0.00189)	(0.00195)	(0.00241)	(0.00360)	(0.00339)
Longitude	0.0335***	0.0802***	-0.0110**	0.00906**	0.0329***	0.0210***	0.0310***	0.0100***	0.00191	-0.0183***	-0.0192***	0.0985***	0.166***
	(0.00180)	(0.00518)	(0.00445)	(0.00454)	(0.00278)	(0.00214)	(0.00281)	(0.00265)	(0.00249)	(0.00228)	(0.00391)	(0.00553)	(0.00513)
Altitude (km)	0.154***	0.356***	0.0713***	-0.0818***	-0.00726	-0.125***	-0.0775***	0.166***	0.0747***	-0.0481***	0.265***	0.665***	0.593***
	(0.00967)	(0.0299)	(0.0242)	(0.0248)	(0.0142)	(0.0123)	(0.0155)	(0.0138)	(0.0133)	(0.0142)	(0.0232)	(0.0337)	(0.0311)
Distance to sea (km)	-0.000586***	-0.00252***	-0.00285***	-0.00137***	8.14e-05	0.000406***	0.000719***	0.000413***	0.000436***	-1.03e-05	-0.000463***	-0.000833***	-0.00105***
	(3.86e-05)	(0.000113)	(9.54e-05)	(0.000101)	(5.85e-05)	(4.51e-05)	(5.51e-05)	(5.34e-05)	(5.10e-05)	(4.81e-05)	(7.85e-05)	(0.000115)	(0.000107)
Constant	0.184	-8.349***	1.303***	1.140**	0.300	2.560***	1.437***	3.881***	4.920***	7.986***	8.643***	-6.023***	-15.59***
	(0.186)	(0.537)	(0.464)	(0.469)	(0.294)	(0.220)	(0.291)	(0.278)	(0.262)	(0.241)	(0.411)	(0.564)	(0.521)
Observations	217,440	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120	18,120
R-squared	0.011	0.184	0.227	0.061	0.033	0.020	0.031	0.036	0.024	0.059	0.159	0.063	0.136

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C1: List of hydropower plants used to calculate IV

No.	Dam	Year of operation	Installed capacity (MW)	2005 Survey	2015 Survey
1	Da Nhim	1964	160	x	x
2	Thac Ba	1971	120	x	x
3	Tri An	1991	400	x	x
4	Hoa Binh	1994	1920	x	x
5	Vinh Son	1994	66	x	x
6	Thac Mo	1995	150	x	x
7	Song Hinh	2000	70	x	x
8	Yaly	2000	720	x	x
9	Da Mi	2001	175	x	x
10	Ham Thuan	2001	300	x	x
11	Can Don	2003	78	x	x
12	Sesan 3	2006	260		x
13	Quang Tri	2007	64		x
14	A Vuong	2008	210		x
15	Dai Ninh	2008	300		x
16	Tuyen Quang	2008	342		x
17	Binh Dien	2009	44		x
18	Buon Kuop	2009	280		x
19	Buon Tua Srah	2009	86		x
20	Plei Krong 1	2009	100		x
21	Sesan 4	2009	360		x
22	Song Ba Ha	2009	220		x
23	Song Con	2009	63		x
24	Ban Ve	2010	320		x
25	Cua Dat	2010	97		x
26	Huong Dien	2010	71		x
27	Son La	2010	2400		x
28	Song Tranh 2	2010	190		x
29	Srepok 3	2010	220		x
30	An Khe - Kanak	2011	173		x
31	Dak R'Tih	2011	144		x
32	Dong Nai 3	2011	180		x
33	A Luoi	2012	170		x
34	Bac Ha	2012	90		x
35	Dak Mi 4	2012	190		x
36	Dong Nai 4	2012	340		x
37	Ban Chat	2013	220		x
38	Hua Na	2013	180		x
39	Khe Bo	2013	100		x
40	Nam Chien	2013	200		x

Notes: x indicates corresponding plants used to construct IV for each survey.

Table C2: Hydropower generation model

VARIABLES	Average generation by month (MWh/day)
Installed capacity (MW)	6.187***
	(1.083)
Upstream installed capacity (MW)	1.383**
	(0.608)
Inflow to dam; SWAT simulation (m^3/s)	4.167***
	(0.458)
Inflow to dam; SWAT simulation squared (m^6/s^2)	-0.000204**
	(8.63e-05)
Observations	2,984
R-squared	0.878
Number of groups	40
Dam FE	Y

Driscoll-Kraay standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D1: Robustness check: Alternative IVs

	Survey 2005, $\rho = 1$						Survey 2015, $\rho = 10$					
VARIABLE	Baseline	Geodetic	Realproduction	Noupcap	7month	Timedummy	Baseline	Geodetic	Realproduction	Noupcap	7month	Timedummy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Performance												
Revenue; log												
Outage (Vol); log	-0.27	-0.30	0.37	-0.23	-0.36	-0.32	-0.73***	-0.73***	-0.51**	-0.49**	-0.73***	-0.68***
	(0.44)	(0.44)	(1.14)	(0.44)	(0.53)	(0.46)	(0.25)	(0.24)	(0.21)	(0.23)	(0.25)	(0.26)
Observations	1,132	1,132	1,132	1,132	1,132	1,132	915	915	915	915	915	915
First stage F-statistic	14.32	13.06	1.61	14.60	14.11	13.27	17.68	18.28	17.25	15.36	17.68	10.30
SW χ^2 test	15.46***	14.10***	1.74	15.76***	15.23***	14.33***	19.54***	20.20***	19.07***	16.98***	19.54***	11.38***
AR χ^2 test	0.45	0.58	0.15	0.32	0.54	0.55	5.67**	5.68**	3.01*	2.33	5.67**	4.13**
TFP YKL model; log												
Outage (Vol); log	-0.25	-0.27	-0.01	-0.24	-0.33	-0.27	-0.55**	-0.53**	-0.58***	-0.46**	-0.55**	-0.61*
	(0.22)	(0.22)	(0.39)	(0.21)	(0.27)	(0.23)	(0.23)	(0.23)	(0.21)	(0.19)	(0.23)	(0.37)
Observations	977	977	977	977	977	977	421	421	421	421	421	421
First stage F-statistic	9.59	8.92	2.94	9.90	8.22	8.95	3.62	3.67	3.90	3.86	3.62	1.58
SW χ^2 test	10.36***	9.63***	3.17*	10.69***	8.88***	9.67***	4.15**	4.21**	4.46**	4.42**	4.15**	1.81
AR χ^2 test	2.50	2.81*	0	2.35	3.10*	2.26	3.97**	3.88**	5.13**	3.14*	3.97**	2.57
TFP YKLM model; log												
Outage (Vol); log	-0.13*	-0.16**	0.04	-0.12*	-0.21**	-0.11	-0.53	-0.52	-0.33	-0.49*	-0.53	-0.57
	(0.07)	(0.07)	(0.23)	(0.07)	(0.08)	(0.09)	(0.41)	(0.41)	(0.30)	(0.28)	(0.41)	(0.65)
Observations	966	966	966	966	966	966	392	392	392	392	392	392
First stage F-statistic	9.01	8.43	2.97	9.26	7.85	8.46	2.71	2.76	2.45	2.65	2.71	1.03
SW χ^2 test	9.73***	9.10***	3.21*	10.01***	8.48***	9.14***	3.13*	3.18*	2.83*	3.05*	3.13*	1.19
AR χ^2 test	3.66*	5.96**	0.04	3.57*	7.07***	1.50	4.66**	4.64**	1.47	6.25**	4.66**	1.53
Panel B: Energy inputs												
Generator use; log												
Outage (Vol); log	0.35***	0.36***	-0.48	0.31***	0.22**	0.49**	0.31***	0.31***	0.28***	0.23**	0.31***	0.37***
	(0.11)	(0.10)	(0.86)	(0.09)	(0.09)	(0.20)	(0.10)	(0.09)	(0.10)	(0.10)	(0.10)	(0.12)
Observations	1,035	1,035	1,035	1,035	1,035	1,035	605	605	605	605	605	605
First stage F-statistic	11.38	10.69	1.58	11.56	10.34	10.40	10.14	10.56	10.73	8.58	10.14	8.81
SW χ^2 test	12.28***	11.54***	1.71	12.48***	11.16***	11.23***	11.34***	11.81***	12.00***	9.60***	11.34***	9.85***
AR χ^2 test	8.26***	8.39***	0.87	8.15***	4.84**	9.20***	6.49**	6.82***	5.17**	2.72*	6.49**	10.32***
Energy cost; log												
Outage (Vol); log	-0.76**	-0.78**	0.66	-0.73**	-0.95**	-0.97**						
	(0.33)	(0.33)	(1.57)	(0.33)	(0.42)	(0.40)						

(cont.)

	Survey 2005, $\rho = 1$						Survey 2015, $\rho = 10$					
VARIABLE	Baseline	Geodetic	Realproduction	Noupcap	7month	Timedummy	Baseline	Geodetic	Realproduction	Noupcap	7month	Timedummy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Observations	1,136	1,136	1,136	1,136	1,136	1,136						
First stage F-statistic	16.47	14.90	1.68	16.76	16.46	14.90						
SW χ^2 test	17.78***	16.09***	1.82	18.10***	17.77***	16.09***						
AR χ^2 test	7.63***	8.66***	0.30	7.26***	8.14***	9.74***						
Electric cost; log												
Outage (Vol); log							-1.03**	-1.04**	-1.06***	-1.11***	-1.03**	-1.15***
							(0.43)	(0.43)	(0.32)	(0.29)	(0.43)	(0.44)
Observations							841	841	841	841	841	841
First stage F-statistic							19.32	19.97	20.32	17.92	19.32	11.01
SW χ^2 test							21.44***	22.16***	22.55***	19.88***	21.44***	12.22***
AR χ^2 test							9.31***	9.63***	11.90***	15.51***	9.31***	9.17***
Fuel cost; log												
Outage (Vol); log							-0.62	-0.64	-0.64	-0.64	-0.62	-1.08
							(0.55)	(0.55)	(0.41)	(0.42)	(0.55)	(0.76)
Observations							458	458	458	458	458	458
First stage F-statistic							13.59	13.93	10.63	8.58	13.59	6.11
SW χ^2 test							15.57***	15.96***	12.18***	9.83***	15.57***	7.00***
AR χ^2 test							1.78	1.88	3.61*	3.81*	1.78	5.13**

Panel C: Other inputs

Material cost; log												
Outage (Vol); log	-0.27	-0.30	0.03	-0.23	-0.32	-0.32	-0.62	-0.60	-0.85*	-0.62*	-0.62	-0.67
	(0.70)	(0.68)	(0.92)	(0.69)	(0.84)	(0.74)	(0.40)	(0.39)	(0.46)	(0.38)	(0.40)	(0.50)
Observations	1,124	1,124	1,124	1,124	1,124	1,124	529	529	529	529	529	529
First stage F-statistic	13.63	12.43	1.62	13.84	13.36	12.74	8.85	9.13	8.91	7.49	8.85	6.30
SW χ^2 test	14.72***	13.43***	1.75	14.95***	14.43***	13.76***	10.03***	10.34***	10.09***	8.48***	10.03***	7.14***
AR χ^2 test	0.17	0.23	0	0.12	0.16	0.22	3.34*	3.22*	3.71*	2.97*	3.34*	2.25
Labor cost; log												
Outage (Vol); log	0.04	0.01	0.34	0.06	-0.01	0.02	-0.64	-0.63	-0.62	-0.72*	-0.64	-0.72
	(0.18)	(0.19)	(0.53)	(0.18)	(0.22)	(0.19)	(0.52)	(0.51)	(0.38)	(0.40)	(0.52)	(0.56)
Observations	1,129	1,129	1,129	1,129	1,129	1,129	845	845	845	845	845	845
First stage F-statistic	13.45	12.55	1.60	13.80	12.87	12.87	14.10	14.56	14.73	13.64	14.10	8.35
SW χ^2 test	14.52***	13.55***	1.73	14.90***	13.90***	13.90***	15.64***	16.16***	16.34***	15.13***	15.64***	9.26***
AR χ^2 test	0.04	0.01	0.73	0.08	0	0.01	2.75*	2.68	5.48**	8.12***	2.75*	3.41*

The regressions include control variables as those in Table 2, and dummies for sectors. The dependent variables are in bold. The endogenous variable is power outage volume (Vol). First stage F-statistic: Kleibergen-Paap rk Wald F-statistic that is robust to heteroskedasticity. Underidentification test: SW χ^2 test (Sanderson-Windmeijer first stage χ^2 test). Weak-instrument robust tests for significant endogenous variables : AR χ^2 test (Anderson-Rubin Wald χ^2 test). Robust standard errors clustered at the province level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table D2: Robustness check : Sale loss, and firm heterogeneity

Sample	First stage			Second stage					
	HAI(ρ)		*F-statistic*	*Outage (Vol); log*		*Obs*	*Cluster*	*SW χ^2 test*	*AR χ^2 test*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Survey 2005 (ρ = 1)									
Baseline (Whole sample)	-11.65***	(3.08)	14.32	-0.27	(0.44)	1,132	24	15.46***	0.45
Manufacturers	-11.67***	(3.18)	13.50	-0.28	(0.44)	1,119	24	14.52***	0.50
Generator	-12.44***	(3.58)	12.09	0.14	(0.42)	379	24	13.78***	0.09
Non-generator	-8.83**	(3.50)	6.38	-0.86	(0.53)	734	24	6.98***	4.14**
Elec-intensive	-11.28***	(3.40)	11.01	-0.42	(0.53)	874	24	11.86***	0.79
Non-Elec-intensive	-13.27***	(5.08)	6.82	-0.01	(0.42)	258	24	7.95***	0
Exporters	-17.90***	(4.71)	14.45	-0.39	(0.31)	524	24	16.10***	2.63
Non-exporters	-6.07**	(2.63)	5.32	0.37	(0.80)	608	24	5.92**	0.24
Survey 2015 (ρ = 10)									
Baseline (Whole sample)	-1.76***	(0.42)	17.68	-0.73***	(0.25)	915	19	19.54***	5.67**
Manufacturers	-1.94***	(0.55)	12.36	-0.88***	(0.31)	643	19	13.78***	8.05***
Generator	-1.47**	(0.72)	4.10	-1.48**	(0.71)	224	19	5.08**	8.14***
Non-generator	-2.06***	(0.57)	13.23	-0.90***	(0.25)	416	18	15.25***	10.84***
Elec-intensive	-1.53***	(0.55)	7.72	-1.32***	(0.45)	580	19	8.59***	12.57***
Non-Elec-intensive	-2.29***	(0.24)	94.74	-0.10	(0.29)	335	18	109.49***	0.10
Exporters	-2.83***	(0.80)	12.61	-0.82***	(0.21)	251	18	15.82***	16.60***
Non-exporters	-1.55***	(0.33)	22.36	-0.51*	(0.26)	664	19	25.12***	2.22

Note: The regressions replicate the baseline 2SLS estimates as those in Table 4 (first stage), and 5 (second stage) for various sub-samples. The dependent variable is revenues (log). The endogenous variable is power outage volume (*Vol*). **First stage F-statistic:** Kleibergen-Paap rk Wald F-statistic that is robust to heteroskedasticity. **Underidentification test:** SW χ^2 test (Sanderson-Windmeijer first stage χ^2 test). **Weak-instrument robust tests for significant endogenous variables** : AR χ^2 test (Anderson-Rubin Wald χ^2 test). Robust standard errors clustered at the province level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table D3: Robustness check: Survey 2015

		Handle fiscal year	Weights		
VARIABLE	Baseline	Add dummy	Weak	Median	Strong
	(1)	(2)	(3)	(4)	(5)
		Panel A: Performance			
Revenue; log					
Power outage; log	-0.73***	-0.55*	-1.13***	-1.03**	-1.03**
	(0.25)	(0.29)	(0.41)	(0.40)	(0.41)
Observations	915	915	915	915	915
First stage HAI(10)	-1.76***	-1.64***	-2.13***	-2.19***	-2.19***
	(0.42)	(0.41)	(0.49)	(0.46)	(0.47)
First stage F-statistic	17.68	15.98	18.91	22.67	22.02
SW χ^2 test	19.54***	17.68***	20.90***	25.06***	24.33***
AR χ^2 test	5.67**	3.71*	5.04**	4.51**	4.40**
TFP YKL model; log					
Power outage; log	-0.55**	-0.34	-0.15	-0.11	-0.09
	(0.23)	(0.27)	(0.18)	(0.17)	(0.17)
Observations	421	421	421	421	421
First stage HAI(10)	-1.53*	-1.60**	-4.24***	-4.37***	-4.55***
	(0.80)	(0.74)	(0.71)	(0.74)	(0.75)
First stage F-statistic	3.62	4.67	35.37	35.17	37.31
SW χ^2 test	4.15**	5.37**	40.52***	40.29***	42.74***
AR χ^2 test	3.97**	1.44	0.66	0.42	0.28
TFP YKLM model; log					
Power outage; log	-0.53	-0.45	-0.04	-0.04	-0.04
	(0.41)	(0.40)	(0.18)	(0.17)	(0.17)
Observations	392	392	392	392	392
First stage HAI(10)	-1.50	-1.43*	-4.44***	-4.56***	-4.75***
	(0.91)	(0.80)	(0.80)	(0.83)	(0.84)
First stage F-statistic	2.71	3.16	31.14	30.36	32.38
SW χ^2 test	3.13*	3.65*	35.90***	35.00***	37.33***
AR χ^2 test	4.66**	2.83*	0.07	0.05	0.04
		Panel B: Energy inputs			
Generator use; log					
Power outage; log	0.31***	0.21	0.19**	0.19**	0.19**
	(0.10)	(0.13)	(0.08)	(0.08)	(0.08)
Observations	605	605	605	605	605
First stage HAI(10)	-1.86***	-1.83***	-3.93***	-4.03***	-4.14***
	(0.58)	(0.58)	(0.55)	(0.53)	(0.52)
First stage F-statistic	10.14	10.11	52.04	57.51	62.60
SW χ^2 test	11.34***	11.33***	58.21***	64.32***	70.02***
AR χ^2 test	6.49**	2.13	4.52**	5.27**	5.56**
Electric cost; log					
Power outage; log	-1.03**	-0.84*	-0.87**	-0.91**	-0.90**

(cont.)

VARIABLE	Baseline	Handle fiscal year Add dummy	Weights Weak	Median	Strong
	(1)	(2)	(3)	(4)	(5)
	(0.43)	(0.45)	(0.40)	(0.39)	(0.38)
Observations	841	841	841	841	841
First stage HAI(10)	-1.85***	-1.76***	-2.12***	-2.18***	-2.18***
	(0.42)	(0.41)	(0.52)	(0.49)	(0.49)
First stage F-statistic	19.32	18.55	16.40	19.87	19.44
SW χ^2 test	21.44***	20.62***	18.20***	22.05***	21.57***
AR χ^2 test	9.31***	5.19**	7.20***	8.45***	8.28***
Fuel cost; log					
Power outage; log	-0.62	-0.71	0.24	0.23	0.23
	(0.55)	(0.61)	(0.23)	(0.23)	(0.23)
Observations	458	458	458	458	458
First stage HAI(10)	-2.42***	-2.33***	-4.48***	-4.55***	-4.60***
	(0.66)	(0.63)	(0.50)	(0.54)	(0.54)
First stage F-statistic	13.59	13.84	79.05	69.99	72.24
SW χ^2 test	15.57***	15.90***	90.58***	80.20***	82.77***
AR χ^2 test	1.78	1.88	0.98	0.98	0.95
		Panel C: Other inputs			
Material cost; log					
Power outage; log	-0.62	-0.43	-0.20	-0.23	-0.21
	(0.40)	(0.42)	(0.50)	(0.49)	(0.48)
Observations	529	529	529	529	529
First stage HAI(10)	-1.87***	-1.87***	-4.01***	-4.10***	-4.16***
	(0.63)	(0.58)	(0.76)	(0.76)	(0.76)
First stage F-statistic	8.85	10.39	27.81	28.91	29.56
SW χ^2 test	10.03***	11.79***	31.50***	32.75***	33.49***
AR χ^2 test	3.34*	1.29	0.17	0.23	0.20
Labor cost; log					
Power outage; log	-0.64	-0.74	-0.15	-0.16	-0.13
	(0.52)	(0.58)	(0.34)	(0.33)	(0.32)
Observations	845	845	845	845	845
First stage HAI(10)	-1.62***	-1.48***	-1.91***	-1.98***	-1.98***
	(0.43)	(0.42)	(0.44)	(0.41)	(0.41)
First stage F-statistic	14.10	12.33	18.76	23.57	22.70
SW χ^2 test	15.64***	13.70***	20.82***	26.15***	25.19***
AR χ^2 test	2.75*	2.70	0.23	0.26	0.19

The regressions include control variables as those in Table 2, and dummies for sectors. The dependent variables are in bold. The endogenous variable is power outage volume (Vol). First stage F-statistic: Kleibergen-Paap rk Wald F-statistic that is robust to heteroskedasticity. Underidentification test: SW χ^2 test (Sanderson-Windmeijer first stage χ^2 test). Weak-instrument robust tests for significant endogenous variables : AR χ^2 test (Anderson-Rubin Wald χ^2 test). Robust standard errors clustered at the provincea level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Bibliography

Ackerberg, D. A., Caves, K. and Frazer, G. (2015), 'Identification Properties of Recent Production Function Estimators', Econometrica **83**(6), 2411–2451.

Ackerberg, D., Lanier Benkard, C., Berry, S. and Pakes, A. (2007), Chapter 63 Econometric Tools for Analyzing Market Outcomes, Vol. 6, Elsevier, pp. 4171–4276.

Allcott, H., Collard-Wexler, A. and O'Connell, S. D. (2016), 'How Do Electricity Shortages Affect Industry? Evidence from India', American Economic Review **106**(3), 587–624.

De Loecker, J. (2011), 'Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity', Econometrica **79**(5), 1407–1451.

EAU-WB (2017), Firm Level Productivity Estimates: Methodological Note, Report, Enterprise Analysis Unit - World Bank.

Katayama, H., Lu, S. and Tybout, J. R. (2009), 'Firm-level productivity studies: Illusions and a solution', International Journal of Industrial Organization **27**(3), 403–413.

Kaur, S. (2014), 'Nominal Wage Rigidity in Village Labor Markets', National Bureau of Economic Research Working Paper Series **No. 20770**.

Levinsohn, J. and Petrin, A. (2003), 'Estimating Production Functions Using Inputs to Control for Unobservables', The Review of Economic Studies **70**(2), 317–341.

Li, J. and Heap, A. D. (2008), 'A Review of Spatial Interpolation Methods for Environmental Scientists', Geoscience Australia **Record 2008/23**.

Marschak, J. and Andrews, W. H. (1944), 'Random Simultaneous Equations and the Theory of Production', Econometrica **12**(3/4), 143–205.

Matsuura, K. and Willmott, C. J. (2009), 'README: Terrestrial Precipitation: 1900-2014 Gridded Monthly Time Series Version 4.01 '.

Maureen, D. A. and Jean, P. P. (2000), 'GIS-based construction of baseline climatologies for the Mediterranean using terrain variables', Climate Research **14**(2), 115–127.

Nalder, I. A. and Wein, R. W. (1998), 'Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest', Agricultural and Forest Meteorology **92**, 211–215.

Olley, G. S. and Pakes, A. (1996), 'The Dynamics of Productivity in the Telecommunications Equipment Industry', Econometrica **64**(6), 1263–1297.

Pavcnik, N. (2002), 'Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants', The Review of Economic Studies **69**(1), 245–276.

Sarsons, H. (2015), 'Rainfall and conflict: A cautionary tale', Journal of Development Economics **115**, 62–72.

Solow, R. M. (1957), 'Technical Change and the Aggregate Production Function', The Review of Economics and Statistics **39**(3), 312–320.

Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G. and McKendry, I. G. (2006), 'Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density', Agricultural and Forest Meteorology **139**(3-4), 224–236.

Van Beveren, I. (2012), 'Total factor productivity estimation: A practical review', Journal of Economic Surveys **26**(1), 98–128.

Wooldridge, J. M. (1999), 'Asymptotic Properties of Weighted M-estimators for variable probability samples', Econometrica **67**(6), 1385–1406.

Wooldridge, J. M. (2001), 'Asymptotic properties of weighted M-estimators for standard stratified samples', Econometric Theory **17**(2), 451–470.

Wooldridge, J. M. (2009), 'On estimating firm-level production functions using proxy variables to control for unobservables', Economics Letters **104**(3), 112–114.

Wooldridge, J. M. (2010), Econometric analysis of cross section and panel data, 2 edn, The MIT Press, Cambridge, Massachusetts, London, England.