The London School of Economics and Political Science

Electrification and Industrial Development in Indonesia

Dana Kassem

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, June 2018

To my parents.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of about 42,900 words.

Acknowledgements

The work in this thesis would not have been possible without the help and support of many people. I am intellectually indebted to my PhD advisors - Oriana Bandiera and Robin Burgess. Robin has been involved at every stage of this project from formalizing the research topic, to supporting grant applications, and bringing this project together at the end. I thank him for his unwavering high standards that motivated me to work harder. Oriana has been a great mentor and is an exceptional role model; I look up to her scholarly rigour, acute thinking, and thoughtfulness. I thank Pasquale Schiraldi, Gharad Bryan, Mark Schankerman, Greg Fischer, Alessandro Gavazza, and many others at the STICERD and the department for helpful discussions and comments. I thank Jane Dickson, Rhoda Frith, and Mark Wilbor for their support. I gratefully acknowledge financial support from the IGC, STICERD, and the Kuwait Programme at the LSE Middle East center.

I thank the PLN team who welcomed me into their office and made this project possible. I thank Suroso Isnandar for taking the time to answer my original request and arranging my visit. I thank Muhammad Ikbal Nur for allowing me to use PLN resources and for many helpful discussions. I thank Musa Partahi Marbun and Ahmad Yusuf Salile for supporting me with data collection and for being good friends.

I am grateful for the friendship of Pedro Alves, Michel Azulai, Kieu-Trang Nguyen, Roberto Sormani, Manuel Staab, Xuezhu Shi, and Panos Mavrokonstantis. Michel deserves special gratitude for long discussions that greatly improved some of the work in this thesis. I thank the alumna of the cool office, Giulia Zane, Junichi Yamazaki, and Shiu Bo for many research discussions and even more pints. Daniel Osorio Rodriguez and Oliver Pardo were great friends during the early years. When writing these acknowledgments, I remember with sadness Rui Zhang, who left us too early.

My PhD journey was enriching both intellectually and on a personal level. But struggling with anxiety made my experience less pleasant. I would not have been able to make it without my very solid and reliable support system. I thank Arthur Seibold, my partner in life and in academia, for being there through all the midday and midnight panic attacks, all so selflessly, even when he had his own worries. I thank my best friend, Anna Youssef, who was always there for me on endless Skype calls, reminding me that there is more to life, and making me laugh when I was feeling low. Finally, my deepest gratitude is to my parents. I thank them for always being patient, loving, and proud. I thank them for providing me with opportunities throughout my whole life, and prioritizing my education over everything else. Without them I wouldn't be where I am today.

Abstract

Economists and policymakers have long believed that access to electricity is essential for industrial development, and ultimately growth. Despite this consensus, there is limited evidence of this relationship. In this thesis, I ask whether electrification causes industrial development. I study the effect of the extensive margin of electrification (grid expansion) on the extensive margin of industrial development (firm entry and exit). I combine newly digitized data from the Indonesian state electricity company with rich manufacturing census data. To deal with endogenous grid placement, I build a hypothetical transmission grid based on colonial incumbent infrastructure and geography. The main instrumental variable is the distance to this hypothetical grid. I examine the effect of electrification on local industrial development. To understand when and how electrification can cause industrial development, I shed light on an important economic mechanism - firm turnover. I find that electrification causes industrial development, represented by an increase in the number of manufacturing firms, manufacturing workers, and output. Electrification increases firm entry rates, but also exit rates. Overall, electrification creates new industrial activity, as opposed to reorganizing it across space. I then evaluate the impact of electrification on firm-level performance. I find that connected firms are larger, more likely to exit, and younger. This is consistent with higher turnover at the market level. I look at the implications of the previous results on industry productivity. Higher turnover rates lead to higher average productivity and induce reallocation towards more productive firms. This is consistent with electrification lowering entry costs, increasing competition and forcing unproductive firms to exit more often. Without the possibility of entry or competitive effects of entry, the effects of electrification are likely to be smaller. I use detailed product-level production data to structurally estimate a quantity-based production function, which when combined with price data, allows me to estimate marginal cost. Electrification substantially reduces the cost of production of existing products and their prices. While mark-ups don't change for incumbent firm-product pairs, the average mark-up increases in the market. This is due to a selection effect where products produced post access have higher mark-ups. These products are "new" and are more likely to be differentiated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The idea that electrification causes industrial development dates back as far as Lenin[1]. Even today, many governments and aid agencies[2] invest in energy infrastructure projects, especially in developing countries. In 2017, the Indonesian government invested around $1.8 billion in electricity, 7% out of its total budget for infrastructure. The Kenyan government is currently investing $2.1 billion in the grid expansion to rural areas. The Kenyan policymakers expect this investment "to enhance industrialization and emergence of [...] industries". There is consensus among policymakers that access to electricity is an essential ingredient for industrial development, which is considered a fundamental driver of growth.

However, recent economic evidence shows that the benefits of electrification are not as large as previously thought[3]. If public funds are limited, this presents an argument against investing in energy infrastructure and instead in favor of allocating funds to other types of public expenditure such as health or education. In fact, electrification in various African countries has increased substantially over the last decades, but these countries have not witnessed industrial development. So in this thesis I ask, does electrification cause industrial development? Or do these investments have little impact on the pace of industrial development?

To answer this question, I use a rapid, government-led grid expansion during a pe-

[1] Lenin (1920)"*Communism is Soviet power plus the electrification of the whole country.*" Lenin believed that electrification would transform Russia from a *"small-peasant basis into a large-scale industrial basis"*

[2] The World Bank has committed to lending $6.3 billion to the Energy and Mining sector worldwide. From *The World Bank Annual Report 2017*, http://www.worldbank.org/en/about/annual-report.

[3] e.g. Lee, Miguel, and Wolfram (2016) who focus on residential electrification

riod of rapid industrialization in Indonesia. I travelled multiple times to Indonesia and put together a comprehensive data-set covering a period of 11 years from 1990 to 2000 from various current and historical sources. I first map the expansion of the electric transmission grid over time and space in Java, the main island in Indonesia. I then map manufacturing activity in 25,000 administrative areas for more than 29,000 unique firm observations in Java, where 80% of Indonesian manufacturing firms are located. These data allow me to understand *when* and *how* electrification affects industrial development.

The research in this thesis is the first to examine the effect of the extensive margin of electrification (grid expansion) on the extensive margin of industrial development (firm entry and exit). The effect of the extensive margin of electrification, i.e. extending the electric grid to new locations, has been studied on employment (Dinkelman (2011)) and general development-level indices (Lipscomb, Mobarak, and Barham (2013)). Other papers have estimated the demand and cost of rural electrification for households in a controlled environment (Lee, Miguel, and Wolfram (2016)). The link between electrification and firms has been studied on the intensive margin and is mostly focused on the effect of shortages on firm outcomes (e.g. Allcott, Collard-Wexler, and O'Connell (2016)). Variation in shortages creates short-run firm responses by affecting the input price of electricity which in turn affects the firm's production decision on the intensive margin. The evidence on the intensive margin of electrification and industrial development is important, but the effect of the extensive margin of electrification on industrialization is potentially different, and of greater relevance to those interested in long run development. Changes on the extensive margin of electrification, meaning whether the firm can be connected to the electric grid or not, can create long-run firm responses by affecting the extensive margin of firm decisions, namely, entry and exit.

An economic mechanism through which electrification potentially affects industrial development is therefore firm turnover, driven by the entry and exit of firms. Electrifying a new location can influence firms' entry and exit decisions in that particular location. This changes the composition of firms in the market, and hence, average productivity. Whether or not electrification enhances or decreases manufacturing productivity is therefore a question that requires empirical verification.

Indonesia is an appropriate setting to answer this research question. For historical reasons, the Indonesian power sector remained underdeveloped compared to countries with a similar GDP[4]. In 1990, Java, the most developed and densely populated island in Indonesia, was only around 40% electrified. The island has since witnessed a massive and successful government-led effort to expand access to electricity up until the year 2000. During that period, transmission capacity in Java quadrupled and electrification ratios increased to more than 90%. At the same time, Indonesia experienced fast growth in the manufacturing sector. This allows me to match modern type firm-level micro data with sufficient recent variation in access to the grid to detailed data on the electrification infrastructure.

Establishing a causal link between electrification and industrial development is empirically challenging. In any emerging economy, infrastructure and industrialization occur simultaneously, and separating demand-side from supply-side factors is difficult. This poses an empirical challenge in identifying the effect of electrification on industrial outcomes. The empirical strategy I implement in this thesis tries to make progress on this issue by using an instrumental variable strategy inspired by the transportation infrastructure literature[5]. I exploit a supply-side natural experiment based on the need of the state electricity monopoly to have a single interconnected electricity grid in Java. I construct a hypothetical interconnected electric transmission grid that is a function of incumbent disconnected electrification infrastructure built by Dutch colonial electric utilities and geographic cost factors. The hypothetical grid abstracts from endogenous demand factors that could be driving the expansion of the grid and focuses on cost factors only. The use of the colonial infrastructure also means that the incumbent infrastructure is unlikely to be correlated with economic forces in 1990. Distance to the hypothetical grid is used to instrument for endogenous access to electricity, conditional on various controls, including other types of infrastructure. A second empirical challenge that is less discussed in the literature is a violation of the Stable Unit Treatment Value Assumption (SUTVA). SUTVA requires that the treatment of one unit does not affect the outcome of other units, in other words, no spillovers or general equilibrium effects. In the context of this paper, this means that electrifying one location should not affect the industrial outcomes of other locations. I address this issue by conducting various empirical tests for general equilibrium effects.

The data-sets used in this paper come from various sources. I collected and digitized

[4] McCawley (1978)

[5] For example, see Banerjee, Duflo, and Qian (2012), Chandra and Thompson (2000), Redding and Turner (2014) and Faber (2014)

spatial data on the electrification infrastructure from the Indonesian state electricity monopoly Perusahaan Listrik Negara (PLN) in Jakarta. This includes data on the location, operation year, and capacity of power plants and transmission substations. To build a time-series, I use administrative documents from PLN. Gaps are then filled from World Bank loan reports from 1969 to 1992. I then construct measures of access to the grid based on the distance from the centroid of a desa to the nearest transmission substation. A desa is the lowest administrative division in Indonesia. To study firm turnover, I construct yearly maps of manufacturing activity in Java, which includes the number of firms, manufacturing output, number of manufacturing workers, and entry and exit rates in any desa in Java. The information on manufacturing activity at the desa level comes form the Indonesian annual manufacturing census 1990-2000. This is a census of Indonesian manufacturing firms with 20 or more employees. The firm-level data is also used to get information on firm output, inputs, exit and entry decisions, as well as to get estimates of revenue productivity. I complement the firm-level data with product-level data where I observe product prices. These data allow me to estimate physical productivity, marginal cost, and mark-up. Together with revenue productivity, these variables will allow me to look at the effect of electrification on different measures of productivity. I then combine productivity estimates with firm market share data to study the effect of electrification on reallocation at an aggregate industry level.

This paper contributes to the literature on infrastructure and development. A strand of literature examines the effect of different types of infrastructure on economic outcomes. These include the effect of dams on agricultural productivity and poverty (Duflo and Pande (2007)), and the effect of transportation (roads, railways, highways) infrastructure on regional economic outcomes (examples include Donaldson (2010), Banerjee, Duflo, and Qian (2012), Faber (2014), Donaldson and Hornbeck (2016), and Gertler, Gonzalez-Navarro, Gracner, and Rothenberg (2014)). In terms of electrification infrastructure, a growing literature studies generally the relationship between energy and development. Ryan (2017) studies the effect of expanding the transmission infrastructure on the competitiveness on the Indian electricity market. In another paper, Ryan (2018) experimentally investigates the relationship between energy productivity and energy demand among Indian manufacturing plants. A subset of the literature evaluates the effects of grid expansion as in Dinkelman (2011) who estimates the effect of electrification on employment in South Africa and Lipscomb, Mobarak, and Barham (2013) where they look at the effect of electrification in Brazil. Rud (2012) looks at the effect of electrification on industrialization in India at the state level. He shows that industrial output in a state increases with electrification.

While these papers focus on the extensive margin of electricity supply, many papers study the relationship between electricity supply and firms on the intensive margin, i.e. shortages. Reinikka and Svensson (1999) show that unreliable power supply in Uganda reduces private investment productivity by forcing firms to invest in generators and other low-productivity substitutes for reliable public provision of power. Fisher-Vanden, Mansur, and Wang (2015) use Chinese firm-level panel data to examine the response of firms to power shortages. They find that firms respond by re-optimizing among inputs, which increases their unit cost of production but allows them to avoid substantial productivity losses. Allcott, Collard-Wexler, and O'Connell (2016) find that electricity shortages in India reduce revenue but have no effect on revenue productivity.

Another strand of literature this paper is related to is the one on productivity and firm dynamics. Many papers study the determinants of firm turnover and its role in reallocating resources from less productive to more productive firms (examples include Syverson (2004), Syverson (2007), Foster, Haltiwanger, and Syverson (2008), Bartelsman, Haltiwanger, and Scarpetta (2013), Nguyen (2014)). An extensive literature as in Tybout (2000), Hsieh and Klenow (2009), and Bloom, Mahajan, McKenzie, and Roberts (2010), aims at explaining the productivity gap between firms in developing countries and firms in developed countries. These differences in productivity across countries imply substantial differences in aggregate performance. Infrastructure is one suggested explanation to the lower productivity level of firms in developing countries, in particular, access to electricity. I contribute to this literature in this thesis by linking infrastructure to reallocation and turnover in explaining the low productivity of firms in developing countries.

My results show that electrification causes industrial development at a local level by increasing manufacturing activity in desas. Access to the grid increases the number of firms, number of workers in manufacturing, and manufacturing output. Interestingly, electrification increases firm turnover by increasing not only entry rates, but also exit rate.

At the firm level, I find that electrification causes average firm size to increase, both in terms of how much output the firm produces and how much inputs it demands.

The results on firm turnover are confirmed in the firm-level analysis. Electrification increases the probability of exit, making it harder for inefficient firms to survive. In addition, electrification shifts the firm age distribution towards younger firms. This is a sign of churning in the industry, created by increased entry (more young firms) and increased exit (firms die more often).

At both the desa-level and the firm-level, I test for general equilibrium effects and I find that electrification does indeed create new industrial activity, as opposed to only relocating economic activity from non-electrified areas to electrified areas. This implies that there are no major violations of SUTVA in this particular setting.

I also find that electrification increases average productivity, consistent with higher firm turnover. I use a decomposition of an aggregate revenue-weighted average productivity following Olley and Pakes (1996). I find that electrification increases allocative efficiency where the covariance between firm productivity and market shares is higher in electrified areas. These results are theoretically consistent with a decrease in the entry cost, suggesting that electrification increases aggregate productivity by allowing more productive firms in the market, increasing firm turnover, and enhancing allocative efficiency.

Finally, I find that electrification decreases average marginal cost of production. Incumbent firm-product pairs experience a substantial decrease in the marginal cost of production which they completely pass through to consumers. However, firm-product pairs that are selected into the market by electrification charge a higher mark-up, possibly because electrification allows firms to produce more differentiated products.

This thesis is formed of 7 chapters. Chapter 2 introduces the new data on the Indonesian electrification infrastructure and lays out the empirical strategy I follow in the rest of the thesis. Chapter 3 presents evidence on the effect of electrification on local industrial outcomes and investigates how electrification affects the organization of industrial activity across space. In chapter 4, I evaluate how electrification affects the performance and survival of firms. Chapter 5 examines the implications of electrification on industry productivity and reallocation. Chapter 6 explores how electrification affects the firm's cost structure and market power. Chapter 7 concludes.

Chapter 2

New Data on the Indonesian Electrification Infrastructure and Empirical Strategy

2.1 Introduction

In the first chapter of this thesis, I present the data and the empirical strategy that are common to all the subsequent chapters. A significant part of the work for this thesis involved collecting new data on the electrification infrastructure in Indonesia. In what follows, I introduce these data in detail and present the various datasets that I have constructed for the analysis. I highlight how these data are useful for understanding the causal effect of electrification and the economic mechanism in play.

The credibility of any empirical work relies on the validity of the empirical strategy used to identify causal effects. I design an instrumental variable approach based on a hypothetical transmission grid that abstracts from endogenous demand factors and focuses on cost factors. Here I present my empirical strategy and my identification assumption, and finally provide a couple of tests for its validity.

I start by providing a brief institutional background to the Indonesian power sector, going back to colonial Indonesia and ending with the collapse of the Suharto regime. I then present the new data I have collected for this project. Finally, I describe the main empirical strategy. These data and empirical strategy form the basis of the analysis that follows in the later chapters.

2.2 Institutional Background

2.2.1 History of the Indonesian Power Sector

Knowing the historical context of the power sector in Indonesia is crucial to understand why the Indonesian electricity supply was underdeveloped, including in Java. During the period of Dutch colonization of Indonesia, access to electricity was unequal and mainly reserved to colonial establishments. Between 1953 and 1957 the three Dutch owned electric utilities in Indonesia were nationalized by the Government. Perusahaan Listrik Negara (PLN), the Indonesian state electricity monopoly, became fully responsible for generating, transmitting and distributing electricity in Indonesia, and still is until today. The transfer was not friendly, and was without a transition period where the new Indonesian management could have been trained by its colonial predecessors and many documents were destroyed in the process. Political unrest, lack of funds, hyperinflation and the lack of qualified management and engineers led to a period of decline in efficiency, poor operating conditions, and inadequate expansion (McCawley (1971)). This in turn led to a large electric supply deficit, which meant low household electrification ratios and that businesses and industries had to rely on self-generation. Power supply in Indonesia was poor even relative to other countries with a similar GDP per capita. To put things into perspective, in 1975, Indonesian GDP per capital was around $216, higher than the GDP per capita in India of $162. However, in the same year, electricity production per capita in Indonesia was only about one-fifth the level in India (McCawley (1978)). Over the next decades, with the help of various international aid agencies, PLN was expanding steadily both in terms of physical and human capital.

2.2.2 Objective of the Government of Indonesia 1990-2000

The main sources of electricity supply in Indonesia in the late 1980s and early 1990s comprised of PLN, the state electricity monopoly, and self-generation (around 40% of generating capacity), mainly by the manufacturing sector. As Indonesia was witnessing an expansion of the PLN generation capacity, the manufacturing sector was shifting from relying exclusively on self-generation towards the use of captive generation for solely on a stand-by basis. Trends in PLN sales and captive power suggested that manufacturing firms, even after incurring the sunk cost of acquiring a generator, prefer grid electricity. This suggests that the marginal price of electricity from the grid is lower that the marginal price of electricity from self-generation. In 1989, the level of electricity consumption per capital was still low in Indonesia

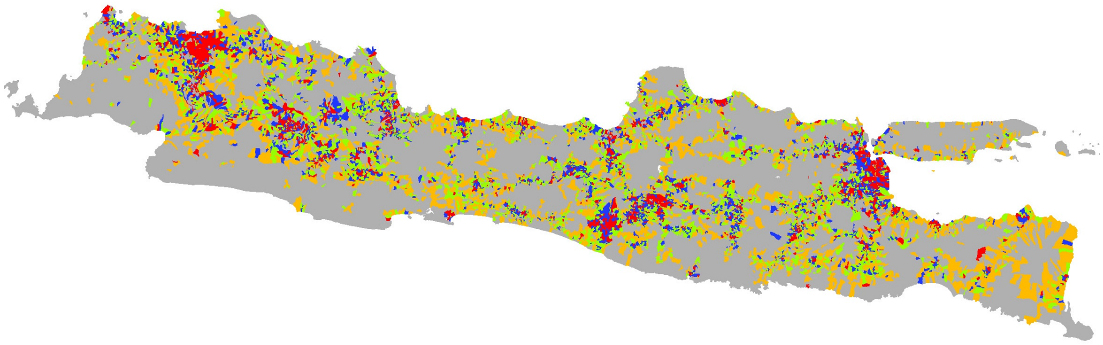
(137.5 kWh) relative to other countries at the same development level and its neighbours (Malaysia 1,076 kWh, India 257 kWh, Philippines 361 kWh, and Thailand 614 kWh.)[1].

This low level of electricity consumption was due to the lack of supply facilities. PLN's investment program in the late eighties was designed to meet the goals set by the Government's Five-Year Development Program (REPELITA V) by 1994. These included a 75% electrification ratio in urban areas, 29% electrification ratio overall, and finally, the substitution of 80% of captive generation by the industrial sector. The objective of the Government at that time was to replace self-generation, i.e. providing grid electricity to non-connected incumbents, as opposed to expanding the grid to industrialize new locations. The subsequent Five-Year Development Program (REPELITA VI 1994-1999) by the Indonesian government had the following objectives for the power sector: (i) provide adequate, reliable, and reasonably priced supply of energy to rapidly growing economy, (ii) conserve and diversify the sources of energy, and (iii) minimize social and environmental adverse impacts. Goal (i) illustrates the simultaneity problem of growing adequate infrastructure provision and economic growth[2]. The government of Indonesia was investing heavily in electricity supply to keep up with a rapidly growing economy, which poses the empirical challenge of identifying the causal effect of the expansion of electricity supply on industrial development. In 1997, the Asian financial crisis hit, followed by the end of the Suharto dictatorship and political unrest, which all led to a lack of funds. Investment in the power sector continued during that period, albeit at a slower pace. By 2000, more than 90% of firms Java had access to electricity.

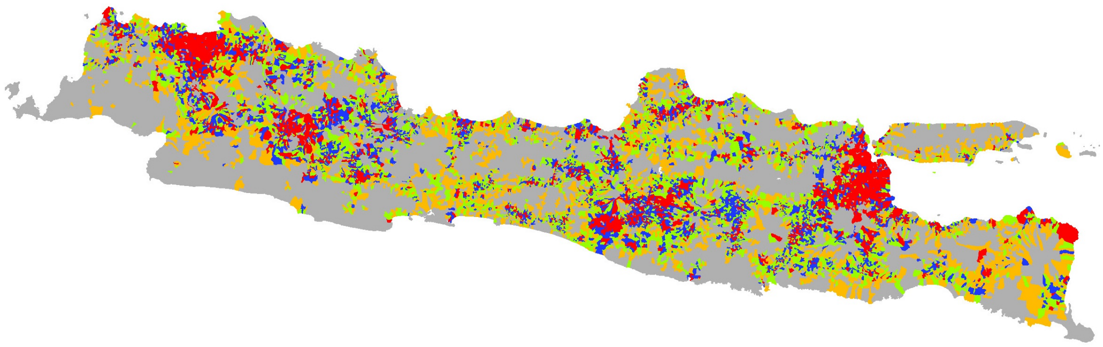
Figure 2.1 presents the dramatic increase in electrification ratios in Java during the sample period. Figure 2.1a shows the spatial distribution of electrification ratios in Java in 1990. Electricity was mostly concentrated in the capital city of Java, Jakarata, but also the cities Bandung, Yogyajakarta, and Surabaya. The expansion of electricity over time can be seen in the increase electrification ratios in 1993 (figure 2.1b), 1996 (figure 2.1c), and finally in the year 2000 (figure 2.1d), when most of Java was fully electrified.

[1] Source: IEA Statistics 2014

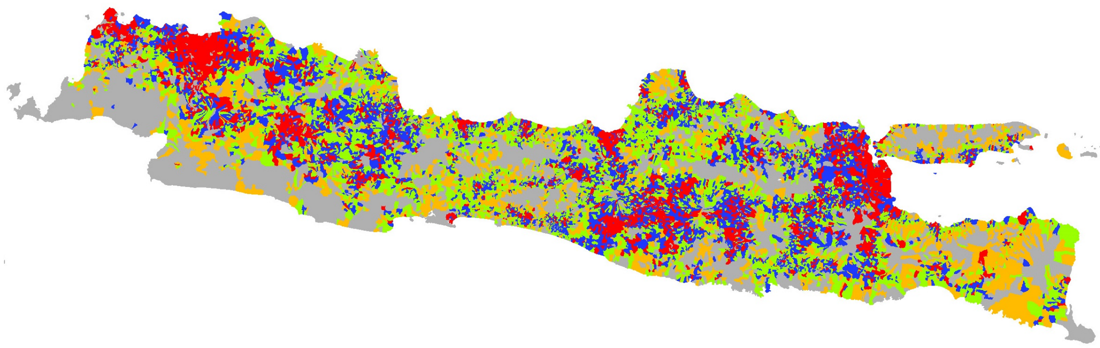
[2] Source: Official planning documents.



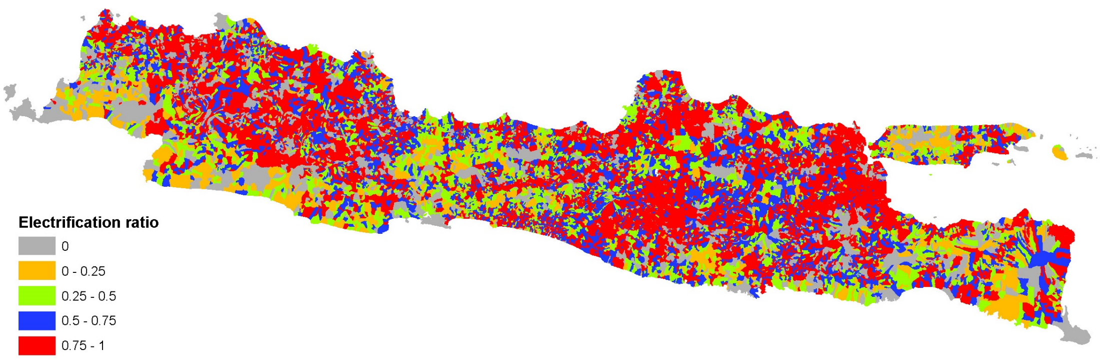
(a) 1990



(b) 1993



(c) 1996



(d) 2000

Figure 2.1: Desa-Level Electrification Ratios 1990 to 2000.
Source: PODES, BPS

2.3 New Data on Electrification in Java, 1990-2000

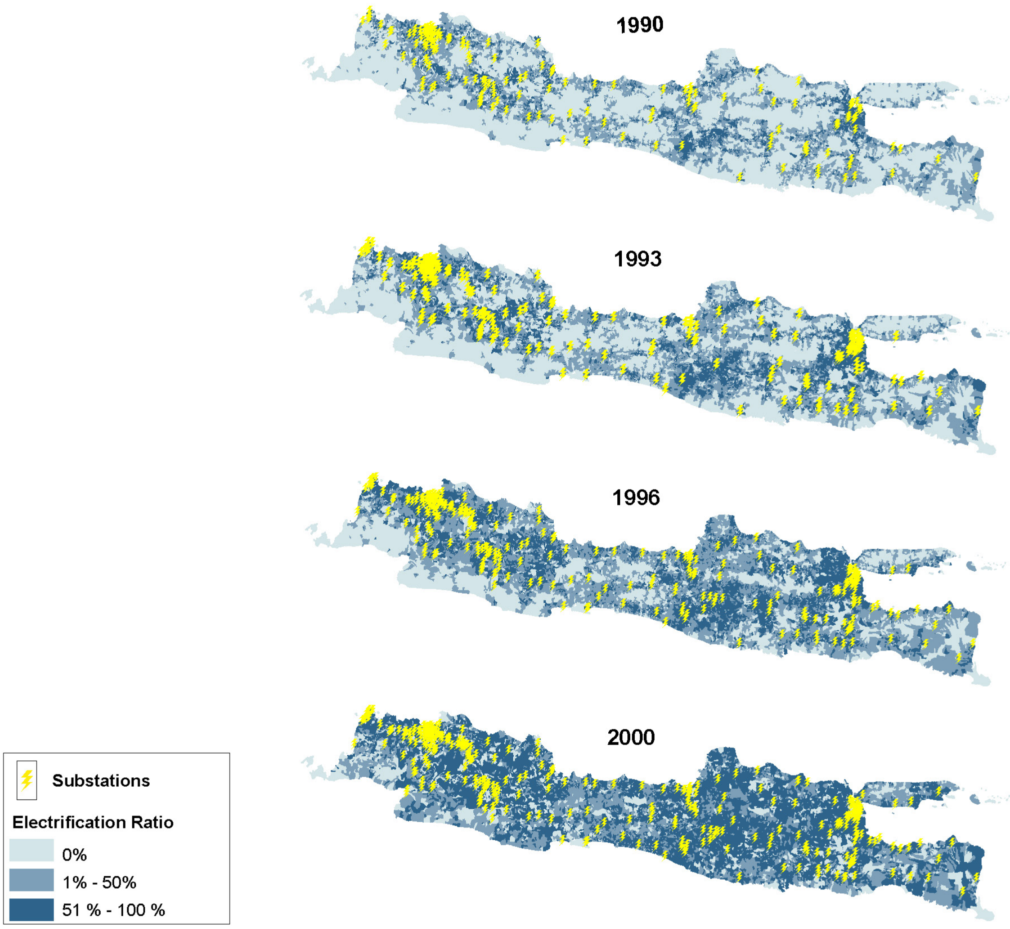
In order to evaluate the impact of electrification on industrial development in Java, I have constructed a new panel data-set on 24,824 Javanese desas, the lowest administrative division in Indonesia. The data-set follows these desas annually from 1990 to 2000, a period during which electrification in Java increased from 40% to almost 100% as can be seen in figure 2.1.

I start by constructing a time-series of the electricity transmission network in Java between 1990 and 2000 using data from various sources. Java is the most dense island in Indonesia with 60% of the population and 80% of manufacturing firms[3]. I travelled multiple times to Jakarta, and I spent a considerable amount of time and resources collecting and digitizing data from current and historical administrative records from PLN. I digitized information on the location, capacity and operation date of equipment within power plants and transmission substations in Java from the PLN Head Office in Jakarta. The main sources of the raw data are (i) inventory tables of transmission transformers within each transmission substation (see figure 2.13 in appendix 2.6 section 2.A.), and (ii) maps (digital, for example figure 2.14 in the appendix and paper figures 2.3 and 2.4) of the transmission network in Java.

To build the time-series from 1990 to 2000, gaps in administrative data were filled using World Bank power project reports, which evaluate electricity infrastructure loans given by the World Bank to Indonesian government between 1969 and 1996. In addition, because location data from PLN is not always accurate, I manually cross-checked power plant and substation coordinates using data downloaded from OSM (Open Street Maps). The resulting data-set is a panel of all transmission substations in Java. Figure 2.2 shows the expansion of the grid during the sample period where the yellow bolts represent transmission substations.

[3] Source: author's calculations.

Figure 2.2: Expansion of the Grid 1990-2000



The expansion of the transmission grid in Java during that period was rapid and substantial as shown by the summary statistics in table 2.1. In 1990, the number of substations was 115. By 2000, there was a total of 279 transmission substations in Java. Total electricity transmission capacity increased from 6620 MVA to 25061 MVA, almost 4 times.

Table 2.1: Summary statistics: Electrification Infrastructure

Variable	**1990**	**2000**
Number of Substations	115	279
Total Capacity(MVA)	6619.58	25061.28

2.4 Industrial Outcomes

There are multiple units of analysis. I start my empirical analysis by looking at the effect of access on desa-level manufacturing outcomes. A desa is the lowest administrative division in Indonesia[4]. Data on desa level boundaries were acquired from BIG, the Indonesian National Mapping Agency. To get information on manufacturing activity in these desas, I use the Indonesian annual census of all manufacturing firms in Indonesia with 20 or more employees, where I observe in which desa each firm is located. I restrict the analysis to firms located in Java, which constitute around 80% of all Medium and Large firms in Indonesia. This allows me to create variables such as the number of manufacturing firms, number of manufacturing workers and total manufacturing output in each desa. The resulting data-set is a yearly balanced panel of all desas in Java from 1990 to 2000. Table 2.2 presents some summary statistics of these desas. On average, around 60% have access to the grid over the sample period. The average number of medium or large firms per desa is less than 1. However, the median is 0. This shows that most desas in fact have zero manufacturing firms since I include all the desas in Java in the sample regardless of whether it has any manufacturing firms or not. The sample of desas includes all the administrative divisions that cover the island of Java, and these could be urban, rural, residential, and so on. Conditional on having a positive number of firms, the average number of firms per desa is around 4 firms. The last three rows of table 2.2 show that there is substantial variation on how large these desas are in terms of population and area. The final total number of desas per year used in the analysis is around $24,000$[5].

[4] There are 4 administrative divisions in Indonesia: province, regency, district and desa.

[5] Some desas were excluded as part of the identification strategy. See section 2.5 for more details.

Table 2.2: Desa-Level Summary Statistics

Variable	Mean	Median	Min	Max
Access	0.58	1	0	1
Number of firms	0.9	0	0	204
Number of firms > 0	4.2	2	1	204
Area (km^2)	5.7	4.3	1	540
Population	4,500	3,332	36	800,000
Pop. Density (per km^2)	2,548	1,451	7.7	36,413
Number of desas		23,770		

I use information from the Desa Potential Statistics (PODES) survey for 1990, 1993, 1996 and 2000. The PODES data-set contains on all Indonesian desas, which I use to get data on desa level characteristics such as population, political status, legal status and most importantly, various infrastructure variables. These include information on the type of infrastructure available in the desa such as railway, motor station, river pier, and airport. In addtion, I use GIS data on cities, waterways, coastline and roads in Java. I measure the distance from each desa (centroid) to each of these geographic features in addition to the nearest electric substation and the hypothetical least cost grid. I also use data on elevation to measure land gradient at each location. This data is used to construct a digital map of desas in Java with various desa-level characteristics over time.

I then take advantage of the richness of information in the firm-level data from the census of manufacturing and analyze the effect of access to electricity on firm-level outcomes. Table 2.3 shows the distribution of firms across industries and access ratios in 1990 and 2000. The industries are ordered by the number of firms in that industry, giving a clear picture of the Indonesian manufacturing sector. The largest five industries are food and beverages, textiles, non-metallic mineral products (e.g. cement, clay, etc..), wearing apparel, and furniture, forming 60% of the manufacturing sector in Java. Between 1990 and 2000, the total number of manufacturing firms in Java has increased by almost 50%. Columns (3) and (4) show the access ratio in 1990 and 2000, respectively. There has been an increase in the access ratio in almost all industries to varying degrees. The only industry that witnessed a decrease in the access ratio is furniture, but that can be explained by the massive entry to the furniture sector, where the number of firms tripled over the decade.

Table 2.3: Industry-Level Summary Statistics

	Observations		Access	
	(1)	(2)	(3)	(4)
Industry	1990	2000	1990	2000
Food and beverages	**2,035**	**2,817**	**0.63**	**0.86**
Textiles	**1,356**	**1,600**	**0.69**	**0.92**
Non-metallic products	**947**	**1,413**	**0.71**	**0.91**
Wearing Apparel, fur	**864**	**1,325**	**0.75**	**0.90**
Furniture	**578**	**1,380**	**0.77**	**0.74**
Rubber and plastic	591	867	0.85	0.96
Tobacco products	812	691	0.22	0.83
Chemicals	524	745	0.90	0.92
Wood products	314	653	0.78	0.88
Fabricated metals	315	612	0.87	0.98
Leather and footwear	239	415	0.87	0.99
Printing and publishing	237	272	0.83	0.99
Machinery and equipment	158	246	0.82	1.00
Paper products	132	301	0.83	0.99
Electrical machinery	131	174	0.99	1.00
Motor Vehicles	121	168	0.91	1.00
Other Transport	106	142	0.55	0.99
Basic metals	76	155	0.96	1.00
Radio, TV equipment	58	112	0.97	0.99
Medical equipment	34	40	0.88	1.00
Coke, petroleum, fuel	2	19	1.00	0.95
Mean	1002	1356	0.70	0.90
Total	9,630	14,199		

The final level of analysis is at the product level. I supplement the firm-level data with product-level data at the 9 digit level where I observe the sales and physical output of each product produced by the firm. I can therefore calculate product price and using structural techniques of estimating production functions, I estimate physical productivity. This product data is however only available from 1994 onward.

2.5 Empirical Strategy

The expansion of the grid is demand driven. In fact, PLN follows a demand forecast methodology where they forecast demand in a certain area and compare it to the existing supply infrastructure. PLN then decides to expand it if they believe there will be a gap between supply and demand in the future. I explain this methodology in detail in Appendix 2.B. Importantly, this methodology implies that the bias in ordinary least square estimates can go either way. On the one hand, more productive regions have higher demand forecasts, which means that OLS will be upward bias. On the other hand, areas with generally poor infrastructure, where firms are less productive, will have a higher gap between demand forecasts and existing supply, meaning that OLS will be downward bias. Another element in the decision of expanding the grid is the cost of construction, which is potentially exogenous.

Using the data described above, I estimate the effect of access to the grid $Access_{vpt}$ on outcome Y_{vpt} of desa v, province p and year t using the following specification:

$$Y_{vpt} = \alpha + \beta Access_{vpt} + \eta \mathbf{V}_{vpt} + \gamma_p + \delta_t + \epsilon_{vpt} \tag{2.1}$$

and the firm-level equivalent where I estimate the effect access $Access_{vpst}$ on outcome y_{ivpst} of firm i in desa v, province p, industry s and year t.

$$y_{ivpst} = \alpha + \beta Access_{vpst} + \nu \mathbf{X}_{ivpst} + \eta \mathbf{V}_{vpt} + \gamma_p + \delta_{st} + \epsilon_{ivpst} \tag{2.2}$$

where $\mathbf{X}_{ivpst}$ is a vector of firm controls, $\mathbf{V}_{vpst}$ is a vector if desa level controls, γ_p are province fixed effects, δ_t are year fixed effects and δ_{st} are industry-by-year fixed effects.

Electricity grids are placed endogenously to industrial outcomes. Even conditional on all the listed controls, estimating the above model by OLS will give biased results. In order to deal with the endogeneity problem, I propose an instrumental variable approach exploiting a supply-side natural experiment. Up until the late 1980's, the electricity grid in Java was not interconnected. My empirical strategy exploits the fact that PLN needed to build an interconnection of the grid, which occurred by the start of my sample period. This interconnection created a change in the probability of receiving electricity in the future in certain desas that lie between two grids. The section below describes how this strategy in detail.

2.5.1 Hypothetical Least Cost Grid

In 1969, electricity grid in Java consisted of 5 different disconnected grids across the island (Figure 2.3). Having disconnected grids is inefficient, prevents load-sharing across regions, and increases the price of supplying electricity. Therefore, the 1970's and the 1980's witnessed a huge and successful effort by PLN with the help of agencies such as the World Bank and the Asian Development Bank to connect the various grids on the island (Figure 2.4). Various transmission lines were built for the main purpose of interconnecting the grid. As a result, desas nearby the lines connecting the grids faced a positive shock to the probability of receiving electricity access in the future as it is cheaper to connect desas that are closer to the existing network. Distance to these transmission lines is therefore a potential instrument for access, however, the lines themselves could be endogenously placed.

Figure 2.3: Java Network 1969

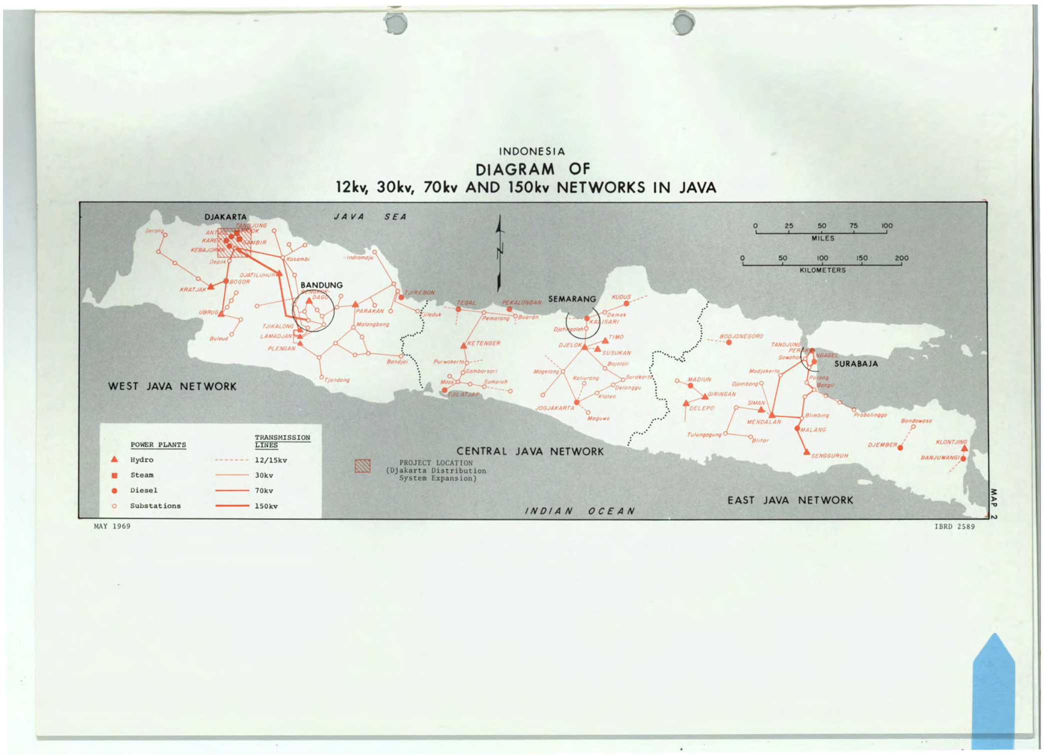


Figure 2.4: Java Network 1989

To deal with the concern that transmission lines could be targeted at areas that are different than others, for example, non-farming land, I create a hypothetical grid to connect the main power plants in the separate grids. In total, I consider 15 power plants which I identify from historical maps as the main power plants in the 5 separate grids. I implement the following procedure to construct the hypothetical least cost grid:

1. For each pixel on the map, I assign a cost value based on elevation and waterway data. Cost is a simple linear function of these two variables.

2. I calculate the least cost path for each pair of power plants based on the cost data.

3. I use Kruskal's algorithm[6] to find the least cost combination of least cost paths such that all power plants are interconnected. The resulting network is the hypothetical least cost transmission grid.

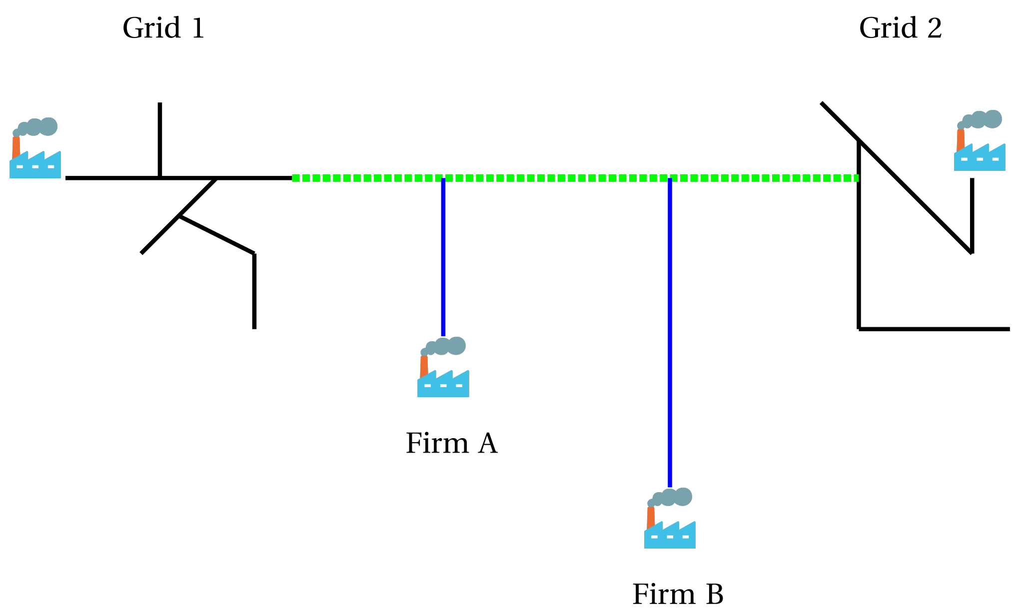
Figure 2.5 shows the resulting hypothetical least cost grid. The distance to the hypothetical least cost grid is then used as the instrumental variable.

[6] Kruskal's algorithm is a minimum spanning tree algorithm. The minimum spanning tree is the spanning tree that has the lowest cost among all the possible spanning trees. The cost of the spanning tree is defined as the sum of the weights of all the edges in the tree.

Figure 2.5: Hypothetical Least Cost Transmission Grid

Figure 2.6 illustrates the empirical strategy in a simplified manner. Consider two disconnected grids Grid 1 and Grid 2. These represent the incumbent infrastructure built by the Dutch electricity company and were existent by 1969. During the 1970s and the 1980s, the two grids became interconnected by the green line. Consider two firms (or desas) A and B that only differ in their distance to the green line. Because Firm A is closer to the green line, it is then more likely to get connected to the electricity grid in the 1990s compared to Firm B. The blue lines therefore represent the instrument. Because of potential concerns regarding the placement of the green line, I create a hypothetical green line that is based solely on cost factors. The hypothetical least cost grid is essentially an instrument for the actual interconnection transmission network.

Figure 2.6: Empirical Strategy



In my empirical strategy, I control for various desa-level characteristics. One concern is that the location of the power plants is endogenous. In Java, many of these power plants are hydroelectric power plants, meaning their location is tied to the natural source. In addition, these power plants have been built by the Dutch electric utilities decades before the start of the sample period[7]. It is likely then that the factors determining the location of these power plants do not directly affect outcomes in 1990 (conditional on controls). Nonetheless, I exclude desas within a certain radius of power plants to deal with the concern that power plants are endogenously located. Power plants are built close to the consumption centers that they are meant to supply electricity to in order to minimize transmission losses. Because consumption centers are typically cities and urban areas, one concern is that the instrument is correlated to distance to the closest city. To alleviate this concern, I include distance to the nearest city as a control variable.

Because most economic activity is located along the coast of the island, many of the power plants are located there as well. One reason is that the coast is flatter and therefore it is cheaper to build there. Furthermore, proximity to coal sources for thermal power plants is crucial. Coal in Indonesia is mostly available in the islands

[7] http://maps.library.leiden.edu/apps/search?code=04693focus

of Sumatera and Kalimantan, which are easily reachable from the north coast because of proximity and good wave conditions in the Java sea. Furthermore, because the coast is flatter, Kruskal's algorithm will favor lines along the coast. It is then important to control for distance to coast in any empirical specification to avoid any threats to exclusion.

Controlling for desa elevation is also necessary because it is correlated with distance to hypothetical least coast grid. Another potential confounder is the possible correlation between distance to the hypothetical grid and the road network in Java. For that reason, controlling for distance to road is important to guarantee the exclusion of the instrument. In all my specification, I control for the distance to the nearest regional road. I also control for the availability of non-energy infrastructure facilities. These include railway station, motor station, river pier, sea port, and airport. In addition to geographic controls, I also control for the desa political status and legal status. Political status is an indicator for whether the desa is the district capital. Legal status of the village refers to whether the desa is governed by an elected official, appointed official, or a traditional chief.

At the firm level, I control for whether the firm is public or private to deal with any favoritism in access towards government owned firms. I also control for firm age, legal status, and export status. The identification assumption is that, conditional on controls, the potential outcomes of desas or firms are independent of their distance to the hypothetical least cost grid.

To summarize, geographic desa controls include distance to coast, elevation, distance to nearest city, and distance to nearest road. Other desa level controls include various infrastructure availability dummies, political status, and legal status. Firm level controls include firm age, export status, legal status and ownership type.

2.5.2 Instrument Variation and Controls

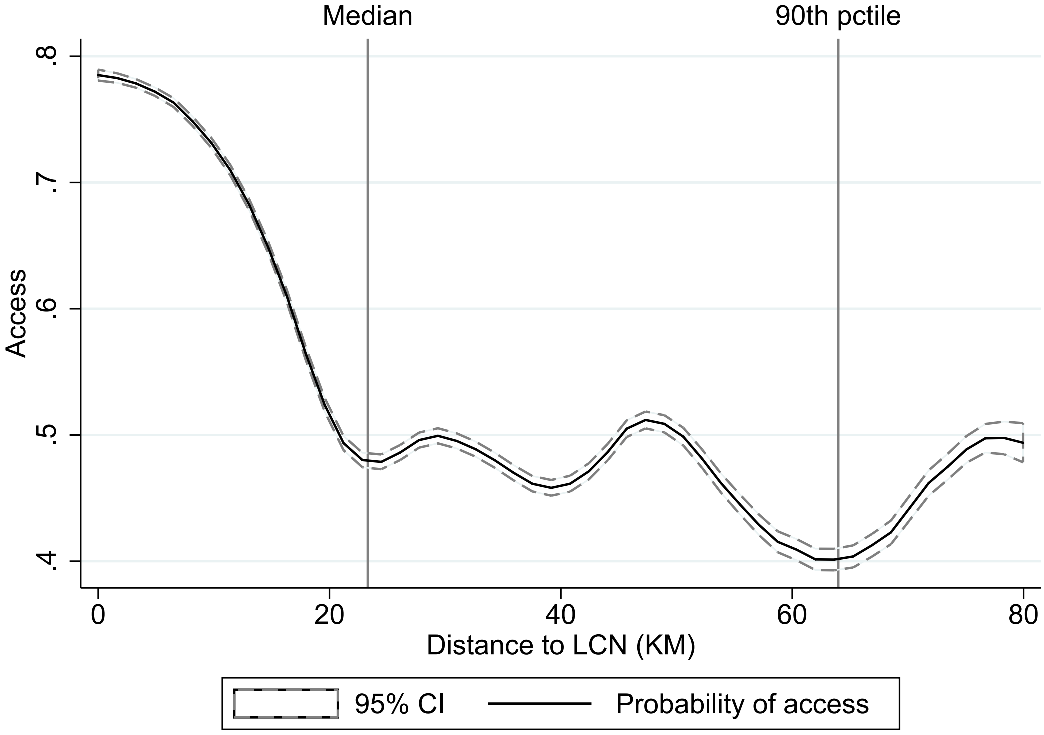
Given that the instrument used to identify the causal effect of electrification is based on geography, what variation is left in the distance to the hypothetical grid after controlling for all geographic characteristics of desas? In other words, conditional on local geography, why is it possible to still have two desas with different distances to the hypothetical grid? The answer is because what matters for the hypothetical least cost grid is *global* geography, not local geography. This is because the hypothetical least cost grid has the objective of minimizing the cost of building the transmission grid, taking the location of the incumbent power plants as given. This is different to using local geography to create the cheapest possible grid and predict access as in Lipscomb, Mobarak, and Barham (2013) where the authors create a least cost grid, including simulated locations of power plants, given the national budget. When taking as given the location of actual power plants, the least cost algorithm will not always choose the flatter areas because in some locations choosing a steeper path might lead to a flatter path further ahead on route to the next power plant. This creates variation in the distance to the hypothetical grid for locations with the same local geographic characteristics.

2.5.3 Desa-Level First Stage

Figure 2.7 plots the unconditional probability of a desa having access to the grid as a function of the distance to the hypothetical least cost grid. The closer a desa is to the hypothetical grid, the more likely it is to have access to the actual grid. The relationship between the probability of access to the actual grid and the instrument is negative. I also plot the median and 90^{th} percentile of the instrument. At large values of the instrument, i.e. for desas very far from the hypothetical, the instrument doesn't predict the probability of access very well. However, this is not much of a concern as there are few observations in that region (beyond the 90^{th} percentile). Figure 2.8 plots the probability of a desa having access to the grid for the years 1990, 1995 and 2000, against the distance to the hypothetical grid. The graph shows that the negative relationship between access and the instrument persists over time. Holding distance to the hypothetical grid fixed, the probability of having access to the grid is increasing over time. This captures the fact that the electricity grid was expanded substantially between 1990 and 2000, increasing access from around 43% of Java's desas to 71%[8].

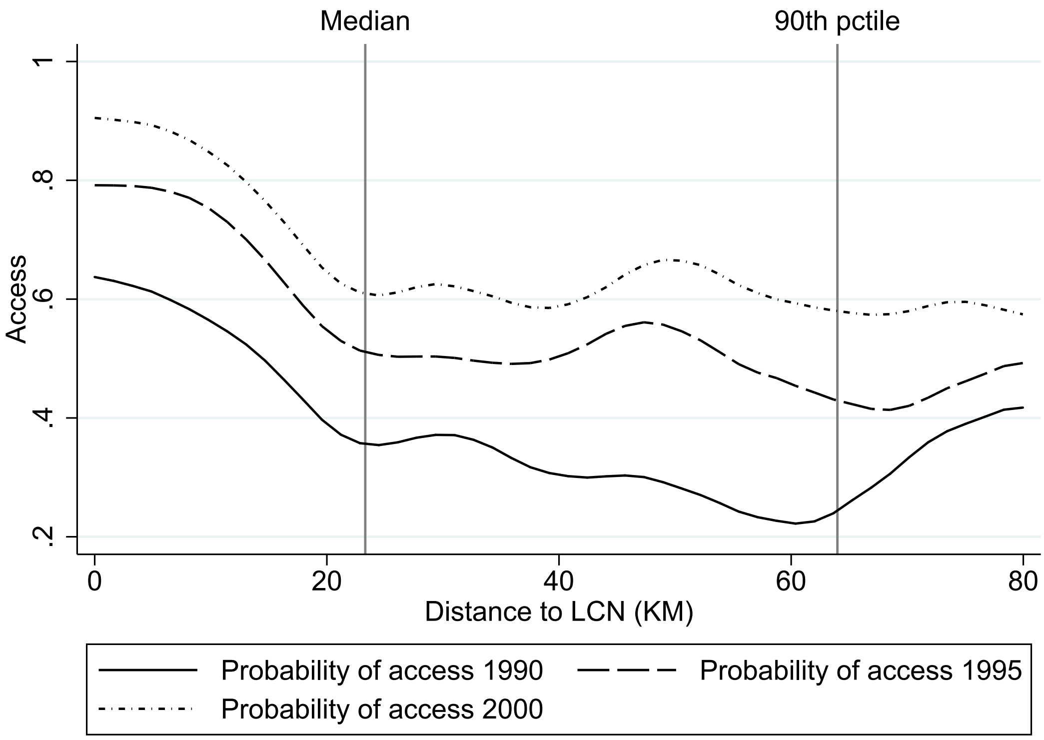
[8] PLN reports an electrification ratio of 50% in 1990.

Figure 2.7: Distance to Hypothetical Grid and Probability of Being Connected



The y-axis presents probability of a desa being connected to the grid, where $Access_{vpt}$ is a dummy variable equal to 1 if a desa is within 15 KM of the nearest transmission substation. The probability is estimated using an Epanechnikov kernel function with a bandwidth of 2.16. The x-axis shows the distance from the desa to the hypothetical least cost grid. The median and the 90^{th} percentile of the distance to the hypothetical grid are shown for reference.

Figure 2.8: Distance to hypothetical gridand Probability of Being Connected, by Year



The y-axis presents probability of a desa being connected to the grid for years 1990, 1995 and 2000, where $Access_{vpt}$ is a dummy variable equal to 1 if a desa is within 15 KM of the nearest transmission substation. The probability is estimated using an Epanechnikov kernel function with a bandwidth of 2.16. The x-axis shows the distance from the desa to the hypothetical least cost grid. The median and the 90^{th} percentile of the distance to the hypothetical grid are shown for reference.

Table 2.4 shows the first stage regression using distance to the hypothetical least cost grid $Z(KM)$ as an instrumental variable and using all the controls discussed above. The dependent variable, $Access_{vpt}$, is an indicator variable equal to one if the desa is within 15 KM[9] of the nearest transmission substation in year t.

The coefficient in column (1) is negative and significant, indicating that the further away a desa is from the hypothetical least cost network, the less likely it is to have access to electricity. The first stage F-statistic is high enough to guarantee relevance of the instrument, avoiding weak instrument bias. The coefficient in column (1) then shows that even conditional on various controls, this difference in means is still significant and distance to the hypothetical grid is a good predictor of access to electricity at the desa level.

[9]This threshold was chosen based on conversations with electrical engineers at the Indonesian state electricity monopoly. The results are not sensitive to this particular choice.

Table 2.4: First Stage Regressions

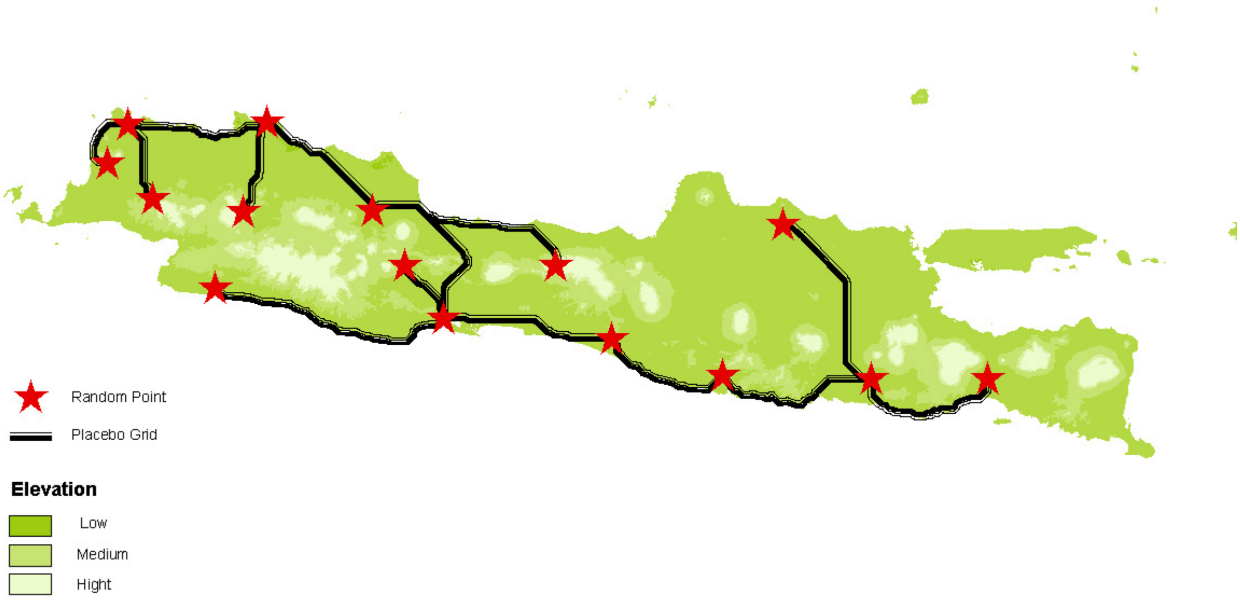
	(1)
	Access_{vpt}
Z (KM)	-0.00165***
	(0.000152)
Distance to city	-0.00263***
	(0.000131)
Distance to coast	5.56e-05
	(0.000149)
Elevation	-0.191***
	(0.00940)
Distance to road	-0.00410***
	(0.000664)
Motorstation	-0.0281**
	(0.0136)
Railway	0.0419**
	(0.0191)
Seaport	-0.0545
	(0.0537)
Airport	0.167***
	(0.0423)
First Stage F	118.7
Observations	261,470
Year FE	✓
Desa Controls	✓
Province FE	✓

Notes: First stage regression of access instrumented with distance to hypothetical least cost grid. Access is defined at the desa level. A desa has $Access_{vpt} = 1$ if it is within 15 Km of the nearest substation. Desa controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political status, and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

2.5.4 Instrument Validity

In this section, I present two exercises that test the validity of the hypothetical least cost grid instrument. First, I create a placebo hypothetical least cost grid that connects some random points in Java using the same least cost algorithm as the one used in the main instrument (figure 2.5). If access to the grid is correlated with the distance to this least cost placebo grid, it would mean that local geography, irrespective of the location of the actual electric transmission grid, is what is driving the correlation between access and the instrument. Figure 2.9 illustrates the placebo hypothetical least cost grid. The origin points to be connected by the algorithm were randomly chosen by the computer. The same algorithm applied to create the hypothetical least cost network using the main incumbent power plants was applied to connect these randomly generated points on a single network.

Figure 2.9: Placebo Least Cost Grid



The second test is based on a Euclidean or straight line version of the least cost grid where instead of connecting the colonial power plants with least cost paths based on geography, I connect them on a network of straight lines, ignoring geography. This version of the hypothetical grid should alleviate any concerns that local geography is what drives the correlation between the instrument and access to the grid as opposed to the incumbent electric infrastructure. Figure 2.10 illustrates the hypothetical Euclidean grid. The power plants connected by the straight lines are the same as in the original hypothetical least cost grid. Each of the power plants was connected to the closest power plant by a straight line, resulting in a single interconnected grid of straight lines.

Figure 2.10: Hypothetical Euclidean Grid

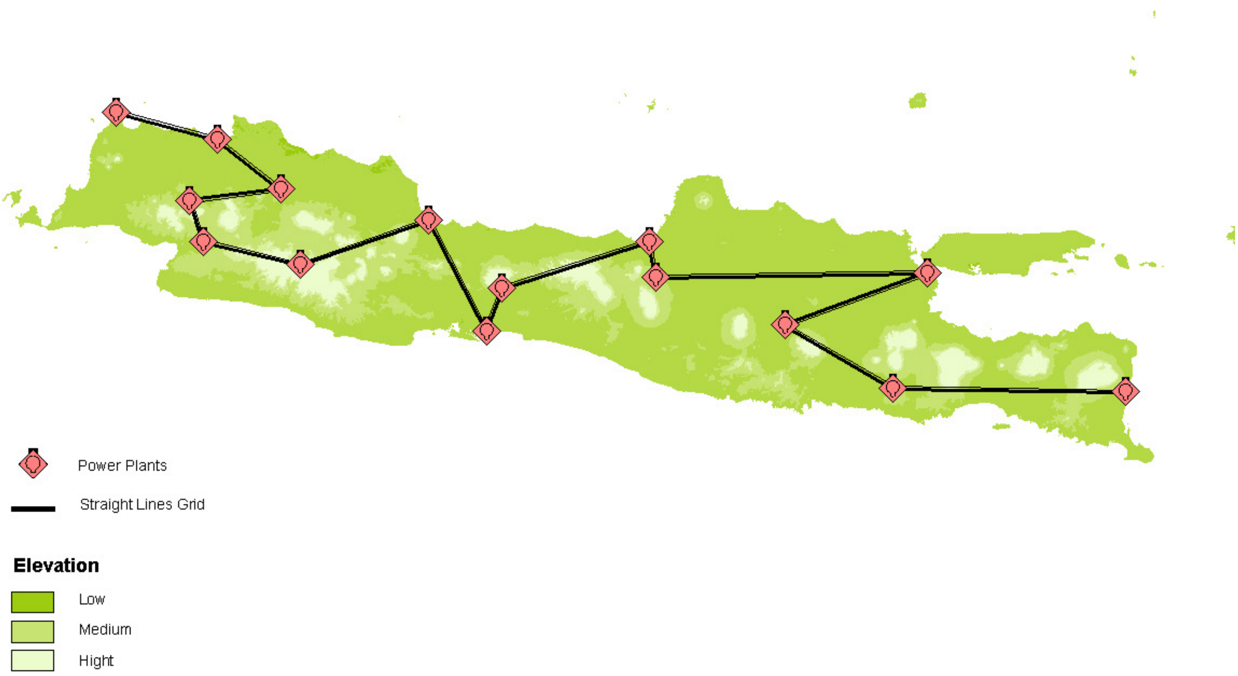


Table 2.5 presents the results of the first stage regressions using these two alternative instruments. The first row shows the coefficient on the instrument, where in each column a different instrument is used. For comparability, column (1) presents again the first stage using the main instrument Z, the distance to the hypothetical least cost grid.

Column (2) presents the results from the first stage regression of access on the placebo instrument. There is no correlation between access to the grid and the distance to the placebo grid and the estimated coefficient is very small and statistically indistinguishable from zero. The first stage F is close to zero. The coefficients on the control variables remain more or less unchanged. The fact that access and distance to the placebo grid are not correlated alleviates the concern that correlation between access and the main instrument is purely driven by geography. The origin points of the hypothetical least cost grid, or the incumbent infrastructure, plays an important role in determining the correlation between access and Z.

Finally, column (3) presents the first stage of access on the distance to the hypothetical Euclidean grid. This grid only takes into account the origin points and abstracts from geography. The coefficient on the instrument in column (3) shows that there is a significant correlation between access and distant to the Euclidean grid. This is reassuring because it suggests that the location of the main power plants is the main driver of the strong first stage regression in the main empirical specification.

Table 2.5: First Stage Regressions-Validity

Dependent Variable		Access_{vpt}	
Instrument	Z	Placebo	Euclidean
	(1)	(2)	(3)
Instrument	-0.00167***	7.08e-05	-0.00153***
	(0.000152)	(0.000115)	(0.000149)
Distance to city	-0.00262***	-0.00330***	-0.00289***
	(0.000131)	(0.000118)	(0.000124)
Distance to coast	5.95e-05	-6.33e-05	-0.000252*
	(0.000149)	(0.000147)	(0.000145)
Elevation	-0.191***	-0.210***	-0.214***
	(0.00941)	(0.00930)	(0.00934)
Distance to road	-0.00410***	-0.00492***	-0.00381***
	(0.000662)	(0.000633)	(0.000668)
Motorstation	-0.0316**	-0.0334**	-0.0307**
	(0.0136)	(0.0138)	(0.0137)
Railway	0.0468**	0.0544***	0.0462**
	(0.0190)	(0.0188)	(0.0189)
Seaport	-0.0575	-0.0565	-0.0488
	(0.0536)	(0.0542)	(0.0555)
Airport	0.168***	0.161***	0.167***
	(0.0423)	(0.0429)	(0.0423)
Riverpier	0.0256	0.0381	0.0351
	(0.0446)	(0.0452)	(0.0454)
First Stage F	118.7	0.380	106.3
Observations	261,470	261,470	261,470
Year FE	✓	✓	✓
Desa Controls	✓	✓	✓
Province FE	✓	✓	✓

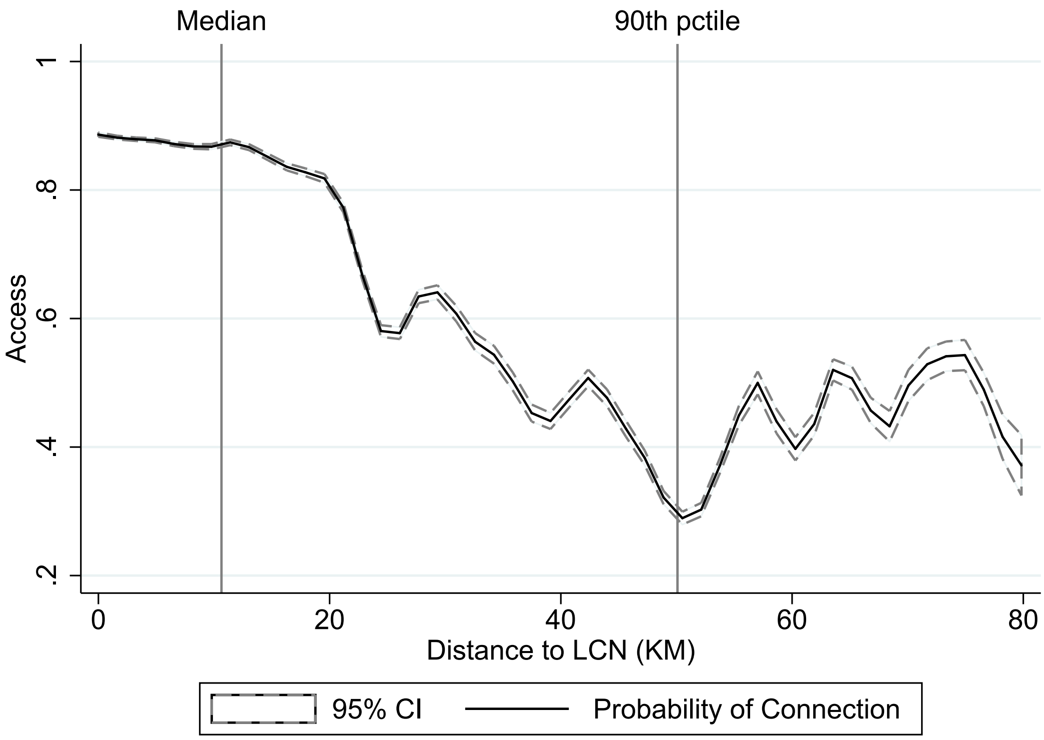
Notes: First stage regressions of access instrumented with distance to hypothetical least cost grid in column (1), placebo least cost grid in column (2), and hypothetical Euclidean grid in column (3). Access is defined at the desa level. A desa has $Access_{vpt} = 1$ if it is within 15 Km of the nearest substation. Desa controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political status, and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

2.5.5 Firm-Level First Stage

Because part of the analysis in the subsequent chapters is at the firm level, and given that firms are located in a subset of the desas, it is necessary to check whether my empirical strategy is still valid at that level. I now check if distance to the hypothetical least cost grid still explains access to electricity at the firm-level. In the current section, I use the same definition of access, $Access_{vpt}$. This is an indicator is equal to one if an firm is located in a desa within $15km$ of the nearest transmission substation. Based on the results from the previous section, firms are located in desas that are on average closer to the hypothetical least cost grid. One concern is therefore whether the instrument is still strong enough.

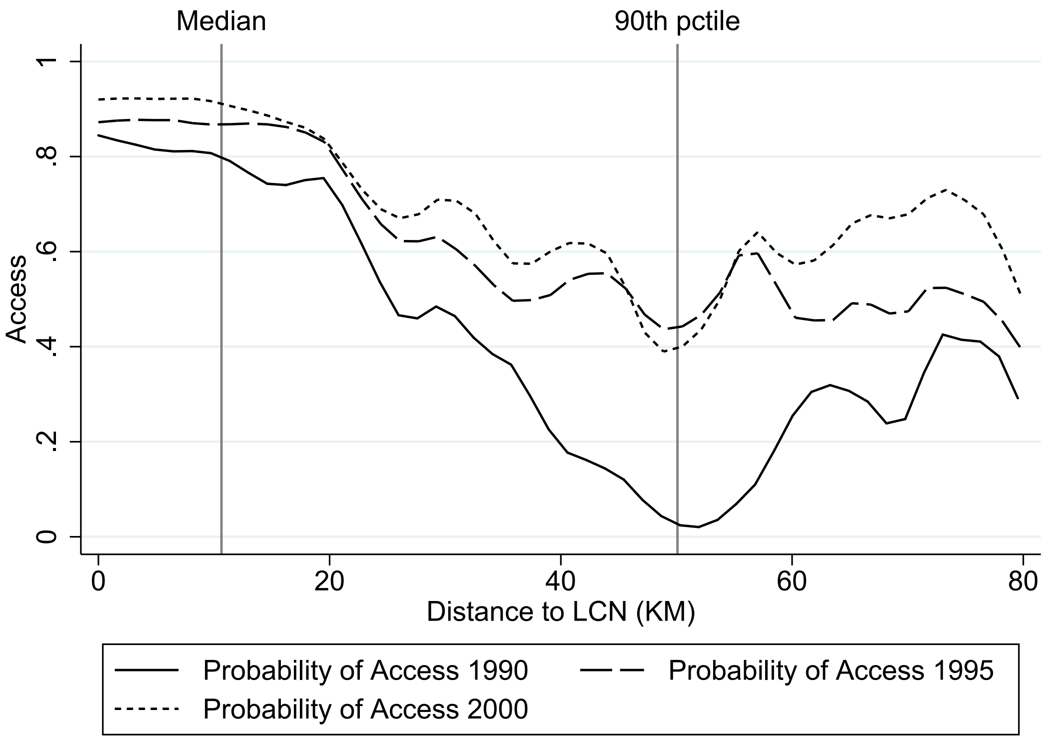
Figures 2.11 and 2.12 show again a negative relationship between the unconditional probability of having access and distance to the least cost network, which is consistent over time.

Figure 2.11: Distance to Hypothetical Grid and Firm Access



The y-axis presents probability of a firm being in a desa with access to the grid, where $Access_{vpt}$ is a dummy variable equal to 1 if a desa is within 15 KM of the nearest transmission substation. The probability is estimated using an Epanechnikov kernel function with a bandwidth of 2.49. The x-axis shows the distance from the desa to the hypothetical least cost grid. The median and the 90^{th} percentile of the distance to the hypothetical grid are shown for reference.

Figure 2.12: Distance to Hypothetical Grid and Firm-Level Access, by Year



The y-axis presents probability of a firm being in a desa with access to the grid for years 1990, 1995 and 2000, where $Access_{vpt}$ is a dummy variable equal to 1 if a desa is within 15 KM of the nearest transmission substation. The probability is estimated using an Epanechnikov kernel function with a bandwidth of 2.49. The x-axis shows the distance from the desa to the hypothetical least cost grid. The median and the 90^{th} percentile of the distance to the hypothetical grid are shown for reference.

Column (1) of table 2.6 show the first stage regressions of access on z_{vt}, the distance to the hypothetical least cost grid. In addition to the above controls defined at the desa-level, I include firm-level controls and year-by-industry fixed effects. The coefficient in column (1) is negative and significant and the first stage F-statistic is high. The instrument is therefore still relevant.

Table 2.6: First Stage Regression - Firm Level

	(1) Access_{vpt}
Z (KM)	-0.00296***
	(0.000460)
Distance to city	-0.00320***
	(0.000304)
Distance to coast	0.00163***
	(0.000455)
Elevation	-0.0858**
	(0.0401)
Distance to road	-0.000329
	(0.000524)
Motorstation	-0.00699
	(0.0142)
Railway	0.00927
	(0.0220)
Seaport	-0.174***
	(0.0646)
Airport	0.0203
	(0.0174)
First Stage F	41.55
Observations	141,615
YearxIndustry FE	✓
Desa Controls	✓
Province FE	✓
Firm Controls	✓

Notes: First stage regression of access instrumented with distance to hypothetical least cost grid. Access is defined at the desa level. A desa has $Access_{vpt} = 1$ if it is within 15 Km of the nearest substation. Desa controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political status, and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

2.6 Conclusion

In this chapter, I presented the data on the expansion of the electric transmission grid that I have collected as part of this project. This unique data on the electrification infrastructure in Indonesia, combined with rich manufacturing census data, allow me to evaluate the effect of electrification on industrial development.

In order to estimate the causal effect of electrification on industrial outcomes, I have designed an empirical strategy that exploits a supply-side natural experiment based on the need of a single interconnected transmission grid on the island of Java. I build a hypothetical least cost transmission grid as a function of incumbent colonial electric infrastructure and geographic cost factors. This hypothetical grid is the basis of the empirical strategy applied in the following chapters. I start by looking at the effect of electrification on local industrial activity in the next chapter.

Appendix

2.A Figures

Figure 2.13: Example of Inventory Table of Transmission Transformers.

Sektor : *Madiun* Hal : 3 / 3

ROUTE DARI - KE		No.	Teg [kV]	Jenis Konduktor	Kapasitas [Amp]	Kapasitas [MVA]	Panjang [km]	Tower [buah]	Operasi [Tahun]	Keterangan
Banaran	Mojoagung	1	150	ACSR.330	740	192	27,60	83	01/01/83	
Banaran	Mojoagung	2	150	ACSR.330	740	192	27,60		01/01/83	
Banaran	SuryaZigZag	1	150	ACSR.330	740	192	12,20	36	01/01/73	
Bojonegoro	Babat.	1	150	Hawk	600	156	35,30	106	01/01/83	
Bojonegoro	Babat.	2	150	Hawk	600	156	35,30		01/01/83	
Bojonegoro	Cepu	1	150	Hawk	600	156	30,97	97	01/01/83	
Bojonegoro	Cepu	2	150	Hawk	600	156	30,97		01/01/83	
Kerek	Mliwang	1	150	Hawk	600	156	9,00	28	01/01/94	
Kerek	Mliwang	2	150	Hawk	600	156	9,00		01/01/94	
Kerek	SemenTuban 3	1	150	Hawk	600	156	2,02	10	08/10/97	
Kerek	SemenTuban 3	2	150	Hawk	600	156	2,02		08/10/97	
Lamongan	Babat.	1	150	TACSR.240	900	234	12,91	91	01/06/96	Reconductoring Hawk -> TACSR.240 th
Lamongan	Babat.	2	150	TACSR.240	900	234	12,91		01/06/96	Reconductoring Hawk -> TACSR.240 th
Manisrejo	Ngawi	2	150	Hawk	600	156	40,70	16	16/04/94	16 tower u/ Branch Ngawi
Manisrejo	Sragen	1	150	Hawk	600	156	78,67	168	01/01/93	
Manisrejo	SuryaZigZag	2	150	ACSR.330	740	192	61,43		01/01/73	
Sragen	Ngawi	2	150	Hawk	600	156	48,97		01/01/1923	
Tuban	Kerek	1	150	Hawk	600	156	14,06	42	01/01/94	
Tuban	Kerek	2	150	Hawk	600	156	14,06		01/01/95	

Panjang transmisi	1.600,95 kms
Jumlah Tower/Tiang	2786 unit

25 kV = 50.73 kms
70 kV = 846.16 kms
150 kV = 703.62 kms

10/05/01 16:34:57 Asset/OMTrans

Inventory table of operating transmission transformers in the Java-Bali transmission network, April 2001. This table corresponds to the Madiun sub-grid and includes information on the voltage, brand, capacity, origin and destination of the connection, and operation year. *Source: PLN.*

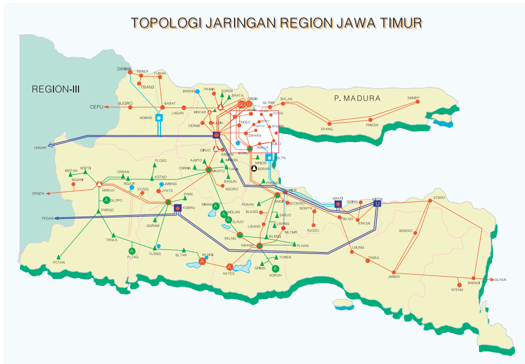
Figure 2.14: Example of current maps of the transmission network in Java.

(a) Java and Banten

(b) West Java

(c) Central Java and Yogyakarta

(d) East Java



Source: Electricity Supply Business Plan (RUPTL) 2006-2015, PLN

2.B Demand Forecasts

2.B.1 Methodology Overview: DKL

The model combines multiple methods; mainly trend projections and estimating elasticities using OLS (referred to as the econometric model by PLN). PLN conducts its forecast at the sectoral level before aggregating at the regional level. In the case of Java, the forecast is aggregated at the system level. PLN considers four sectors: Residential, Commercial, Public and Industrial. For each of these sectors, energy consumption is forecasted as a function of historical PLN data, macroeconomic variables, and elasticities of energy sales in that sector with respect to economic growth.

2.B.2 Residential Sector

- Energy Consumed: $E_t^R = E_{t-1}^R * (1 + \epsilon_t^R * g_t) + \Delta N b_t^R * UK_t^R$ where:

  - ϵ_t^R is the elasticity of residential energy sales (kWh) with respect to regional GDP growth. Elasticities are obtained using the econometric model where they calculate the elasticity either by using actual yearly data or by regressing log sales on log gdp.
  - g_t is the regional GDP growth rate. This is either taken from BPS the Indonesian Statistics Bureau or projected linearly.
  - ΔNb_t^R is the change in the number of residential customers between year t and year $t-1$. For future years, it is the change in the forecasted number of customers between two years. The number of customers is projected linearly using customer factor (the equivalent of elasticity) and population growth rates where CF_t^R is calculated as the elasticity of the number of customers with respect to economic growth[10].
  - $Nb_t^R = Nb_{t-1}^R * (1 + CF_t^R * g_t)$
  - UK_t^R is energy consumption per customer (kWh/hh). The customer is one household.

- In order to forecast electrification ratios, the future number of households in the economy is forecasted using population forecasts and average number of individuals per household and then used with the forecasted number of customers to calculate the implied electrification ratio.

2.B.3 Commercial, Industrial and Public Sectors

Similarly, for each sector i, the goal is to get an estimate of energy consumption. This is done as follows: the number of customers is calculated/projected:

- Energy Consumed: $E_t^i = E_{t-1}^i * (1 + \epsilon_t^i * g_t)$ where:
  - ϵ_t^i is the elasticity of energy sales (kWh) in sector i with respect to regional GDP growth.
  - g_t is the regional GDP growth rate.

- In order to forecast power contracted, average power (VA) per customer is multiplied by the number of new customers in sector, then it is added to the previous year's power contracted:

- $PC_t^i = PC_{t-1}^i + \Delta Nb_t^i * UK$

[10] PLN assumes elasticities if the calculated ones are unreasonable.

- ΔNb_t^i the change in the number of customers between year t and year $t-1$ in sector i
- $Nb_t^R = Nb_{t-1}^R * (1 + CF_t^R * g_t)$
- UK_t^i is energy consumption per customer (kWh/hh) which is the average from historical data.

2.B.4 Forecasted Total Demand and Load Factor

- Total Energy Sales (GWh): $ES_t = E_t^R + E_t^C + E_t^P + E_t^I$
- Forecasted energy sales represent the energy needs of PLN customers
- Required energy production (GWh) needs to take account of inefficiencies such as transmission and distribution losses (L%) and station use(SU%): $P_t = \frac{ES_t}{(1-L-SU)}$
- The final form of demand forecast is called peak load (MW). To calculate that from required production, the load factor is needed: $LF_t = 0.605 * \frac{E_t^R}{ES_t} + 0.7 * \frac{E_t^C + E_t^P}{ES_t} + 0.9 * \frac{E_t^I}{ES_t} < 1$
- Finally, the peak load of the system, which is the goal of this procedure, is:

$$PL_t = \frac{P_t}{365 * 24 * LF_t * 1000}$$

2.B.5 Disaggregation

Because the Java-Bali system is interconnected, forecast is done at the system level. Figure 2.15 shows an example of a demand forecast table for the next 10 years done in 1992 by PLN. The next step is to disaggregate this forecast at the substation level. The way this is done is by looking at the proportion of the load borne by each substation out of the whole system load, and assuming that in the future these proportions will be the same. Then divide the forecasted load according to each substation's proportion. Once the load forecast is calculated for each substation, it is then compared to the capacity of each substation. If the load is greater than 80% of the capacity, then the substation should be extended or a new substation is commissioned.

Figure 2.15: Example of a Demand Forecast Table.

Energy & Load Demand Forecast
JAVA & BALI SYSTEM (HIGH)

Table 3.9

Fiscal Year	1990*	1991#	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Residential														
Population(10^3)	110351.6	112166.3	113998.8	115855.1	117727.9	119592.1	121422.4	123251.1	125097.0	126958.9	128825.2	130670.5	132470.1	134262.8
Growth Rate(%)		1.6	1.6	1.6	1.6	1.6	1.5	1.5	1.5	1.5	1.5	1.4	1.4	1.4
Electr.Ratio(%)	34.0	36.1	41.4	45.5	49.9	54.8	59.9	65.3	71.1	76.8	82.4	87.9	93.4	98.7
No.of Customers(10^3)	7500.0	8093.0	9437.4	10549.2	11760.7	13095.7	14539.1	16100.5	17789.0	19502.7	21230.9	22974.2	24734.2	26512.9
Power Contr./Cust.(VA)	628.3	626.2	625.8	620.1	618.0	616.2	614.6	613.2	611.9	610.9	610.0	609.2	608.6	608.0
Power Contracted(MVA)	4712.0	5099.2	5874.4	6541.5	7268.4	8069.4	8935.4	9872.3	10885.4	11913.6	12950.5	13996.5	15052.5	16119.8
Growth of GDP Total(%)	8.9	7.6	7.9	8.4	7.8	8.2	8.1	8.2	8.1	8.0	7.3	7.3	7.3	7.3
Consump./Cust.(kwh)	906.0	950.9	859.3	886.9	891.2	901.5	913.7	928.7	945.1	957.7	966.4	980.4	999.4	1022.7
PLN SALES (GWh)	6795.3	7695.3	8109.6	9356.6	10481.7	11805.4	13284.9	14953.1	16812.3	18677.4	20517.8	22524.2	24719.2	27113.6
Growth (%)	13.3	13.2	5.4	15.4	12.0	12.6	12.5	12.6	12.4	11.1	9.9	9.8	9.7	9.7
Share to Tot.(%)	30.3	30.4	29.5	27.9	25.6	24.1	22.9	22.2	21.7	21.1	20.5	20.0	19.8	19.5
Commercial														
No.of Customers	244481	264257	308366	335450	364169	394988	427444	461667	497761	533511	568755	603573	638043	672250
Customers Elasticity	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Elasticity	1.8	1.8	1.7	1.7	1.7	1.6	1.6	1.6	1.5	1.5	1.5	1.4	1.4	1.4
Growth of GDP Sec.(%)	9.46	7.01	7.71	7.89	8.02	7.68	7.30	7.14	8.26	8.19	8.10	6.51	6.56	6.58
Power Contr./Cust.(VA)	5688.0	5688.9	6037.1	5692.4	5693.0	5693.6	5694.1	5694.5	5694.9	5695.2	5695.5	5695.8	5696.0	5696.2
Power Contracted(MVA)	1390.6	1610.2	1755.1	1909.5	2073.2	2248.9	2433.9	2629.0	2834.7	3038.5	3239.4	3437.8	3634.3	3829.3
Consump./Cust.(kwh)	7659.7	8518.2	7966.2	8626.8	9370.1	10066.1	10770.4	11510.3	12415.7	13422.8	14536.8	15365.9	16285.8	17295.4
PLN SALES (GWh)	1872.7	2251.0	2456.5	2893.9	3412.3	3976.0	4603.7	5314.0	6180.1	7161.2	8267.9	9274.5	10391.1	11626.8
Growth (%)	17.4	20.2	9.1	17.8	17.9	16.5	15.8	15.4	16.3	15.9	15.5	12.2	12.0	11.9
Share to Tot.(%)	8.36	8.90	8.93	8.64	8.34	8.11	7.93	7.90	7.96	8.08	8.25	8.25	8.31	8.38
Public & Others														
No.of Customers	178556	195096	223148	248626	276296	306692	339450	374781	412871	451417	490183	529191	568479	608098
Customers Elasticity	1.15	1.15	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Elasticity	1.5	1.5	1.5	1.5	1.5	1.5	1.4	1.4	1.4	1.3	1.3	1.3	1.3	1.3
Power Contr./Cust.(VA)	7829.5	7786.9	7799.5	7687.9	7651.7	7619.5	7591.3	7566.4	7544.3	7525.8	7510.0	7496.5	7484.8	7474.5
Power Contracted(MVA)	1398.0	1534.9	1724.7	1911.4	2114.1	2336.8	2576.9	2835.7	3114.8	3397.3	3681.3	3967.1	4255.0	4545.3
Consump./Cust.(kwh)	9485.9	8880.2	9592.3	10085.8	10636.1	11176.1	11668.8	12175.3	12891.7	13596.7	14377.5	14974.7	15646.0	16387.5
PLN SALES (GWh)	1693.8	1732.5	2140.5	2507.6	2938.7	3427.6	3961.0	4563.1	5322.6	6137.8	7047.6	7924.5	8894.4	9965.2
Growth (%)	7.00	2.29	23.55	17.15	17.19	16.64	15.56	15.20	16.65	15.32	14.82	12.44	12.24	12.04
Share to Tot.(%)	7.56	6.85	7.78	7.49	7.18	6.99	6.83	6.79	6.85	6.92	7.04	7.05	7.12	7.18
Industry														
No.of Customers	27204	27547	36228	42584	47507	53569	60988	69337	77221	85005	93428	102551	112856	124027
Elasticity	1.2	1.1	1.1	1.1	1.0	1.0	1.0	1.0	1.0	0.9	0.9	0.9	0.9	0.9
Growth of GDP Sec.(%)	14.45	14.82	13.76	15.95	11.56	12.76	13.85	13.69	11.37	11.20	11.01	10.85	10.69	10.53
Power Contr./Cust.(VA)	187215.1	198881.9	224369.0	232511.6	260958.2	248515.3	236652.4	226338.6	218647.1	212452.7	206912.3	201938.2	197287.0	193118.1
Power Contracted(MVA)	5093.0	5478.6	8128.4	9901.3	12397.3	13312.7	14433.0	15693.7	16884.2	18059.5	19331.4	20709.0	22265.1	23951.9
Energy Ind.Demand(GWh	20366.2	23686.3	26997.2	31733.9	35402.3	39919.6	45448.5	51670.4	57545.3	63345.9	69622.9	76421.5	84100.8	92425.3
Energy Big Cust.(GWh)	0.0	0.0	3515.0	1848.9	3677.8	101.7	0.0	0.0	0.0	0.0	0.0	0	0	0
Captive Power (GWh)	8326.1	10082.5	12195.7	13000.9	11329.6	10090.1 *	9266.8	9265.9	8209.7	6645.7	5289.3	3677.2	3125.1	2370.4
Captive Takeover(GWh)	0.0	0.0	0.0	0.0	0.0	2583.5 *	2220.8	1269.6	2109.8	2391.5	2014.9	2128.6	921.6	1064.0
Share PLN to Tot.(%)	59.1	57.4	54.8	59.0	68.0	74.7	79.6	82.1	85.7	89.5	92.4	95.2	96.3	97.4
PLN SALES (GWh)	12040.1	13603.8	14801.5	18733.0	24072.7	29829.5	36181.7	42404.5	49335.7	56700.2	64333.5	72744.3	80975.7	90054.9
Growth (%)	25.64	12.99	8.80	26.56	28.50	23.91	21.29	17.20	16.35	14.93	13.46	13.07	11.32	11.21
Share to Tot.(%)	53.7	53.8	53.8	55.9	58.8	60.8	62.3	63.1	63.5	63.9	64.2	64.7	64.8	64.9
Total														
No.of Customers (10^3)	7950.2	8579.9	10005.1	11175.9	12448.7	13851.0	15367.0	17006.3	18776.8	20572.7	22383.2	24209.6	26053.6	27917.3
Power Contracted(MVA)	12593.6	13722.9	17482.7	20263.8	23853.1	25967.8	28379.2	31030.8	33719.1	36408.9	39202.6	42110.5	45206.9	48446.1
PLN SALES (GWh)	22401.8	25282.6	27508.1	33491.0	40905.3	49038.5	58031.3	67234.6	77650.6	88676.6	100166.8	112467.4	124980.4	138760.5
Growth Rate(%)	19.4	12.9	8.8	21.7	22.1	19.9	18.3	15.9	15.5	14.2	13.0	12.3	11.1	11.0
T & D Losses(%)+)	15.2	14.9	14.5	14.0	13.5	13.0	12.5	12.0	11.5	11.0	10.5	10.0	9.7	9.4
Energy Sent Out(GWh)	26660.6	29979.7	32463.2	39279.8	47681.4	56813.3	66821.7	76955.7	88345.6	100293.1	112621.3	125706.2	139193.5	153999.1
Plant Use (%)	4.9	4.9	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Production (GWh)	28047.8	31524.4	34171.8	41347.2	50190.9	59803.4	70320.1	81006.0	92995.3	105599.5	118608.7	132380.8	146519.5	162104.3
Load Factor (%)	70.1	76.1	69.1	69.2	69.3	69.4	69.5	69.6	69.9	71.6	71.7	71.8	71.9	72.0
Peak Load (MW)	4565.4	4728.0	5645.3	6821.1	8267.8	9837.0	11550.2	13286.3	15187.3	16842.6	18881.2	21038.5	23258.8	25701.4

MEI 1992 DKL,DIVSIS

* Actual # Estimated (1991/92)

Demand forecasts table doen in 1992 for the Java-Bali system as part of the PLN 10-year business plan. *Source: PLN.*

Chapter 3

Effect of Electrification on Local Industry

3.1 Introduction

In this chapter, I examine the effect of electrification on desa-level industrial outcomes. I investigate what happens to manufacturing activity in the desa when the grid arrives by looking at the number of manufacturing firms, number of workers in manufacturing, and manufacturing output in section 3.2.

In order to understand the mechanisms through which electrification affects local industry, I look at how firm turnover, as measured by the entry and exit rates of firms, is affected by electrification in section 3.3. A change in firm turnover could mean that electrification is changing the composition of firms in the industry by affecting barriers to entry. By focusing on the extensive margin of electrification (grid expansion), the aim is therefore to see whether electrification has any effect of the extensive margin of industrialization (firm entry and exit).

Finally, an important question that arises in any spacial analysis is whether electrification creates new industrial activity or it reorganizes industrial activity across space. I address this question by conducting various empirical tests in section 3.4.

3.2 Electrification and Local Industrial Outcomes

I examine whether the expansion of the grid affected the number of firms, manufacturing employment and manufacturing output at the desa level. To this end, I use the data-set of all desas in Java between 1990 and 2000 which I described in chapter 2. To recap, I superimpose the firm-level data on a digital map of desas in Java to get the number of firms per desa per year, as well as number of workers in manufacturing, and industrial output. I combine that with data desa-level characteristics including non-energy infrastructure and geography variables. The resulting data-set is a balanced panel of $24,824$ desas over 11 years.

3.2.1 Desa-Level Manufacturing Outcomes

The three columns of table 3.1 shows the OLS, IV and reduced-form regression results for three desa-level outcomes as in specification (2.1): number of firms, total number of workers in the manufacturing sector, and total manufacturing output. Because I have many desas that don't have any medium or large manufacturing firm, hence many zero values, I use the level of these variables instead of the log (See table 3.7 in the appendix for results with zero-preserving log transformations).

Table 3.1: Impact of access on desa level outcomes.

Sample: Desa-Level			
	(1)	(2)	(3)
Dependent Variable	No of Firms	No of Workers in Manufacturing	Output Billion IDR
Panel A: OLS			
Access_{vpt}	0.378***	74.64***	3.973***
	(0.0288)	(6.196)	(0.491)
Panel B: IV			
Access_{vpt}	0.887*	513.9***	39.74***
	(0.480)	(113.8)	(8.175)
First Stage F	118.7	118.7	118.7
Panel C: Reduced Form IV			
Z (KM)	-0.00148*	-0.856***	-0.0662***
	(0.000793)	(0.176)	(0.0125)
Observations	261,470	261,470	261,470
Year FE	✓	✓	✓
Province FE	✓	✓	✓
Geo Controls	✓	✓	✓
Mean Dep Var	0.84	110	6.7

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, IV and reduced-form regressions of equation (2.1). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Across all outcome variables, the OLS estimates in Panel A are positive and significant, suggesting that there is a positive correlation between access to electricity and industrial outcomes. Compared to the IV estimates in Panel B, OLS is consistently smaller in magnitude. This result is in line with the infrastructure literature both on electrification (e.g. Dinkelman (2011), and Lipscomb, Mobarak, and Barham (2013)) and transport (Baum-Snow (2007), Duranton and Turner (2012), and Duranton, Morrow, and Turner (2014)) indicating that infrastructure is allocated to less productive areas. This means that the OLS estimates will underestimate the effect of electrification on manufacturing, as the results show. However, the difference in magnitude between the OLS and the IV estimates is surprisingly large. Before discussing potential reasons in section 3.2.2, I first turn to the interpretation of the IV estimates.

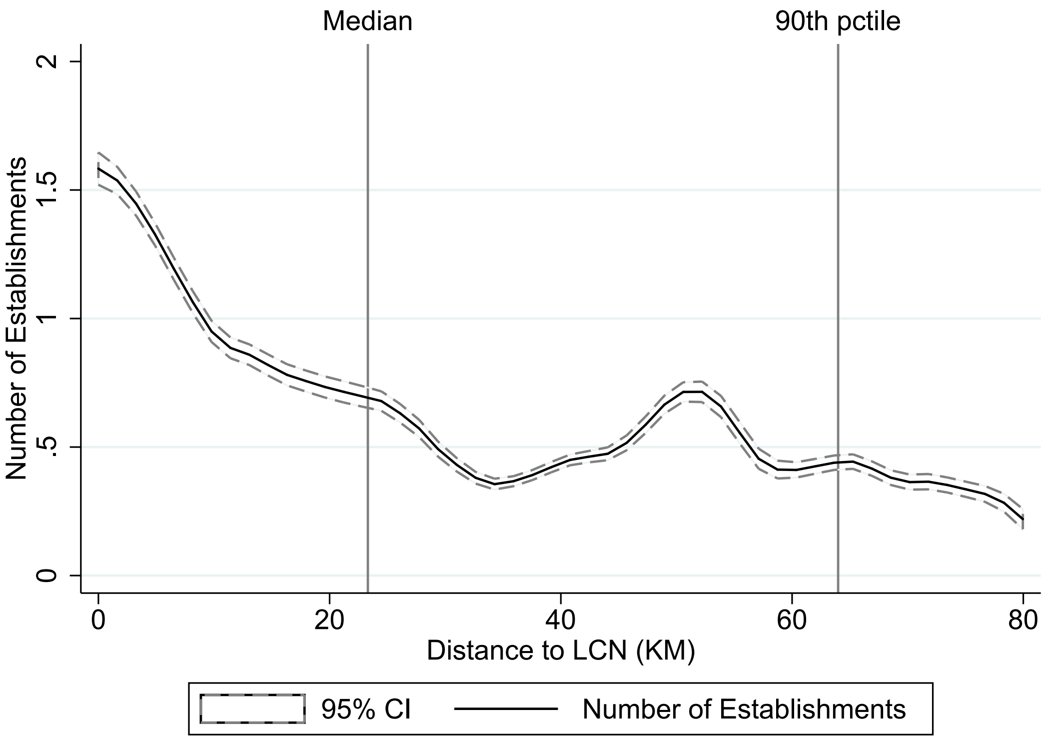
The IV estimates in Panel B are positive and significant. The coefficient in column (1) in panel B says that the causal effect of grid access on the number of firms in a desa is an increase of 0.9 firm. Considering that the average number of firms per desa in the sample is 0.84, this effect is large and around 100% increase over the average. Theoretically, a larger number of firms is associated with a tougher competition. Therefore, electrification potentially intensifies competition by increasing the number of active producers.

Similarly for the number of workers and manufacturing output, the IV estimates in columns (2) and (3) are positive, large and strongly significant. A caveat is that I don't observe the universe of manufacturing firms, but instead I observe the universe of medium and large manufacturing firms with 20 or more employees. To mitigate this issue, for the number of firms, I use the reported start year of production in the survey as opposed to the first year I observe the firm in the data. I take that into account when calculating the total number of firms in a desa which greatly alleviates this issue.[1] As for the total number of workers in manufacturing and manufacturing output, I don't observe any information for these firms before they are in the survey. Therefore coefficients in panel B columns (2) and (3) should be interpreted as the causal difference in the number of workers and manufacturing output between electrified and non-electrified desas with Medium and Large manufacturing firms. Panel C of table 3.1 presents the reduced-form regressions from regressing desa outcomes on the instrument, distance to the hypothetical grid. Coefficients in columns (1), (2) and (3) all show the closer a desa is to the least cost network, the larger the number of firms, number of manufacturing workers and manufacturing output.

Figures 3.1, 3.2 and 3.3 illustrate this negative relationship (unconditional) and show the kernel regression of the of number of manufacturing firms, number of workers and manufacturing output as a function of the distance to the hypothetical least cost grid. The relationship between each of these desa-level outcome variables and the distance to the hypothetical grid is negative, illustrating the reduced-form effect of the instrument on the outcome variables.

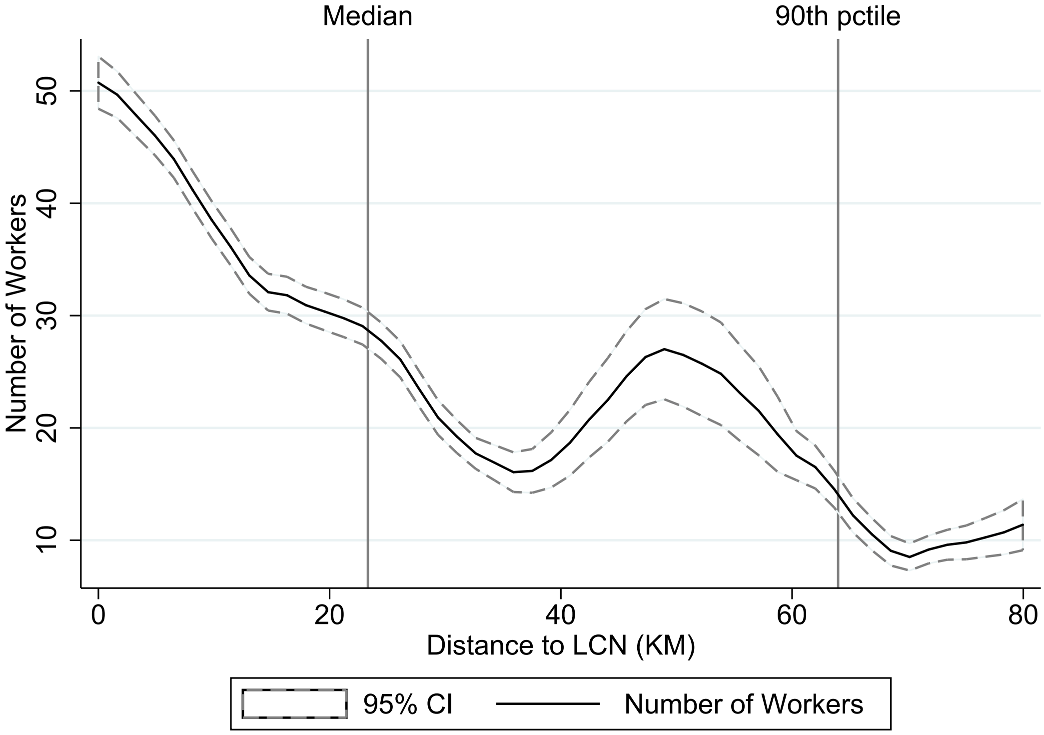
[1] Of course, I still don't observe those firms that exited before they reached the threshold to be included in the survey. This is however not a major concern as these firm are naturally small both in number of workers and probably in production relative to the total manufacturing sector.

Figure 3.1: Distance to Hypothetical Grid and Number of Manufacturing Firms



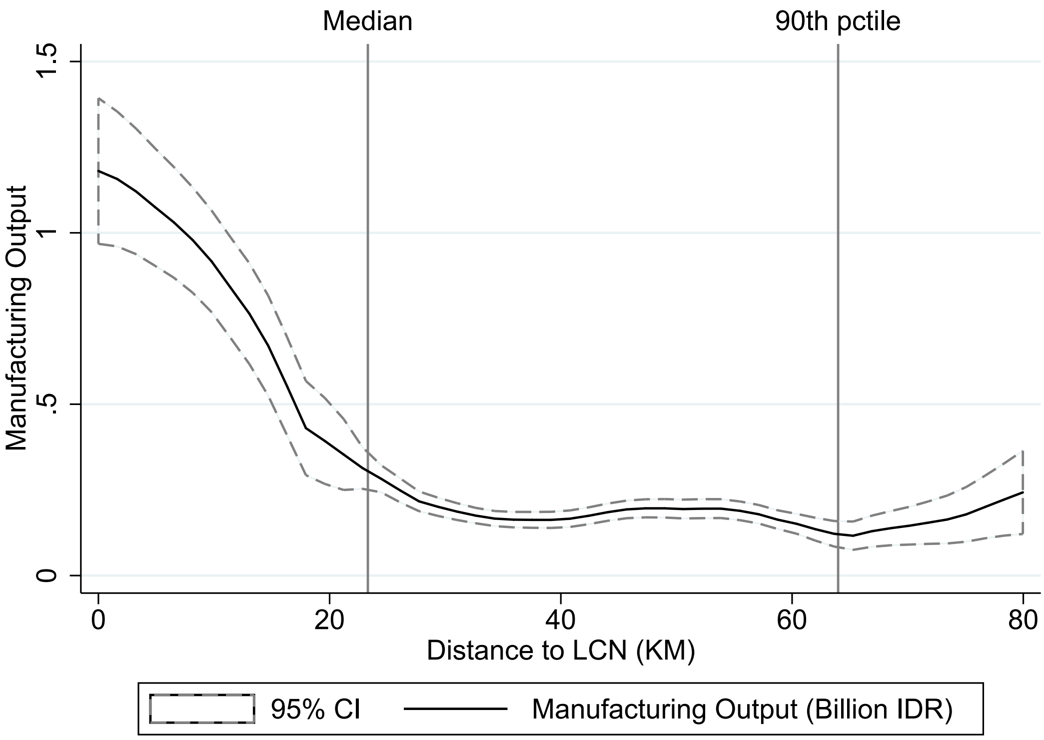
The y-axis presents the number of manufacturing firms at the desa level as a function of the distance of that desa to the hypothetical least cost grid. This is estimated using an Epanechnikov kernel function with a bandwidth of 2.42. The x-axis shows the distance from the desa to the hypothetical least cost grid. The median and the 90^{th} percentile of the distance to the hypothetical grid are shown for reference.

Figure 3.2: Distance to Hypothetical Grid and Number of Manufacturing Workers



The y-axis presents the number of manufacturing workers at the desa level as a function of the distance of that desa to the hypothetical least cost grid. This is estimated using an Epanechnikov kernel function with a bandwidth of 3.35. The x-axis shows the distance from the desa to the hypothetical least cost grid. The median and the 90^{th} percentile of the distance to the hypothetical grid are shown for reference.

Figure 3.3: Distance to Hypothetical Grid and Manufacturing Output



The y-axis presents the manufacturing output (Billion IDR) at the desa level as a function of the distance of that desa to the hypothetical least cost grid. This is estimated using an Epanechnikov kernel function with a bandwidth of 5.02. The x-axis shows the distance from the desa to the hypothetical least cost grid. The median and the 90^{th} percentile of the distance to the hypothetical grid are shown for reference.

3.2.2 Magnitude of Estimated Coefficients.

The direction of the OLS bias I find is common in the infrastructure literature as discussed in the previous section. However, the difference in magnitudes between the IV estimates and the OLS estimate is rather large, and calls for a discussion. I will present and discuss four potential reasons for this difference.

The first and most concerning reason is a violation of the exclusion restriction. The validity of any instrumental variable strategy rests on the assumption that the instrument is excluded, meaning that the instrument only affects the outcome variable through its effect on the endogenous treatment variable. In this setting, this means that the distance to the hypothetical grid, conditional on controls, only affects industrial outcomes through its effect on access to the actual grid. Unfortunately this assumption cannot be directly tested and we would have to rely on economic reasoning to understand how likely it is that there is a violation. There are largely two types of variables that could affect both the distance to the hypothetical least cost grid and industrial outcomes. The first is other types of infrastructure such as access to roads. The second group is local geography. To ensure that the exclusion restriction is not violated, I include an extensive set of controls for both types of variables in all empirical specifications, as outlined in the first chapter of this thesis. In addition to geographic and infrastructure controls, I also control for other political and economic characteristics. The results from section 2.5.4 in chapter 2 with the placebo grid and the Euclidean grid alleviate this concern and show that local geography does not drive the correlation between access and the distance to the least cost grid.

To test whether there are other time-invariant factors that could be driving the correlation between the instrument and access, I run specification (2.1) again but including desa-level fixed effects:

$$Y_{vpt} = \alpha + \beta Access_{vpt} + \eta \mathbf{V}_{vpt} + \gamma_d + \delta_{pt} + \epsilon_{vpt} \tag{3.1}$$

where γ_d is the desa fixed effect and δ_{pt} is a province-by-year fixed effect.

Since the instrument is also time-invariant, I interact it with year dummies. The variation used here is different than in table 3.1: when instrumenting the the distance to the hypothetical grid interacted with year dummies, I exploit *time* variation in how the instrument explains access. I still include all the time-varying desa-level controls as before. Results are presented in table 3.2. As before, the OLS estimates in panel A are downward biased. The IV estimates in panel B show that electrification causes industrial outcomes to increase. Panel C presents the reduced form regression of outcomes on the instrument Z interacted with time dummies. The coefficients indicate that the closer a desa is to the hypothetical least cost grid, the more industrial activity it has, and this relationship is consistent over time.

Given this rich set of controls and the evidence from the various empirical tests presented in this chapter and the previous chapter, it is unlikely that a violation of the exclusion restriction is driving the difference in magnitudes between the IV and OLS estimates.

Table 3.2: Impact of access on desa level outcomes - fixed effects.

Sample: Desa-Level			
	(1)	(2)	(3)
Dependent Variable	No of Firms	No of Workers in Manufacturing	Output Billion IDR
		Panel A: OLS	
Access$_{vpt}$	-0.0450***	-6.532*	-2.812***
	(0.0112)	(3.829)	(0.571)
		Panel B: IV	
Access$_{vpt}$	3.010***	942.2***	127.4***
	(0.338)	(113.3)	(16.20)
First Stage F	73.83	73.83	73.83
		Panel C: Reduced Form IV	
Zx1991	-0.000926***	-0.389***	-0.0132***
	(9.21e-05)	(0.0443)	(0.00238)
Zx1992	-0.00151***	-0.623***	-0.0325***
	(0.000148)	(0.0671)	(0.00479)
Zx1993	-0.00194***	-0.731***	-0.0637***
	(0.000182)	(0.0901)	(0.0223)
Zx1994	-0.00250***	-0.911***	-0.0600***
	(0.000226)	(0.0950)	(0.00927)
Zx1995	-0.00326***	-0.903***	-0.0853***
	(0.000283)	(0.227)	(0.0124)
Zx1996	-0.00386***	-1.287***	-0.0751***
	(0.000335)	(0.117)	(0.0106)
Zx1997	-0.00410***	-1.279***	-0.0870***
	(0.000379)	(0.128)	(0.0116)
Zx1998	-0.00417***	-1.222***	-0.258***
	(0.000428)	(0.129)	(0.0319)
Zx1999	-0.00461***	-1.328***	-0.313***
	(0.000452)	(0.137)	(0.0433)
Zx2000	-0.00513***	-1.468***	-0.390***
	(0.000483)	(0.151)	(0.0498)
Observations	261,470	261,470	261,470
Desa FE	✓	✓	✓
ProvinceYear FE	✓	✓	✓
Geo Controls	✓	✓	✓
Mean Dep Var	0.84	110	6.7

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, IV and reduced-form regressions of equation (3.1). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, distance to road, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

The second possible reason is a technical one that is somewhat common in two-stage least square (2SLS) strategies with a binary endogenous variable, access in this case. If the first stage of the 2SLS estimation gives predicted values for the binary endogenous variable that are outside the $[0,1]$ range, then this could lead to inflated second stage coefficients. This is not the case in this paper, where the 1^{st} and the 98^{th} percentiles of the predicted values in the first stage are between 0 and 1[2].

The third reason, which is the most likely reason, is a compliers' issue. Given that I am estimating a local average treatment effect of access on industrial outcomes; this difference in magnitudes is potentially driven by a complier sub-population of desas that would benefit *more* from electrification. For instance, is it possible that compliers are different from the average electrified desa in Java. This is because the decision to electrify a desa is affected by political and socioeconomic conditions. Complier desas are those desas that get access to the grid because the cost of extending the grid to them is low, and not because of confounding political, economic, or social reasons. Given that the compliance of these desas is based on the low cost of electricity provision, it may well be that these desas will experience higher returns to electrification. Second, the compliers in my empirical strategy are more likely to have firms in more electricity intensive industries, and these industries would naturally benefit more from electrification.

The fourth possible reason is measurement error. Measurement error in the access variable could lead to an attenuation bias in the estimated OLS coefficient. I am not able to rule this out, especially that the access definition in this chapter is a rough one. However, results from the firm-level analysis in the next chapter, where I use a more accurate definition of access and still get a large difference between IV and OLS estimates, indicate that measurement error is unlikely to be severe in this case.

Now that I have discussed reasons for the large difference between OLS and IV estimates, it is important to ask whether the IV estimates are sensible. In other words, are the IV estimates too large, irrespective of how they compare to the OLS estimates? Looking at the bottom two rows of table 3.1, it is clear that the unconditional average number of firms is low. This is driven by the fact that many desas have zero firms. Conditional on having a positive number of firms (bottom row), the effect of

[2] Source: author's calculation.

access on the number of workers in manufacturing and manufacturing output do not appear so large. In fact, the estimated IV coefficients for these variables is similar to the difference between desas that have zero firms and the average desa with a positive number of firms. Therefore, the effect of electrification on local industry is comparable to and could be interpreted as moving from a desa with no firms to the average industrialized desa.

3.3 Electrification and Firm Turnover

The availability of the grid in a desa may affect the attractiveness of this particular desa to entrepreneurs who are considering to start a firm. As shown in section 3.2, electrification causes the total number of firms in a desa to increase. I now investigate the role of entry and exit as drivers of this increase.

Columns (1) and (2) of table 3.3 look at the effect of access on firm turnover. The first outcome is entry rate, defined as the ratio of entrants to the total number of firms. The second outcome variable is the exit rate, defined as the ratio of exiting firms to the total number of firms. These outcomes are only defined for desas with a positive number of firms. As before, the OLS estimates in panel A are positive and smaller in magnitude than the IV estimates in panel B, and are therefore downward biased. Focusing on panel B, the IV estimate in column (1) show that access to the grid increases firm entry rate by around 10%. Interestingly, in column (2), the coefficient on access shows that the exit rate *also* increases due to electrification, although by a smaller amount than the entry rate. This is consistent with the an increase in the total number of manufacturing firms from column (1) in table 3.1. Electrification therefore increases firm turnover, leading to more churning in a given desa. Higher churning is a sign of efficiency where firm selection into and out of the desa is at work.

Table 3.3: Impact of access on desa level turnover.

Sample: Desa-Level		
	(1)	(2)
Dependent Variable	Entry Rate	Exit Rate
Panel A: OLS		
Access$_{vpt}$	0.00719***	0.00171***
	(0.00263)	(0.000581)
Panel B: IV		
Access$_{vpt}$	0.106***	0.0157**
	(0.0284)	(0.00658)
First Stage F	58.39	58.39
Panel C: Reduced Form IV		
Z (KM)	-0.000249***	-3.68e-05**
	(6.00e-05)	(1.50e-05)
Observations	54,210	54,210
Year FE	✓	✓
Province FE	✓	✓
Geo Controls	✓	✓
Mean Dep Var	0.07	0.01

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, IV and reduced-form regressions of equation (2.1). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

These findings suggest that the extensive margin of electrification induces long-run firm responses; entry and exit. Interpreting the results in this section, the extensive margin of electrification therefore affects the extensive margin of industrialization, or firm entry, by increasing entry rates. In a competitive environment, more entry can lead to more exit as relatively unproductive incumbents will be less likely to survive. Therefore, electrification also increases exit rates.

3.4 Electrification and Relocation of Industrial Activity

The results in the previous section indicate that electrification increases industrial activity at the desa-level by attracting more firms. To learn about the aggregate effect of electrification, one important question is thus whether these firms are new firms or whether they are firms that have relocated from other non-electrified desas. In particular, it is interesting to understand if these firms would have existed anyway, regardless of electrification. In the case where firms would relocate, the effect of electrification would be a reorganization of economic activity across the island as opposed to creation of *new* economic activity; meaning that the aggregate effect of electrification is small or negligible.

Put differently, a potential concern is that the stable unit treatment value assumption (SUTVA) is violated in the identification strategy in this analysis. SUTVA requires that the treatment applied to one unit does not affect the outcome for another unit. If electrifying one desa (or firm) will create firm relocation or business stealing for competitors (because of lower prices), then SUTVA is violated. The presence of these spillovers across different desas complicates the interpretation of my results. Electrifying one desa can have an effect on firms in other desas, and these effects are likely to be negative. What I estimate as the average difference between electrified and non-electrified desas could be therefore a combination of creation of new economic activity and displacement of economic activity from those that don't get electrified (or are already electrified) to desas that get newly electrified.

In the following subsections, I attempt to address the question of whether electrification creates new economic activity or whether it is relocating economic activity. I start by looking at the possibility of firm relocation.

3.4.1 Relocation of Incumbent Firms

Can electrifying a new desa induce firms in non-electrified desas to close their factories and move them to the newly electrified desa? This could happen if a firm finds in profitable to do so, i.e. when the cost of relocation is smaller than the benefit of relocating. Firms choose to locate in certain desas presumably because the benefits from being in that location are the highest for that particular firm (e.g. local knowledge, home bias, etc.), so moving would be costly, in addition to the physical relocation costs.

Unlike a network of highways or subways, access to the electrification infrastructure is not restricted to particular locations such as a train station or a highway entrance. There is no technological limit on where the grid can go. In the context the island of Java, even if a desa is faraway from the grid at a certain point in time, it will eventually be connected to the grid. Given that this is a period of rapid expansion of the grid in Java, eventually all desas became connected to the grid. So unless the firm is really impatient, the benefit of moving to an electrified desa today versus waiting to get access in the future is unlikely to be a profitable action. Confirming this insight, I observe no firm movements across desas in the dataset[3],[4].

Finally, the evidence from desa-level regressions in table 3.3 column (2) shows that there is more exit in electrified desas. If firms were shutting down their factories in non-electrified desas and moving them to electrified desas, then the exit rates would be higher in non-electrified desas. Results show the opposite. This result on exit rates is thus evidence against exit of firms from non-electrified desas to electrified desas.

3.4.2 Empirical Tests

To test whether relocation of firms is important in this context, I perform three main empirical tests. Given the technology argument made above and the rapid grid expansion, relocation is likely to happen at a local geographic level where the benefits from being in different desas are comparable within a certain proximity. This argument applies both to incumbent firms as well as entrants. In fact, it is expected for these local spillover effects to be larger for entrants since these do not need to incur a physical cost of relocation.

First, I estimate equation (2.1) at the district[5]-level, a higher administrative division than a desa[6]. If spillovers are prominent, then the estimates should be smaller at the district-level. Table 3.4 presents the OLS and IV results. For comparability with the desa-level results in table 3.1, I use the average number of firms, average number of manufacturing workers and average manufacturing output in a district as opposed to the total[7] in columns (1), (2) and (3) as the dependent variables. In columns (4)

[3] Less that 5% of the firms change desas between 1990-2000. I exclude these firms from the analysis.

[4] Another possibility is that entrepreneurs could be closing their factories in non-electrified desas and opening new factories producing *different* products in electrified desas. In this case, the firm will show up with a new firm identifier in the data, and it will be counted it as an exiting firm from the non-electrified desa and a new entry in the electrified desa. However, since I don't observe the identity of the owners, it is not possible for me to track this firm. Given that it is producing a different product, it wouldn't be unreasonable to consider this firm as a new firm.

[5] Kecamatan in Bahasa

[6] The average number of desas per district is 16.

[7] Results are similar when using the total then dividing by average number of desas in a district.

and (5), I present the results for the entry and exit rates, defined as the total number of entrants and exiting firms divided by the total number of firms at the district-level, respectively. Comparing to the desa-level results, the effect of access on these industrial outcomes at the district level is very close to the effect at the desa-level. The estimated coefficients are if anything somewhat larger that the estimated coefficients from table 3.1, meaning that relocation of economic activity within district is unlikely. The IV results in Panel B therefore confirm that spillovers or relocation of economic activity are not prominent in this context.

Table 3.4: Impact of access on district level outcomes.

Sample: District-Level					
	(1)	(2)	(3)	(4)	(5)
Dependent Variable	No. of Firms	No. of Workers in Manufacturing	Output Billion IDR	Entry Rate	Exit Rate
Panel A: OLS					
Access_{dt}	0.447***	3.312***	85.95***	0.00738*	0.00254***
	(0.0716)	(0.818)	(13.51)	(0.00395)	(0.000756)
Panel B: IV					
Access_{dt}	1.616*	39.05***	617.5***	0.101***	0.0143*
	(0.846)	(13.43)	(229.6)	(0.0367)	(0.00737)
First Stage F	20.12	20.12	20.12	19.73	19.73
Observations	17,941	17,941	17,941	13,407	13,407
Year FE	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓
Mean Dep Var	1.08	153	8.9	0.072	0.009

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, and IV regressions of equation (2.1) at the district level. Geographic controls are defined at the district level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the district level. Access is defined as a dummy equal to 1 if at least 50% of desas in the district are within 15Km of the closest substation.

Second, I test if an increase in the number of neighboring desas that switch from being non-electrified to electrified in a certain year negatively affects the number of firms and the number of entrants in desas that are not electrified and that remain so. If there are any relocation effects, I would be expect them to be largest for this sub-sample.

I run the following specification where I test the effect of N^S_{vpt}, the number of switching neighboring desas on desa outcome Y_{vpt}, conditional on the total number of neighboring desas N_{vp} defined as the number of desas within a 7 km radius of the desa.

$$Y_{vpt} = \alpha + \beta N^S_{vpt} + \theta N_{vp} + \mu Z_{vp} + \eta \mathbf{V}_{vpt} + \gamma_p + \delta_t + \epsilon_{vpt} \tag{3.2}$$

Of course, N^S_{vpt} is endogenous. I instrument N^S_{vpt} with the average distance of neighboring desas to the hypothetical grid[8], conditional on the desa's distance to the least cost hypothetical grid Z_{vp}.

[8]Variation in the shape of the grid across space means that the average neighbors distance to the grid and the desa's own distance to the grid are not perfectly collinear. Interacting the IV with time dummies also helps with power.

Table 3.5: Relocation of Economic Activity Desa-Level

Sample: Desa-Level				
	(1)	(2)	(3)	(4)
Dependent Variable	No of Firms	No of Entrants	Entry Rate	Exit Rate
Panel A: OLS				
N^S_{vpt}	0.0082	-0.000199	-0.00035	-9.5e-05
	(0.0051)	(0.00033)	(0.00049)	(8.45e-05)
N_{vp}	0.0097***	0.00066***	-0.000160	-3e-05
	(0.00132)	(0.000104)	(0.000104)	(2.00e-05)
Z(KM)	-0.00019	-3.2e-06	-0.0002*	2.9e-06
	(0.000814)	(6.82e-05)	(0.000106)	(2.10e-05)
Panel B: IV				
N^S_{vpt}	-0.0177	0.00349	0.000424	-0.00114
	(0.0135)	(0.00349)	(0.00363)	(0.0008)
N_{vp}	0.0101***	0.00061***	-0.0002	-8e-06
	(0.00135)	(9.60e-05)	(0.000129)	(2.66e-05)
Z(KM)	-0.00019	-4.31e-06	-0.000203*	5e-06
	(0.00081)	(6.8e-05)	(0.0001)	(2e-05)
First Stage F	40.60	40.60	12.44	12.44
Observations	113,312	113,312	15,446	15,446
Year FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓
Mean Dep Var	0.39	0.03	0.08	0.006
Mean N^S_{vpt}	0.38	0.38	0.53	0.53
Mean N_{vt}	35	35	42	42

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, and IV regressions of equation (3.2). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Table 3.5 shows the OLS and IV results for this first test. Panel B column (1) shows the IV estimate for the effect of an increase in the number of switching neighbors on the number of firms in the desa. The coefficient is statistically indistinguishable from zero and is small in magnitude. Give the mean number of switching neighbors in a given year for a given desa, this says that when one neighbor gets electricity in a certain year, the number of firms decreases by 0.007 firms; approximately zero. The coefficient in Panel B column (2) shows the same IV regression for the number of entrants. The estimated effect is small and insignificant, but also positive. This shows that if a neighboring desas gets electrified, that does not decrease the number of entrants in the non electrified desa. Columns (3) and (4) panel B show the IV estimates for entry and exit rates. Results indicate that there is no effect of switching neighbors on firm turnover. In the appendix to this chapter, section 3.A, I show the same test in table 3.8 for the sub-sample restricting the sample to positive number of switching neighbors, where the effects should be larger. The results are similar and do not show any evidence for local spillovers.

Finally, I repeat the desa-level analysis from equation (2.1) but jointly estimating the main effect of access $Access_{vpt}$ and the spillover effect N^C_{vpt}. N^C_{vpt} is defined as the number of connected neighboring desas. I also condition on the total number of neighboring desas N_{vp}.

$$Y_{vpt} = \alpha + \beta Access_{vpt} + \mu N^C_{vpt} + \theta N_{vp} + \eta \mathbf{V}_{vpt} + \gamma_p + \delta_t + \epsilon_{vpt} \tag{3.3}$$

The coefficient on N^C_{vpt} will therefore measure the effect of having an additional electrified neighboring desa on desa outcome Y_{vpt}. If $\hat{\beta}$ and $\hat{\mu} * \bar{N^C_{vpt}}$ sum up to zero, where $\bar{N^C_{vpt}}$ is the average number of connected neighboring desas, then the effect of electrification evaluated at the average number of connected neighbors is only a relocation one. Otherwise, if the sum of $\hat{\beta}$ and $\hat{\mu} * \bar{N^C_{vpt}}$ is larger than zero, then electrification creates *new* economic activity. As before, I instrument access with the desa's own distance to the hypothetical grid, and the number of connected neighbors by the average distance of neighbors to the hypothetical grid, both interacted with time dummies to aid with power.

Table 3.6: Access and spillover effects at the desa-level.

Sample: Desa-Level					
	(1)	(2)	(3)	(4)	(5)
Dependent Variable	No. of Firms	No. of Workers in Manufacturing	Output Billion IDR	Entry Rate	Exit Rate
Panel A: OLS					
Access_{vpt}	0.234***	21.03	2.816**	0.00959**	-0.00033
	(0.0577)	(13.46)	(1.170)	(0.00395)	(0.0009)
N^C_{vpt}	0.0014	1.607***	0.049	-3.8e-05	4.9e-05***
	(0.00161)	(0.389)	(0.0357)	(7.6e-05)	(1.8e-05)
N_{vp}	0.00804***	-0.621***	-0.0620***	-8.2e-05	-8.4e-06
	(0.00134)	(0.190)	(0.0125)	(7.6e-05)	(1.7e-05)
Panel B: IV					
Access_{vpt}	2.001**	545.3**	100.7***	0.152***	0.031**
	(0.886)	(222.5)	(19.68)	(0.053)	(0.0148)
N^C_{vpt}	-0.0318	-6.407	-2.916***	-0.002**	-0.0007***
	(0.0249)	(5.775)	(0.555)	(0.001)	(0.0003)
N_{vp}	0.0277	3.617	1.998***	0.00127*	0.0006***
	(0.0172)	(3.982)	(0.389)	(0.00076)	(0.0002)
First Stage F	39.63	39.63	39.63	5.078	5.078
Observations	261,470	261,470	261,470	54,210	54,210
Year FE	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓
Mean Dep Var	0.84	110	6.7	0.07	0.01
Mean N^C_{vpt}	27.8	27.8	27.8	39.6	39.6
Mean N_{vt}	43	43	43	52	52
P-value of joint effect	0.0016	0.00	0.011	0.0013	0.17

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV regressions of equation (3.3). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level. The p-value in the last row corresponds to the null of H0: $\hat{\beta} + 27.8 * \hat{\mu} = 0$.

Table 3.6 presents the OLS and IV results of equation (3.3). Focusing on the IV results in panel B, the estimated coefficients across all industrial outcomes are comparable to the IV results in table 3.1. The effect of access on industrial outcomes is positive and significant. On the other hand, the IV estimate for the effect of the number of connected neighbors N^C_{vpt} is small and negative, but not always significant. It is significant only in columns (3), (4) and (5). This indicates that spillovers are stronger in the output market, consistent with high relocation costs of firms and workers. The last row of table 3.6 presents the p-value of the joint test where the null is $H0$: $\hat{\beta} + \hat{\mu} * \bar{N^{C}_{vpt}} = 0$. The null is rejected in columns (1) to (4). This indicates that indeed electrification does create new economic activity, and the effects are not restricted to relocation of economic activity.

3.5 Conclusion

I find that electrification causes industrial activity to increase in the desa, manifested by an increase in the number of manufacturing firms, workers in manufacturing, and manufacturing output. Highlighting the economic mechanism in play, I show that electrification increases entry rates, but also exit rates. Electrification therefore increases firm turnover, a potential sign of a healthier market allocation mechanism where there is tougher competition in the market. I find that the increase in industrialization at the desa-level is mostly due to the creation of new manufacturing activity rather than a reorganization of economic activity across space.

An important conclusion from the empirical analysis in this chapter is that the extensive margin of electrification driven by the expansion of the grid is a significant driver of industrialization. Electrification leads to the creation of new firms, which in turns forces inefficient incumbents to exit. Uncovering this economic mechanism - firm turnover - is informative for policy: it clarifies when and how electrification can lead to industrial development. It also implies that if there are other significant non-energy barriers to entry that prevent extensive margin effects of industrialization, the effect of electrification is likely to be smaller.

Appendix

3.A Additional Results

Table 3.7: Impact of electrification on desa level industrial outcomes - Log transformations.

	Sample: Desa-Level					
	(1)	(2)	(3)	(4)	(5)	(6)
	Log(1+N)	Log(h(N))	Log(1+W)	Log(h(W))	Log(1+Y)	Log(h(Y))
	Panel A: OLS					
Access_{vpt}	0.103***	0.131***	0.329***	0.368***	0.126***	0.149***
	(0.00553)	(0.00699)	(0.0171)	(0.0193)	(0.00611)	(0.00725)
	Panel B: IV					
Access_{vpt}	0.210**	0.266**	0.918***	0.961***	0.774***	0.906***
	(0.0983)	(0.125)	(0.307)	(0.346)	(0.125)	(0.147)
First Stage F	118.7	118.7	118.7	118.7	118.7	118.7
	Panel C: Reduced Form IV					
Z (KM)	-0.000350**	-0.000443**	-0.00153***	-0.00160***	-0.00129***	-0.00151***
	(0.000162)	(0.000206)	(0.000500)	(0.000566)	(0.000181)	(0.000214)
Observations	261,470	261,470	261,470	261,470	261,470	261,470
Year FE	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV and Reduced-Form regressions of two different measures of log transformations that preserve zeros for the number of firms N, number of workers in manufacturing W and total manufacturing output Y. The first transformation is a $log(1+X)$. The second transformation is $log(h(X))$ where $h(X) = X + (X^2+1)^{\frac{1}{2}}$ following Liu and Qiu (2016). Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Table 3.8: Relocation of Economic Activity Desa-Level; Positive Number of Switching Neighbors.

Sample: Desa-Level				
	(1)	(2)	(3)	(4)
Dependent Variable	No of Firms	No of Entrants	Entry Rate	Exit Rate
Panel A: OLS				
N^S_{vpt}	0.0178**	5.59e-05	-0.000462	-8.16e-05
	(0.00851)	(0.000490)	(0.000958)	(0.000181)
N_{vp}	0.0107***	0.000735***	-3.70e-05	1.60e-05
	(0.00226)	(0.000268)	(0.000441)	(9.51e-05)
Z(KM)	0.00481**	0.000108	8.38e-05	1.69e-06
	(0.00221)	(0.000210)	(0.000667)	(0.000146)
Panel B: IV				
N^S_{vpt}	0.0211	0.00559	0.00513	-0.00109*
	(0.0301)	(0.00681)	(0.00496)	(0.000573)
N_{vp}	0.01000	-0.000452	-0.00130	0.000244
	(0.00714)	(0.00146)	(0.00123)	(0.000163)
Z(KM)	0.00486**	0.000196	0.000580	-8.81e-05
	(0.00225)	(0.000209)	(0.000829)	(0.000161)
First Stage F	8.309	8.309	5.556	5.556
Observations	4,636	4,636	706	706
Year FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓
Mean Dep Var	0.42	0.02	0.07	0.006
Mean N^S_{vpt}	9.2	9.2	11.6	11.6
Mean N_{vt}	43	43	52	52

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, and IV regressions of equation (3.2), restricting the sample to those desas with a positive number of switching neighbors. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Chapter 4

Electrification and Firm Performance

4.1 Introduction

So far, results show that the expansion of the electricity grid caused an increase in manufacturing activity and increased firm turnover in Java. Is this increase in manufacturing due just to an increase in the number of manufacturing firm or is firm size also affected by access? In other words, does electrification increase industrial activity by attracting the same type of firms or are the firms in electrified areas are different in terms of their performance? To answer this question, I make use of the firm-level manufacturing census and I analyze the effect of access at the desa-level on firm outcomes.

I start by looking at the effect of access on firm output and inputs. I then look at whether firm survival is affected by access for consistency with the turnover results from the previous chapter. Finally, I check if there are any business stealing effects at the firm-level as a test of spillovers.

4.2 Electrification and Firm-Level Outcomes

In this section, I test how average firm performance measures respond to access to electricity. I investigate whether firms in electrified desas are different. I begin by looking at firm output and inputs. I then look at how electrification is affecting firm turnover by looking at firm-level exit probabilities and the age distribution of firms. If firm turnover is a mechanism through which electrification affects manufacturing activity, the probability of exit would be higher and firms would be younger in electrified desas.

4.2.1 Output and Inputs

I first present the estimation results of specification (2.2) for different firm-level outcome variables. Table 4.1 shows the OLS, IV and reduced-form versions of specification (2.2) for the log values of firm-level deflated sales, deflated capital, wage bill, number of workers, energy bill and quantity of electricity consumed in kWh. The treatment variable here again is $Access_{vpt}$, instrumented with Z_v, the distance to the hypothetical least cost grid in kilometers. Table 4.1 panel A presents the OLS results which indicate a positive relationship between average output and inputs and access. The OLS estimates are smaller in magnitude that the IV estimates as before. Panel B shows that electrification causes an increase in average firm output and production inputs. The IV coefficients are all positive and significant at the 1% level. Looking at the first column of Panel B, the causal effect of access on average firm sales is large and positive. Columns (2) to (4) show that access also causes firm input demand for capital and labor (wage bill and number of workers) to increase substantially, with a larger effect on capital relative to labor. Perhaps not surprisingly, the effect on the energy bill in columns (5), which include both spending on electricity and fuels, is the largest. Column (6) shows that firms with access to the grid do indeed consume a substantially greater quantity of electricity in kWh. The fact that electricity consumed increases by more than the increase in the energy bill reassuringly means that the unit price of electricity is lower in electrified areas. Panel C presents the results from the reduced-form regressions. Across all columns, being closer to the hypothetical grid causes all firm-level outcomes to be significantly larger. For robustness, table 4.8 in appendix 4.5 section 4.A repeats the same analysis but using a different definition for access; $Connected_{it}$, This is a dummy variable defined at the firm-level instead of the desa-level and is equal to one if a firm is observed consuming a positive amount of grid electricity in the census. There is still a strong first stage of this different definition of access on the instrument, and the results are similar to those in table 4.1.

Table 4.1: Impact of access on the sales and inputs at the firm level.

Sample: Firm-Level						
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable (Log)	Sales	Capital	Wage Bill	Nb Workers	Energy Bill	Electricity (kWh)
Panel A: OLS						
Access	0.466***	0.416***	0.348***	0.197***	0.447***	0.499***
	(0.0592)	(0.0592)	(0.0422)	(0.0275)	(0.0888)	(0.0933)
Panel B: IV						
Access	2.511***	3.417***	1.788***	1.169***	4.015***	5.125***
	(0.615)	(0.648)	(0.403)	(0.266)	(0.781)	(1.256)
First Stage F	41.55	41.55	41.51	41.55	40.48	30.89
Panel C: Reduced Form IV						
Z (KM)	-0.00665***	-0.00933***	-0.00505***	-0.00336***	-0.0114***	-0.0110***
	(0.00134)	(0.00139)	(0.00102)	(0.000661)	(0.00180)	(0.00204)
Observations	141,615	141,615	141,615	141,615	139,481	120,453
IndustryxYear FE	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, IV, and reduced form regressions of equation (2.2). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Relative to the existing literature, the most readily comparable results to what I find are from Allcott, Collard-Wexler, and O'Connell (2016). In their paper, the authors look at the effect of shortages on firm-level outcomes. They find that a 1 percentage point increase in shortages causes a 1.1% decrease in within firm sales. Access to electricity can be thought of as a 100 percentage points decrease in shortages, which would then translate into a 200% increase in sales revenue[1]. Compared comparable to the Allcott, Collard-Wexler, and O'Connell (2016) result, the effect of electrification on average sales in the desa is much larger. This means that in addition to the within firm effect of electrification on sales, there are large selection effects. The size of the effect confirms the fact that the extensive margin of electricity supply has a bigger effect on the industrial sector relative to the effect of the intensive margin. One explanation is that electrification is likely to reduce entry costs by more relative to improvements in the reliability of electricity supply. If sunk costs of entry are significantly affected by electrification, the effect on average firm outcomes will be larger, because of selection. Lower barriers to entry would attract more entrepreneurs across the whole productivity distribution, leading to tougher selection and therefore more productive firms on average. Allcott, Collard-Wexler,

[1] $\Delta y = \exp(1.1) - 1 = 2$

and O'Connell (2016) also find that shortages do not affect labor input. In contrast, I find a large effect of access on average number of manufacturing workers in the desa, confirming that the extensive margin of electricity has a more considerable effect on the industrial sector.

4.2.2 Input Substitution

I now investigate how electrification affects the firm's input substitution patterns. Electricity is an input of production that is primarily used to power machinery. As electricity becomes cheaper with access, a production technology with substitution across inputs predicts that the firms should substitute away for the other inputs and more towards electricity. An interesting question is therefore whether electrification affects the demand for different inputs differently.

Table 4.2 shows how access to the grid affects firm-level input ratios. As in Table 4.1, the OLS estimates in Panel A are positive but smaller in magnitude relative to the IV estimates in panel B. Column (1) Panel B shows access causes the capital-labor ratio of the firm to increase. From columns (2) and (3), both the energy-capital and energy-labor ratios increase, but the second increases three times as much. This explains the increase in the capital-labor ratio. All these results depict a particular input substitution pattern where capital and energy are complimentary and labor and energy are more substitutable (or at least, there is less substitution between capital and energy than labor and energy).

Table 4.2: Electrification and the firm's input ratios.

	(1)	(2)	(3)
Dependent Variable	log(K/L)	log(E/K)	log(E/L)
Panel A: OLS			
Connected	0.067*	0.0523	0.119*
	(0.038)	(0.0632)	(0.0652)
Panel B: IV			
Access	1.630***	0.808*	2.360***
	(0.392)	(0.463)	(0.541)
First Stage F	41.51	40.48	40.46
Panel C: Reduced Form IV			
Access	-0.0048***	-0.0024*	-0.0069***
	(0.00092)	(0.0013)	(0.0013)
Observations	141,598	139,678	139,664
IndustryxYear FE	✓	✓	✓
Province FE	✓	✓	✓
Geo Controls	✓	✓	✓
Firm Controls	✓	✓	✓

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Results from OLS, IV, and reduced form regressions. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

There are two theoretical reasons that could be driving these differential responses to electrification across inputs. The first is input substitution and different degrees of substitutability between products. When the unit price of an input of production decreases, the overall marginal cost of production decreases, leading to an increase across all input demands, and the increase would be highest for the input which prices has decreased. This is one possible interpretation of the results observed in table 4.1. But if capital is more complementary to electricity than labor, then a decrease in the price of electricity will lead to a larger increase in demand for capital relative to the increase in the demand for labor; thus increasing the capital-labor ratio. If capital and electricity are more complimentary than labor and electricity, when the unit price of electricity falls, this will lead to substitution away from capital and labor towards electricity, but more so for labor. In other words, just as observed in table 4.2, a lower unit price of electricity leads to an increase in the ratios of electricity to the other inputs of production, but the electricity-labor ratio will increase by more than the electricity-capital ratio.

All these effects of electrification can be explained by a decrease in the unit price of electricity and differential substitution patterns, without any changes in the production technology, i.e. the production function coefficients are the same. In chapter 5, I structurally estimate a production function allowing for flexible substitution patterns to plausibility of the above interpretation. A second reason why these substitution patterns might emerge is a technological effect where electrification changes the production function of the firm. I explore this possibility in more detail in chapter 6. In the remainder of this chapter, I focus on readily observable firm-level outcome variables and relegate the analysis that requires prior structural estimation to chapters 5 and chapter 6. Next, I evaluate the effect of electrification on firm survival.

4.2.3 Effect of Access on Incumbent Firms

The estimated coefficients in tables 4.1 and 4.2 represent the average causal difference between outcomes of firms in electrified desas and non-electrified desas. It combines the effect of access on incumbent firms as well as the selection effect of access where electrification potentially systematically more productive firms or less productive firms. To get a sense of how much of the estimated effect of access on firm outcomes is driven by selection of different firms versus an effect on incumbents, I estimate equation 2.2 with firm fixed effects:

$$y_{ivpst} = \alpha + \beta Access_{vpst} + \nu \mathbf{X}_{ivpst} + \eta \mathbf{V}_{vpt} + \gamma_i + \delta_{st} + \epsilon_{ivpst} \tag{4.1}$$

where γ_i is a firm fixed effect. As with the desa-level regression with fixed effect, I use an interaction of the same instrument with time dummies. This is because the hypothetical least cost instrument does not vary over time and will not be able to identify within firm effects. Table 4.3 presents these results. The OLS estimates in panel A are biased towards zero. Focusing on the panel B, column (1), the estimated coefficient of the causal effect of electrification on the incumbents' sales revenue is positive and significant. Electrification causes the firm's sales to increase by 18%. While there is a significant positive effect of access to electricity on firms, this effect is less than a tenth of the estimated coefficient estimated in table 4.1 resulting from specification (2.2). The difference between (4.1) and (2.2) is that the the first estimates the effect of electrification *within* firm, or on incumbents who switch from not being connected to being connected to the grid, while the second estimates the causal effect of electrification on *average* firm outcomes across desas. Therefore, the results in table 4.3 do not include the effect of selection, while the results in table 4.1 do. Given that the estimated effect of electrification on the sales revenue of incumbents is around a tenth of the estimated effect including selection, this indicates that the selection effects of electrification are substantial and drive most of the increase in manufacturing output at a local level.

Looking at columns (2) and (3) in panel B, the effect of electrification on capital and wages is positive and smaller in magnitude than the effects estimated without the fixed effects, although the results are statistically insignificant. This is not too surprising as capital and labor could face some adjustment costs that hinder the firm from adjust its production process in the short and medium run. The coefficient in column (4) on the number of workers is negative, but not significant. One interpretation of the negative sign, although not significant, could be that these switching incumbents are becoming less labor intensive. These results are in line with Allcott, Collard-Wexler, and O'Connell (2016).

Finally columns (5) and (6) in panel B show that electrification causes the switching incumbents to consume more electricity, as expected. Together with the results from columns (1) to (4), all these results point to a strong selection mechanism that is driving the increase in local industrial outcomes.

Table 4.3: Impact of access on the sales and inputs at the firm level.

Sample: Firm-Level						
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable (Log)	Sales	Capital	Wage Bill	Nb Workers	Energy Bill	Electricity (kWh)
Panel A: OLS						
Access	0.0263	-0.0458**	-0.0163	-0.0128	0.0361	0.0697**
	(0.0193)	(0.0211)	(0.0197)	(0.00987)	(0.0228)	(0.0323)
Panel B: IV						
Access	0.186**	0.00850	0.0781	-0.0317	0.407***	0.287**
	(0.0945)	(0.110)	(0.0845)	(0.0519)	(0.124)	(0.130)
First Stage F	21.95	21.95	21.94	21.95	21.80	22.17
Observations	133,349	133,349	133,334	133,349	131,495	113,655
Firm FE	✓	✓	✓	✓	✓	✓
IndustryxYear FE	✓	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, and IV regressions of equation 4.1. The reduced-form results are omitted for ease of exposition. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

4.3 Electrification and Survival

I now examine whether electrification affects turnover in the economy. In other words, does the expanded access to electricity increase firm selection the desa? I start by investigating the effect of electrification on the probability of exit. I estimate a linear probability model where I regress an exit dummy on access, instrumented with distance to the hypothetical and controlling for desa-level and firm-level characteristics as above. Before presenting the results, a discussion about how exit is defined is necessary. I define exit in period t as a dummy variable equal to one if the firm drops out of the census in period $t+1$. Because this is a census of firms with 20 or more employees, it could be that the firm did not actually exit the market, but instead shrank below the size threshold. For that reason, I restrict the definition of exiting firms to those who are not in the survey in year $t+1$ and have at least 25 employees, which is the 25^{th} percentile of size in the data.

Table 4.4 shows results from the OLS, IV and reduced-form regressions. Column (1) panel A presents the OLS estimate of the effect of access on the average exit probability. The coefficient is positive and significant indicating that the probability of exit and being in an electrified desa are positively correlated. The corresponding IV regression is in column (1) panel B, and as before, the magnitude of the OLS estimate is smaller than the IV estimate. The coefficient shows that the causal effect of electrification on selection is an increase of around 5% in the probability of exit. Column (1) panel C show the reduced-form regressions of exit on the distance to the hypothetical least cost grid, showing that the closer the firm is to the least cost network, the more likely it is to exit. This suggests that survival in the industry is less likely in electrified desas.

Table 4.4 column (2) shows the effect of electrification on the age distribution of firms. It presents the OLS, IV and reduced form regressions of a dummy variable $young_{it}$ equal to 1 if a firm is below the median age. Results show that firms in electrified desas are on average younger. This finding is consistent with electrification shifting the age distribution of firms towards younger firms by (i) increasing entry, therefore having more younger firms, and (ii) increasing exit, therefore shortening the average firm age in the desa.

Table 4.4: Electrification, exit, and the age distribution.

Sample: Firm-Level		
	(1)	(2)
Dependent Variable	Exit	Young
Panel A: OLS		
Access_{vpt}	0.0077***	0.0371***
	(0.002)	(0.0148)
Panel B: IV		
Access_{vpt}	0.049**	0.242**
	(0.016)	(0.099)
First Stage F	41.55	41.55
Panel C: Reduced Form IV		
Z (KM)	-0.000144***	-0.000718**
	(3.82e-05)	(0.000281)
Observations	141,615	141,615
IndustryxYear FE	✓	✓
Province FE	✓	✓
Geo Controls	✓	✓
Firm Controls	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, IV and reduced-form regressions for a young dummy access to electricity defined at the desa level. Young is a dummy equal to 1 if the firm's age is below the median age. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Interesting conclusions can be drawn from the results in table 4.4. The decrease in the survival probability and the shift in the age distribution of firms are sign of tougher selection in the market. Electrification makes it less likely for a firm to survive in the market, consistent with the results in chapter 3 where electrification increases exit rates. Also consistent with the results in chapter 3 on higher entry rates, electrification shifts the age distribution of firms towards younger firm. These results at the firm level confirm that electrification increases firm turnover, as shown at the desa-level in chapter 3. This is a sign of healthy churning where markets become better at weeding out the inefficient firms with electrification. The implications on average industry productivity will be examined in the next chapter.

4.4 Spillovers

As with the desa-level analysis, a threat to identification in the empirical analysis in this chapter is a violation of the SUTVA assumption, or spillovers. In reality, firms in certain desa can sell their output in different desas. Results from the firm-level analysis show that connected firms sell more. Another interesting question is therefore whether these firms are stealing business from unconnected firms. In other words, is there any creation of new output in response to electrification, or is production moving from non-electrified desas to electrified desas?

Given that the data I use in this chapter is more detailed and I can observe the industries in which firms operate, it is interesting to test for spillovers in response to electrification within industries. If there are any spillovers or general equilibrium effects, they are strongest and more detectable within industry. This is because results point at a competitive effect of electrification, where electrification intensifies competition and leads to tougher selection, which theoretically happens within an industry.

To check if spillovers or business stealing effects are present in my context, I run three tests. The extent to which these spillovers exist and might differ by industry depends on various factors. These factors include how easy it is to transport the products and how spread out geographically the demand is.

First, it depends on the type of goods produced and their tradability. For example, we except these spillovers to minimal in the context of non-tradable goods. To test this, I estimate the effect of access on firm sales in the non-tradable sectors[2]. I consider certain products to be non-tradables because of their heavy weight which involves significantly large transportation costs. Table 4.5 presents the IV results for this exercise. I find a coefficient of 2.3, which is very close to the estimate found using the whole sample in table 4.1 panel B column (1). This shows that in a setting where business stealing effects or spillovers should be minimal because of large transportation costs, electrification still increases average firm sales. This indicates that there is some new economic activity being generated from electrification.

[2]These are two three-digit industries (263 and 264 ISIC Rev3). They include the following categories: refractory bricks, clay products, clay bricks, clay tiles, structural clay, cement, lime plaster, gyps

Table 4.5: Effect of electrification on sales of nontradables

Dependent Variable	(1) Sales
Access	2.277**
	(0.907)
First Stage F	12.80
Observations	11,462
IndustryxYear FE	✓
Province FE	✓
Geo Controls	✓
Firm Controls	✓
*** p<0.01, ** p<0.05, * p<0.1	

Notes: Results from OLS and IV regressions. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Second, I test for general equilibrium effects by regressing firm sales on the number of switching neighboring districts:

$$y_{ivpst} = \alpha + \beta M_{vpst} + \eta \mathbf{X}_{ivpst} + \theta Z_{vpt} + \eta \mathbf{V}_{vpst} + \gamma_p + \delta_{st} + \epsilon_{ivpst} \tag{4.2}$$

The idea is that if spacial spillovers exist, then the number of switching districts around the firm desa should affect firm revenue negatively. Here the assumption is that trade costs are infinite for further away districts. It is a strong assumption but it is supposed to capture that trade costs increase with distance. If there are spillovers, they will be strongest between neighboring districts. Because the number of switching neighbors is endogenous, I instrument for it with the average distance to hypothetical least cost grid in the district, conditional on the firm's own distance to the least cost network[3].

Table 4.6 shows the corresponding OLS and IV regressions. Column (2) presents the IV regression. The coefficient on number of switching firms is negative statistically insignificant. This rejects the presence of spacial spillovers. Even if the coefficient

[3]These two distances are not collinear given the variation in the shape of the hypothetical least cost grid.

were to be significant, the implied effect is very small[4] (0.3) relative to the effect of access I find in table 4.1 column (2) Panel B and cannot explain more than 13% of the difference in average sales between electrified and non-electrified firms.

Table 4.6: Testing For Spillovers

	(1)	(2)
	OLS	IV
Dependent Variable	Sales	Sales
(Log)		
Nb switching neighbors	0.149***	-0.108
	(0.0234)	(0.153)
Z (KM)	-0.00776***	-0.00668***
	(0.00129)	(0.00156)
Observations	141,615	141,420
First Stage F		45.02
IndustryxYear FE	✓	✓
Province FE	✓	✓
Geo Controls	✓	✓
Firm Controls	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV regressions. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Finally, I look for spillovers within narrowly defined industries across the whole island. I run the same test as in equation 4.2 at the industry level. The number of switchers in a certain industry is counted in the whole island, as opposed to the neighboring geographic locations as in the previous test, but within a 5-digit industry. Results from IV regressions are presented in table 4.7. The coefficient in column (1) is the estimated effect of an increase in the number of switching competitors on the sales of non-switchers for all industries pooled together is statistically zero. I estimate this relationship again by industry. Across all industry, the estimated coefficients are negative but very small. In column (2), I run the same test for non-tradables and I find a precisely estimated effect of zero. This is not surprising as spillovers are not expected in this particular type of industries. For textiles in col-

[4]This is equal to the estimated coefficient -0.108 times the average number of switching neighbors, which is around 3

umn (4), I also find a precisely estimated zero.

Table 4.7: Testing For Spillovers within a 5-digit industry

Dependent Variable	Log Sales			
	(1)	(2)	(3)	(4)
Industry	All	non-tradables	Food & Bev.	Textiles
Number of Switching Competitors	0.00181 (0.00542)	-0.000866 (0.00346)	-0.0234 (0.0253)	-0.000858 (0.0509)
First Stage F	86.47	124.5	91.83	50.64
Observations	113,115	10,861	24,329	15,317
Mean RHS	10.1	20.5	6.2	5.8
	(5)	(6)	(7)	(8)
	Apparel & Footwear	Furniture	Rubber & plastic	All
Number of Switching Competitors	-0.0164 (0.0102)	-0.0305 (0.0443)	-0.0270 (0.0412)	
Access				2.057*** (0.497)
Observations	16,058	10,836	6,887	113,115
First Stage F	340.5	45.80	477.2	35.11
Mean RHS	11.8	11.9	2.5	

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

Notes: Results from IV regressions. The dependent variable is log sales. The first column shows the regression of the whole sample of firms. The RHS variable is the number of switching competitors. A switching competitor is a firm in the same 5-digit industry that switches from being without access to having access to the grid. Columns (2) - (7) shows the same regression for each of the top 6 largest industries separately. Column (8) presents the effect of access on sales of all firms in the 6 largest industries. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

For the other industries in columns (3) Food and Beverages, (5) Apparel, (6) Furniture and (7) Rubber, the coefficients are statistically indistinguishable from zero, and the magnitudes are small. Using the mean number of switching competitors in the industry in the row "Mean RHS" and the estimated coefficients in columns (1) and (8), spillovers can explain only around 10% of the effect of access. Doing the same back of the envelope calculation for the highest among these estimated coefficients for furniture in column (6), spillovers can explain at most around 20% of the estimated effect.

Based on the results in this section, I conclude that spillovers are not a major concern in this setting. The evidence for business stealing effects is fairly limited, and results from chapter 3 show no evidence for spillovers at the extensive margin (firm entry and relocation). A potential reason why spillovers are limited is as follows. Given the large number of desas (23,000 per year), and the large number of firms (16,104 per year on average), such spillover effects could be negligible or undetectable because each unit is too small to affect its competitors if Java is considered as one single market. The fact that despite the absence of spillovers there are positive and significant large effects of electrification strongly suggests that electrification does indeed create new industrial activity.

4.5 Conclusion

In this chapter, I find that electrification not only increases industrial activity at a local level by increasing the number of firms, but it also affects the performance of firms in the market. Firms in electrified desas are larger, both in terms of how much they sell but also in how much inputs they consume. Consistent with higher turnover at the desa-level, firms in electrified desas face a higher probability of exit, and are more likely to be younger. As for business stealing effects, they appear to be minimal in the particular case of Indonesian manufacturing. While violations of SUTVA cannot be completely ruled out, evidence in this chapter shows that if there are any issues in identification related to spillovers, these issues are not severe and are unlikely to invalidate the analysis in this chapter.

Appendix

4.A Additional Results

Table 4.8: Impact of connection on the sales and inputs at the firm level.

Sample: Firm-Level						
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable (Log)	Sales	Capital	Wage Bill	Nb Workers	Energy Bill	Electricity (kWh)
Panel A: OLS						
Connected_{it}	0.610***	0.686***	0.393***	0.186***	0.675***	0.0394
	(0.0532)	(0.0529)	(0.0354)	(0.0249)	(0.0796)	(0.0676)
Panel B: IV						
Connected_{it}	3.531***	4.805***	2.512***	1.644***	6.050***	29.22**
	(0.716)	(0.912)	(0.589)	(0.396)	(1.165)	(12.29)
First Stage F	33.82	33.82	33.86	33.82	27.08	8.044
Panel C: Reduced Form IV						
Z (KM)	-0.00665***	-0.00933***	-0.00505***	-0.00336***	-0.0114***	-0.0110***
	(0.00134)	(0.00139)	(0.00102)	(0.000661)	(0.00180)	(0.00204)
Observations	141,615	141,615	141,615	141,615	139,481	120,453
IndustryxYear FE	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, IV, and reduced form regressions of equation (2.2). $Connected_{it}$ is a dummy equal to one if the firm is observed consuming a positive quantity of grid electricity. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Chapter 5

Electrification, Productivity, and Reallocation

5.1 Introduction

So far, I have presented evidence on how electrification affects industrial development. An interesting question arising from this evidence is whether electrification has any implications on industry productivity. In particular, in this chapter I am interested in finding out if there are any productivity enhancing effects of electrification at the industry level.

In order to do this, I will need to estimate productivity. The literature on structural estimation of production functions has provided us with a set of tools to recover consistent productivity estimates. I will use different methods from the frontier of this literature and I will compare the results using the respective productivity estimates. When estimating productivity, I will specifically pay attention to how access to electricity might threaten these methods and I will present solutions. In particular, one issue that needs to be dealt with when estimating the production function and backing out productivity estimates for the purpose of finding the causal effect of electrification on average productivity relates to survival. As I have shown in chapters 3 and 4, electrification affects firm survival. Hence, the fact that electrification affects selection in the market needs to be addressed when estimating productivity.

Before turning the empirical model, I first present a theoretical model to motivate the empirical analysis and to clarify the different channels through which electrification can affect aggregate industrial productivity.

5.2 Theoretical Motivation

5.2.1 How can Access to Electricity Affect Productivity?

The purpose of this section is to lay out conceptually the different ways electrification can affect the firm and industry outcomes, keeping in mind the Indonesian context. During the years that the study covers (1990-2000), almost all Indonesian manufacturing firms were using electricity in their production process, but if they were not connected to the grid then they had to rely on self-generation. Since electricity is an input of production, gaining access to the grid will affect the price of the electricity input that the firm faces. Self-generation affects the firm's cost structure in at least two ways. First, in order to start production, the firm needs to incur the cost of buying a generator, which can be hefty, especially for industrial use. This means that electrification can affect the entry costs of a firm. Second, access to the grid will allow to the firm to buy electricity at a cheaper price than the self-generation price, therefore affecting the marginal cost of the firm. To fix ideas, I do not think of access to electricity as directly affecting within firm productivity (productivity is not a function of access), however, electrification can affect selection in the market which in turn affects the average productivity of surviving firms. In the next section, I will present an industry model to understand how each of these channels will affect selection in the market and the implications on average industry productivity.

I present below a model of a monopolistically competitive industry à la Syverson (2007) and Melitz and Ottaviano (2008) to illustrate the effects of the grid expansion on the manufacturing sector. The goal is to analyze selection, allowing for competition effects. As the grid reaches more areas, the entry decision of firms in these areas will be affected through a reduction in the sunk cost of entry. In addition, as more firms in the market are getting connected, and thus becoming more efficient, this will affect the survival of incumbents (and expected value of entry) as a higher proportion of more efficient firms in the market means more intense competition.

5.2.2 Demand

Consider an industry with a continuum of firms of measure N, each indexed by i. Firm i produces a differentiated variety in the market. Consumers have utility U defined over these differentiated varieties indexed by i in set I and a Hicksian composite commodity:

$$U = H + \int_{i\in I} \alpha q_i di - \frac{1}{2}\eta\Big(\int_{i\in I} q_i di\Big)^2 - \frac{1}{2}\gamma \int_{i\in I} {q_i}^2 di \tag{5.1}$$

where H is the consumption of the Hicksian composite good and q_i is the consumption of variety i. The demand parameter $\eta \in (0,1)$ represents the degree of substitutability between different varieties. Utility maximization implies the following demand function:

$$q_i = \frac{\alpha}{\eta N + \gamma} + \frac{\eta N}{\gamma(\eta N + \gamma)}\bar{p} - \frac{1}{\gamma} p_i \tag{5.2}$$

where $\bar{p} \equiv \frac{1}{N}\int_{i\in I} p_i$ is the average price in the market conditional on survival. Define p^{max} as the highest price consumers are willing to pay which can be calculated from setting demand in equation (5.2) to zero:

$$p^{max} = \frac{\gamma\alpha}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} \tag{5.3}$$

The residual demand for product i from (5.2) can therefore be written as:

$$q_i = \frac{1}{\gamma}(p^{max} - p_i) \tag{5.4}$$

5.2.3 Production

On the production side, consider a single input technology[1] where firm i produces according the the following production function:

$$q_i = \phi_i x_i \tag{5.5}$$

where ϕ_i is the firm's physical productivity and x_i is the input of production which is supplied inelastically at a constant[2] price w. Therefore, firm i's marginal cost is $c_i = \frac{w}{\phi_i}$. Combined with the demand form, the profit maximizing price is:

$$p(c_i) = \frac{1}{2}(p^{max} + c_i) \tag{5.6}$$

[1]The assumption of a single input production process is without loss of generality when considering a multiple input production function with constant returns to scale.

[2]This simple representation is meant to capture that although firms are heterogeneous in their productivity they face the same price of electricity which is set by the state, either directly (price per kWh or price of fuel). This is true in the case of Indonesia where the energy sector is heavily regulated and the price is the same everywhere in the country.

The equilibrium profit is:

$$\pi(c_i) = \frac{1}{4\gamma}(p^{max} - c_i)^2 \tag{5.7}$$

Firm i will stay in the market as long as $\pi(w, \phi_i) \geq 0$. This gives the cut-off level of marginal cost c^* such that the firm will not want to stay in the market if its marginal cost exceeds it:

$$c^* = p^{max} = \frac{\gamma\alpha}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} \tag{5.8}$$

Firm price, mark-up and quantity can therefore be written as:

$$p(c_i) = \frac{1}{2}(c^* + c_i) \tag{5.9}$$

$$\mu(c_i) = \frac{1}{2}(c^* - c_i) \tag{5.10}$$

$$q(c_i) = \frac{1}{2\gamma}(c^* - c_i) \tag{5.11}$$

Firm price is increasing in its own marginal cost, but more efficient firms charge relatively higher markups and produced relatively more. The more efficient the marginal firm is (lower c^*), the tougher competition is, reducing firm prices, markups and quantity demanded, conditional of the firm's own marginal cost. The cutoff c^* then implies implies a cutoff level for firm productivity:

$$\phi^* = \frac{w}{c^*} \tag{5.12}$$

Firms with productivity $\phi_i < \phi^*$ will not be profitable and will exit the market. Therefore, $p^{max} = \frac{w}{\phi^*}$.

5.2.4 Long Run Equilibrium

In the long run, a large number of ex-ante identical potential firms decide whether to enter the market. Before observing their productivity, potential entrants have to pay a sunk cost of entry s. They then receive a productivity draw from a distribution $G(\phi)$ with support $[\underline{\phi}, \infty]$. In equilibrium, the expected value of entry should be equal to zero for positive entry to occur:

$$V^e = \frac{w^2}{4\gamma}\int_{\phi^*}^{\infty}\left(\frac{1}{\phi^*} - \frac{1}{\phi}\right)^2 dG(\phi) - s = 0 \tag{5.13}$$

Equation (5.13) pins down ϕ^* which summarizes the equilibrium. The equilibrium mass of firms N is determined using equations (5.6) and (5.8).

5.2.5 Predictions

The goal of this exercise is to see how the equilibrium cut-off changes with access to electricity. This can be studied through comparative statics with respect to two parameters. The first is the input price w. Access to the grid reduces the per-unit price of electricity. The second is the sunk cost of entry s. Entry to a location where the grid hasn't arrived is potentially more expensive as the firm will need to purchase its own generator. Starting with comparative statics with respect to w, and using the implicit function theorem:

$$\frac{d\phi^*}{dw} = -\frac{\partial V^e / \partial w}{\partial V^e / \partial \phi^*} > 0 \tag{5.14}$$

since $\partial V^e / \partial \phi^* < 0$ and $\partial V^e / \partial w > 0$. Therefore, a decrease in w will lead to a lower productivity cut-off. Intuitively, as the input price is lower, a firm that wasn't able to survive before will be able to do so now. As for the sunk cost of entry, the cutoff ϕ^* is decreasing in s since the derivative of the value function with respect to s is -1:

$$\frac{d\phi^*}{ds} = -\frac{\partial V^e / \partial s}{\partial V^e / \partial \phi^*} < 0 \tag{5.15}$$

This says that if access to electricity reduces the sunk cost of entry, then this will increase the average productivity in the industry. The intuition is as follows. If access to electricity lowers barriers to entry, more firms will enter the market, across the whole productivity distribution. This intensifies competitive pressure and makes it more difficult for relatively unproductive firms to survive in equilibrium.

In order to understand how average industrial outcomes could be affected by electrification, it is useful to focus the analysis on changes in the marginal cost cutoff c^*. This is because although the effect of access on ϕ^* is interesting, what ultimately determines the equilibrium outcomes is a combination of input prices and firm productivity, i.e. the marginal cost of the firm. Revisiting the comparative statics with respect to input price w and sunk cost of entry s gives the following predictions. The effect of a decrease in w on c^* is ambiguous. Although ϕ^* increases with a decrease in w, this doesn't necessarily mean that the marginal cost of the marginal firm c^* is lower. The overall effect depends on the relative effects of the decrease in w and increase in ϕ^*. As for the sunk cost of entry, conditional on w, a decrease in s unambiguously leads to a decrease c^*.

Define the average marginal cost of surviving firms $\bar{c} = \frac{1}{1-G(\phi^*)}\int_{\phi^*}^{\infty}\frac{w}{\phi}dG(\phi)$. Given a distribution of productivity $G(.)$, the averages of firm outcomes in equations (5.9)-(5.11) conditional of survival are:

$$\bar{p} = \frac{1}{2}(c^* + \bar{c}) \tag{5.16}$$

$$\bar{\mu} = \frac{1}{2}(c^* - \bar{c}) \tag{5.17}$$

$$\bar{q} = \frac{1}{2\gamma}(c^* - \bar{c}) \tag{5.18}$$

where $\bar{z} = \frac{1}{1-G(\phi^*)}\int_{\phi^*}^{\infty} z(\phi)dG(\phi)$. Intuitively, $\bar{c}$ is increasing in c^*. If the marginal firm is more efficient (lower c^*), then the average firm efficiency in the industry is higher (lower $\bar{c}$). Equation (5.16) predicts that the average observed prices conditional of firm survival is lower when c^* is lower. Equations (5.17) and (5.18) however give an ambiguous prediction on a change in c^* on average markups and quantities. On the one hand, a lower c^* means tougher competition in the market, reducing firm markups and quantities produced. However, tougher selection also means that the set of surviving firms are more efficient (lower $\bar{c}$), and as seen from equations (5.10) and (5.11), more efficient firms charge relatively higher markups and produced more. Which effects dominates depends on the distribution of productivity $G(.)$ and its support.

Recall that in equilibrium, the zero profit condition states that the profit of the marginal firm should be equal to zero. This condition requires that $c^* = p^{max}$:

$$c^* = \bar{p} + \frac{\gamma(\alpha - \bar{p})}{\eta N + \gamma} \tag{5.19}$$

The equilibrium mass of active firms as a function of c^* is therefore:

$$N = \frac{2\gamma(\alpha - c^*)}{\eta(c^* - \bar{c})} \tag{5.20}$$

These equations state that tougher competition (lower c^*) is associated with a higher mass of active firms N and a lower average price[3] $\bar{p}$. To see this[4], suppose N increases, and that surviving firms don't change their prices following entry, keeping $\bar{p}$ constant. From equation (5.19), c^* will decrease. From equation (5.16), $\bar{p}$ will

[3]An implicit assumption here is that $\alpha > c^*$ which implies that α is greater than $\bar{p}$ and $\bar{c}$.
[4]The intuition is the same as in Combes, Duranton, Gobillon, Puga, and Roux (2012).

decrease as result, which further decreases c^*. In addition, the model predicts that firm exit rates unambiguously increase when the marginal cost cutoff c^* is lower. The probability of survival, which is equal to $\tilde{G}(c^*) = 1 - G(\frac{w}{\phi^*})$, is decreasing in c^*. Intuitively, tougher competition is associated with tougher selection where conditional on its own efficiency, a firm's probability of survival is lower.

The relationship between access to electricity and firm-level and industry-level outcomes can be interpreted through the lens of the model. The averages of firm outcomes in (5.16)-(5.18) correspond to the respective observed firm outcomes in the data. If access to the grid reduces fixed cost of entry, the model predicts that access will lead to tougher selection in the market induced by entry of a larger number of firms. In addition, the model predicts that higher exit rates are associated with tougher selection and a higher efficiency cutoff.

Finally, equations (5.14) and (5.15) state that average physical productivity ϕ increases if barriers to entry are lower, but decreases in response to an increase in the input price. This sharp prediction is informative regarding the channels through which access to electricity is affecting the manufacturing sector.

The insights from the model will therefore guide the empirical analysis in the subsequent sections and help interpret the results. Table 5.1 summarizes the predictions of the model, split by the different channels :

Table 5.1: Model Predictions.

Effect of Electrification			
#	Outcome	barriers to entry $s \downarrow$	input price $p_x \downarrow$
1	Effect on competition	$\phi^* \uparrow$	$\phi^* \downarrow$
2	Average revenue productivity $TFPR = \phi * p$	? $\bar{\phi} \uparrow \& \bar{p} \downarrow$	? $\bar{\phi} \downarrow \& \bar{p}?$
3	Average marginal cost $\bar{c}$	$\downarrow$	?
4	Probability of exit	$\uparrow$	$\downarrow$
5	Age distribution	young	old
6	Firm turnover	$\uparrow$	$\downarrow$

The simplicity of the model, which is useful to guide the empirical analysis, means that the model abstracts from many features that are potentially important. I discuss the limitations of the model and how they affect the results in appendix 5.A.

5.3 Measuring Productivity

5.3.1 Methodology and Literature Review

Consider the following revenue-based Cobb-Douglas production function in logs defined at the industry level:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_e e_{it} + \phi_{it} + \epsilon_{it} \tag{5.21}$$

where y_{it} is output, k_{it} if capital, l_{it} is the wage bill and e_{it} is total spending on electricity. ϕ_{it} is firm i's productivity in year t. It subsumes the constant term. Finally, ϵ_{it} is an i.i.d. random shock. This equation is the basis of the empirical framework and will be estimated separately for each industry.

The classic endogeneity challenge in estimating equation (5.21) arises from the fact that ϕ_{it} is observable by the firm when it is choosing its fully flexible inputs such as labor and electricity but not to the econometrician. This leads to biased OLS estimates of the vector of the production function coefficients $\boldsymbol{\beta}$. This is the *simultaneity bias*. Another concern here is the survival bias. Only surviving firms are observed in the data. Empirically, it is observed that larger firms can survive larger shocks. This creates a positive correlation between firm size, measured by capital, and the survival probability conditional on the same shock, which will lead to a bias in the OLS estimates of the coefficient on capital. This is the *survival bias*.

Economists have been trying to solve the endogeneity problem arising from simultaneity bias for decades. Griliches and Mairesse (1995) and Ackerberg, Caves, and Frazer (2015) present a great summary of the early literature as well as the frontier which I will briefly reiterate here. I will further discuss these biases in relation to access to electricity.

Among the earliest attempts to account for these biases are the fixed effects approach, which requires the strong assumption of time-invariant productivity ϕ_i, and the instrumental variable approach (Mundlak (1961), Hoch (1962)) using input prices as instruments under the assumption of perfectly competitive input markets. The instrumental variable approach did not work very well in practice first because typically input prices do not vary sufficiently within an industry across firms and if they do it is because of things like differentiated input qualities, making these instruments invalid.

5.3.2 Simultaneity Bias

The simultaneity bias arises because firms choose their inputs to maximize their profits, potentially after observing at least some information about their own productivity ϕ_{it} that is unobservable to the econometrician.

The frontier literature on dealing with the simultaneity bias was initiated by Olley and Pakes (1996). Their insight was that an observable choice variable of the firm, investment in their case, can be used as a proxy to infer unobserved firm productivity. Using the first order condition of the firm's choice of investment, conditional on capital, the investment function can be inverted to recover ϕ_{it}. Levinsohn and Petrin (2003) propose using materials spending or electricity spending as a the proxy to avoid problems arising from the lumpiness of investment. They also argue that these more flexible inputs are more likely to satisfy the monotonicity assumption needed when inverting the proxy to recover productivity. Both these papers suggest a two-step estimation algorithm, the validity of which crucially relies on timing assumptions. Specifically, at time t when the firm observes its productivity ϕ_{it}, capital k_{it} is predetermined, hence it is a state variable, and the other inputs (labor l_{it} and e_{it} electricity) are fully flexible and are chosen after the firm observes ϕ_{it}.

The first step proceeds as follows. Using the F.O.C. of the proxy variable ι_{it}, the optimal choice of ι_{it} can be written as a function of productivity and state variables. For simplicity assume for now that the only state variable is capital:

$$\iota_{it} = h(\phi_{it}, k_{it})$$

A necessary condition for ι_{it} to be a valid proxy is for its demand function $h(.)$ to be strictly increasing in ϕ_{it} conditional on k_{it}. This is the monotonicity assumption which states that more productive firms will invest more or consume more inputs. Another assumption required for the proxy to be invertible is a scalar unobervable, i.e. productivity is the only variable in the proxy function that is unobservable to the econometrician. Using a first order condition of the firm's optimization problem, the proxy can be inverted to infer productivity. Now ϕ_{it} can be recovered by inverting the $h(.)$ function:

$$\phi_{it} = h^{-1}(\iota_{it}, k_{it}) \tag{5.22}$$

Substituting back in equation (5.21):

$$y_{it} = \beta_l l_{it} + \beta_e e_{it} + \kappa(\iota_{it}, k_{it}) + \epsilon_{it} \tag{5.23}$$

where

$$\kappa(\iota_{it}, k_{it}) = \beta_k k_{it} + \phi_{it} = \beta_k k_{it} + h^{-1}(\iota_{it}, k_{it}) \tag{5.24}$$

Given that $h^{-1}(\iota_{it}, k_{it})$ includes capital, the coefficient on capital β_k cannot be separately identified from the coefficient on capital in $h^{-1}(.)$, hence the $\kappa(.)$.

Equation (5.23) is the first stage estimation equation where the $\kappa(.)$ function is estimated nonparametrically and is approximated by a second-order polynomial in the proxy variable and capital. This is the control function approach. Only β_l and β_e are estimated in this step of the algorithm. In the case where electricity spending is the proxy, or $\iota_{it} = e_{it}$, then only β_l is estimated in this step. In what follows, let v_{it} represent the free variable inputs that are not used as a proxy and are estimated in the first stage. An estimate $\hat{\kappa}(.)$ is also obtained from this step.

The second stage of the estimation algorithm aims at estimating the remaining production function coefficients (β_k in addition to β_e in case electricity spending is used as a proxy) and to correct for the survival bias. An important assumption here is that ϕ_{it} follows a first-order Markov process where productivity today only depends on productivity in the previous period and a random shock:

$$\phi_{it} = g(\phi_{it-1}) + \eta_{it} \tag{5.25}$$

where $g(.)$ is an unknown function and η_{it} is an i.i.d. shock uncorrelated with k_{it-1}. Replacing ϕ_{it} by its law of motion in the production function, the second-stage estimation equation is as follows:

$$y_{it} - \Sigma_v \hat{\beta}_v v_{it} = \beta_k k_{it} + g(\hat{\kappa}(\iota_{it-1}, k_{it-1}) - \beta_k k_{it-1}) + \eta_{it} + \epsilon_{it} \tag{5.26}$$

where ϕ_{it-1} is rewritten in terms of the κ function, capital and the proxy using equation (5.24):

$$\phi_{it-1} = \kappa(\iota_{it-1}, k_{it-1}) - \beta_k k_{it-1}$$

Identification of β_k in this stage relies on the timing assumption that capital in period t, k_{it}, is predetermined and therefore is uncorrelated with $\phi_{it} = g(\phi_{it-1}) + \eta_{it}$. The $g(.)$ function is estimated nonparametrically and is approximated by a polynomial in $\hat{\kappa}_{it-1}$ and k_{it-1}. If e_{it} is used as the proxy, then the second-stage estimation equation is:

$$y_{it} - \Sigma_v \hat{\beta_v} v_{it} = \beta_k k_{it} + \beta_e e_{it} + g(\hat{\kappa}(e_{it-1}, k_{it-1}) - \beta_k k_{it-1} - \beta_e e_{it-1}) + \eta_{it} + \epsilon_{it} \quad (5.27)$$

Ackerberg, Caves, and Frazer (2015) shed light on a potential functional dependence problem in the first stage of the estimation procedure of the OP/LP estimation algorithms. Specifically, labor is likely to be a function of the same variables as the proxy, i.e. ϕ_{it} and k_{it}. Hence, when ϕ_{it} is controlled for nonparametrically, there is little or no variation left to identify the coefficient on labor β_l in the first stage. Ackerberg, Caves, and Frazer (2015) propose the following refinement. Instead of a two-step estimation procedure, they suggest inverting the proxy demand function conditional on labor. No coefficient is estimated in the first step since the control function for productivity also includes l_{it}. Instead, the first stage produces an estimate of predicted output as a function of production function coefficients as follows[5]:

$$y_{it} = \kappa(k_{it}, l_{it}, e_{it}, \iota_{it}) + \epsilon_{it} \quad (5.28)$$

where $\kappa(k_{it}, l_{it}, \iota_{it}) = \beta_l l_{it} + \beta_e e_{it} + \beta_k k_{it} + h(\iota_{it}, k_{it})$.
The rest of the estimation proceeds similarly as before but with the following moment conditions:

$$E\left\{\eta_{it}(\beta_l, \beta_e, \beta_k)\begin{pmatrix} l_{it-1} \\ e_{it-1} \\ k_{it} \end{pmatrix}\right\} = 0$$

These moment conditions are based on the law of motion of ϕ_{it} as in equation (5.25):

$$\eta_{it}(\beta_l, \beta_e, \beta_k) = \phi_{it}(\beta_l, \beta_e, \beta_k) - \phi_{it-1}(\beta_l, \beta_e, \beta_k)$$

and

$$\phi_{it}(\beta_l, \beta_e, \beta_k) = \hat{\kappa}_{it} - \beta_l l_{it} - \beta_e e_{it} - \beta_k k_{it}$$

5.3.3 Survival Bias

Denote the exit decision of firm i in period t by χ_{it}. Every period, after observing ϕ_{it}, firm i decides whether to stay in the market and produce ($\chi_{it} = 1$) or to exit ($\chi_{it} = 0$). In equilibrium, there is a productivity cutoff ϕ_t^* below which firms decide to exit. Empirically, firms with larger capital k_{it} are more likely to survive the same

[5] In case $\iota_{it} = e_{it}$, then $y_{it} = \kappa(k_{it}, l_{it}, e_{it}) + \epsilon_{it}$

shock than firms with less capital. This is the classic survival bias and the survival cutoff potentially depends on capital.

Results from chapters 3 and 4 indicate that electrification affects survival in the market, creating tougher selection. This implies that to obtain consistent estimates, any selection correction should account for the availability of the grid where the firm is located, in addition to the classic survival bias. To address the selection bias, I allow the survival threshold ϕ_t^* to depend on access to the grid in addition to capital.

$$\chi_{it} = \begin{cases} 1, & \text{iff } \phi_{it} > \phi_t^*(k_{it}, access_{it}) \\ 0, & \text{otherwise} \end{cases}$$

This creates a selection bias in the set of surviving firms that are observed in the data and used to estimate the production function coefficients. Olley and Pakes (1996) propose a sample selection correction by estimating the probability of survival as follows:

$$\begin{aligned} Pr(\chi_{it} = 1|\mathscr{I}_{it-1}) &= Pr(\phi_{it} \geq \phi_t^*(k_{it}, access_{it})|\mathscr{I}_{it-1}) \\ &= p_t(\phi_t^*(k_{it}, access_{it}), \phi_{it-1}) \\ &= p_t(\iota_{it-1}, k_{it-1}, i_{it-1}, access_{it}) \end{aligned}$$

where the last equality uses the fact that k_{it} is fully determined by investment[6] and capital in period $t-1$ and that ϕ_{it-1} is a function of ι_{it-1} and k_{it-1} using equation (5.22). The estimation equation conditional on survival is:

$$E[y_{it} - \Sigma_v \hat{\beta}_v v_{it}|\mathscr{I}_{it-1}, \chi_{it} = 1] = \beta_k k_{it} + [\phi_{it}|\mathscr{I}_{it-1}, \chi_{it} = 1] \quad (5.29)$$

$$= \beta_k k_{it} + g(\phi_{it-1}, \phi_t^*(k_{it}, access_{it})) \quad (5.30)$$

Given that k_{it} appears twice on the right hand side of the above equation, a separate estimate for $\phi_t^*(k_{it}, access_{it})$ is required in order to identify β_k. To get this estimate, the probability of survival, which can be estimated using a probit model where the left hand side is a survival dummy with capital and investment on the right hand side. Olley and Pakes (1996) show that the probability of survival can be inverted to recover $\phi_t^*(k_{it})$, conditional on ϕ_{it-1}, which is estimated in the first step of the

[6] In the case where investment is the proxy, this equation becomes: $Pr(\chi_{it} = 1|\mathscr{I}_{it-1}) = Pr(\phi_{it} \geq \phi_t^*(k_{it}, access_{it})|\mathscr{I}_{it-1}) = p_t(k_{it-1}, i_{it-1}, access_{it})$.

algorithm.

Note that including access in the estimation of survival requires the assumption that access is contemporaneously uncorrelated with the innovation term in the law of motion of ϕ. In particular, I assume that access in period t is pre-determined where the planner decides to extend the grid a particular location in period t after observing the productivity of firms in that location ϕ_{it-1}, conditional on access in period $t-1$.

Olley and Pakes (1996) find that when using an unbalanced panel of firms, their correction for sample selection does not affect their estimates. This shows that when using an unbalanced panel the survival bias is negligible. As a result, the subsequent literature including Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015) has ignored this bias and focused mainly on the simultaneity bias. Given that my results show that selection in response to electrification is important in the context of this paper, I will account for both biases in my estimation.

5.4 Results

5.4.1 Firm-Level Revenue Productivity

In this chapter, my goal is to evaluate the effect of electrification on average firm-level productivity. Productivity is defined as the efficiency with which a firm transforms inputs into output. I take the above methodology to the data and discuss some of the data-related issues that arise as necessary. Let $F(.)$ be an industry level production technology. Output quantity Q_{it} of firm i in year t if produced according to $Q_{it} = exp(\phi_{it})F(\mathbf{X}_{it}, \beta)$. Firm productivity is ϕ_{it}, $\mathbf{X}_{it}$ is a vector of production inputs; capital, labor, and electricity. Typically, physical output Q is not observed. Instead we observe firms sales revenue $R_{it} = P_{it} * Q_{it}$. Consider the revenue based production function (in logs):

$$r_{it} = p_{it} + q_{it} = f(\mathbf{x}_{it}, \beta) + \phi_{it} + p_{it} + \epsilon_{it} \tag{5.31}$$

where ϵ_{it} is an error term. Since also prices are unobersvable, the literature typically estimates revenue productivity, or profitability, TFPR, defined as:

$$TFPR_{it} = \phi_{it} + p_{it} \tag{5.32}$$

I first estimate the effect of electrifying a desa on average revenue productivity estimated following Olley and Pakes (1996) by running the following regression:

$$TFPR_{ivpst} = \alpha + \beta Access_{vpst} + \nu \mathbf{X}_{ivpst} + \eta \mathbf{V}_{vpt} + \gamma_p + \delta_{st} + \epsilon_{ivpst} \tag{5.33}$$

I use two estimates of revenue productivity for comparison. The first measure $TFPR_{op}$ is estimated following Olley and Pakes (1996) using investment as a proxy. The second measure $TFPR_{acf}$ is based on the Ackerberg, Caves, and Frazer (2015) refinement as detailed above, using electricity spending as proxy. Production function estimates from each method are presented in tables 5.5 and 5.6 in appendix 6.5.

Results from regression 5.33 with both measures of TFPR, using distance to the hypothetical grid as an instrument, are presented in table 5.2. The OLS results are presented in panel A and the IV results are presented in panel B. As before, the OLS estimate are smaller in magnitude than the IV estimates and are biased towards zero. Using the Olley Pakes TFPR measure in column (1) panel B, the causal effect of electrification on average revenue productivity in a desa is an 18% increase. Column (2) panel B shows that the results are similar when using $TFPR_{acf}$ and are not sensitive to the method used or to the proxy. Panel C presents the reduced form results and indicated that the closer a desa is to the hypothetical least cost grid, the higher average revenue productivity is.

It is worth noting that the goal of this regression is to estimate the causal difference in average revenue productivity between electrified and non-electrified desas. This is theoretically consistent with the empirical model in section 5.2 and the first order Markov process assumption on the law of motion of productivity in equation 5.25. It is common in the literature to run within firm regressions of productivity on a treatment variable and interpret the effect as causal. This is theoretically inconsistent with the Olley and Pakes (1996) and other methodologies which these papers typically follow. This is because these methodologies assumes that productivity only depends on its lag. The assumption is therefore that productivity cannot depend on or be affect by anything that is not its lag. The correct way of doing this would be to incorporate the treatment variable in the law of motion of productivity as in De Loecker (2013).

This criticism does not apply in the case of equation 5.33 for the simple reason that I am not estimating within firm changes in productivity (I do not include firm fixed effects). The interpretation of the coefficients in panel B is a causal difference in TFPR between electrified and non-electrified desas, caused by the competitive effects of electrification.

Table 5.2: Effect of electrification on TFPR.

Sample: Firm-Level		
	(1)	(2)
Dependent Variable	log(TFPR$_{op}$)	log(TFPR$_{acf}$)
Panel A: OLS		
Access$_{vpt}$	0.0184*	-0.008
	(0.0100)	(0.0138)
Panel B: IV		
Access$_{vpt}$	0.177**	0.353**
	(0.089)	(0.138)
First Stage F	43.76	43.76
Panel C: Reduced Form IV		
Z (KM)	-0.000486**	-0.001***
	(0.000236)	(0.00034)
Observations	134,391	134,391
IndustryxYear FE	✓	✓
Province FE	✓	✓
Geo Controls	✓	✓
Firm Controls	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV and Reduced-Form regressions of TFPR on access defined at the desa level. TFPR is measured following Ackerberg, Caves, and Frazer (2015) using electricity spending as a proxy. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

5.4.2 Reallocation at the Regency-by-Industry Level

The evidence so far indicates that electrification increases firm turnover in a desa by allowing more firms in and increasing the probability of exit. This leads to an increase in the average firm productivity in the manufacturing sector. Does electrification improve the reallocation of resources towards more productive firms? To answer this question, I aggregate revenue productivity at the regency-by-industry level. A regency is the second highest administrative division in Indonesia. There are around 100 regencies in Java. On average a regency has 250 desas and around 250 firms per regency. An industry is a two-digit industry classification. I call each regency-by-industry pair a sector. I decompose the sector TFPR index Ω_{st}, defined as the revenue-weighted average of log firm revenue productivity TFPR in an industry s in year t, into an unweighted average and a covariance term (Olley and Pakes (1996)):

$$\begin{aligned}\Omega_{st} &= \sum_{i=1}^{N} S_{it} TFPR_{it} \\ &= \frac{1}{N}\sum_{i=1}^{N} TFPR_{it} \sum_{i=1}^{N}(S_{it} - \frac{1}{N})(TFPR_{it} - \frac{1}{N}\sum_{i=1}^{N} TFPR_{it}) \\ &= \overline{TFPR_{st}} + Ncov(S_{it}, TFPR_{it})\end{aligned} \tag{5.34}$$

where S_{it} is firm i revenue share in sector s. $\overline{TFPR_{st}}$ is the unweighted average of log revenue productivity across all firms in industry s in year t. The Olley-Pakes covariance term measures allocative efficiency. It is higher when more productive firms have larger market shares. I test how electrifying more desas within a regency affects the industry. I define $Access_{st}$ as a dummy = 1 if at least 0.5 of firms are within $15KM$ of the nearest substations. I use a similar identification strategy as at the desa level where I instrument access with the average distance in the industry to the hypothetical grid. The estimation equation is:

$$Y_{st} = \alpha + \beta Access_{st} + \gamma_{pt} + \delta_s + \epsilon_{st} \tag{5.35}$$

where with province-by-year fixed effect and sector fixed effect. Table 5.3 presents the results. The IV estimates in panel B show that access increases both weighted and unweighted productivity at the sector level. In addition, the Olley Pakes covariance term increases with access. This means that electrification increases the covariance between market share and revenue productivity. Reallocation is more efficient in regions-by-industry groups with larger electrified proportions. This is evidence for a firm turnover mechanism where electrification helps reallocating resources towards more productive firms.

Table 5.3: Olley-Pakes Revenue Weighted Productivity Decomposition

Sample: Sector-Level			
	(1)	(2)	(3)
Dependent Variable	Weighted Average $\log(\text{TFPR}_{OP})$	Unweighted Average $\log(\text{TFPR}_{OP})$	Covariance $(\log(\text{TFPR}_{OP}), share)$
Panel A: OLS			
Access_{vpt}	0.140***	0.0114	0.121***
	(0.0249)	(0.0123)	(0.0197)
Panel B: IV			
Access_{vpt}	0.550***	0.261***	0.278**
	(0.163)	(0.0945)	(0.109)
First Stage F	36	36	36
Panel C: Reduced Form IV			
Z (KM)	-0.00213***	-0.000998***	-0.00106***
	(0.000549)	(0.000292)	(0.000410)
Observations	9,899	9,899	9,899
Industry FE	✓	✓	✓
ProvincexYear FE	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parentheses clustered at the sector level

5.5 Electrification and Capital-Biased Technological Change

In this section, I investigate whether electrification leads to a change in the production technology of firms in a way that is biased towards capital. The idea is that if electricity and capital are complementary, then having access to cheaper electricity might lead the firm to invest in machines that are more productive but more electricity intensive. This is one potential explanation for the observed differential input substitution responses between access electricity, capital and labor found in chapter 4 table 4.2. In other words, what could be driving these effects on input ratios is a change in the production function coefficients, instead of only a change in the inputs relative prices.

To test this possibility, I estimate a production function following the above procedure allowing the coefficient on capital, β_k, to be a function of access to the grid. In order to check whether access affects capital differently than labor, I estimate a production function allowing the capital coefficient to depend on access. I follow Olley and Pakes (1996) using investment as a proxy to estimate the following production function:

$$Y = K^{\beta_k + \theta access} L^{\beta_l} M^{\beta_m} \exp(\phi) \exp(\epsilon)$$

where $access$ is an access dummy, ϕ is the hicks-neutral productivity term as before, and ϵ is an exogenous shock unobservable to the firm. Estimation equation:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \theta k_{it} * access_{it} + \phi_{it} + \epsilon_i t$$

Table 6.11 shows there is limited evidence for a change in capital coefficient. The estimate of θ is mostly zero across various industries, apart from two industriess. Standard errors are bootstrapped at the industry level. This suggests that access does not affect firms' investment decision in technology, at least not in a few years. This result is plausible, given that Indonesian manufacturing firms were using electricity in the 1990s, and since evidence from economic history shows that firms are slow in adopting new technologies (see for example Atkeson and Kehoe (2007)).

Table 5.4: Production Function Estimates with Capital Augment Energy

industry	β_l	β_m	β_k	θ
15	0.38***	0.48***	0.06***	0.008*
	(0.011)	(0.012)	(0.008)	(0.003)
16	0.25***	0.66***	0.04*	0.000
	(0.032)	(0.040)	(0.020)	(0.002)
17	0.34***	0.58***	0.04**	0.004
	(0.018)	(0.018)	(0.011)	(0.004)
18	0.37***	0.54***	0.02	0.002
	(0.020)	(0.014)	(0.012)	(0.006)
19	0.31***	0.59***	0.02*	0.004
	(0.021)	(0.020)	(0.018)	(0.011)
20	0.29***	0.64***	0.02	0.001
	(0.017)	(0.016)	(0.019)	(0.006)
21	0.33***	0.58***	0.12***	0.008
	(0.034)	(0.025)	(0.037)	(0.013)
22	0.40***	0.54***	0.04*	0.015*
	(0.029)	(0.023)	(0.019)	(0.007)
24	0.41***	0.52***	0.06*	0.010
	(0.021)	(0.019)	(0.019)	(0.006)
25	0.28***	0.65***	0.05**	0.008
	(0.013)	(0.012)	(0.011)	(0.005)
26	0.47***	0.44***	0.05**	0.005
	(0.014)	(0.009)	(0.015)	(0.003)
27	0.34***	0.57***	0.05	0.030
	(0.045)	(0.038)	(0.066)	(0.039)
28	0.27***	0.62***	0.06*	0.009
	(0.025)	(0.018)	(0.020)	(0.008)
29	0.32***	0.53***	0.08*	0.015
	(0.031)	(0.022)	(0.041)	(0.018)
31	0.35***	0.59***	0.12*	0.010
	(0.054)	(0.049)	(0.061)	(0.011)
32	0.30***	0.56***	0.05	0.012
	(0.053)	(0.040)	(0.084)	(0.073)
33	0.36***	0.55***	0.01	0.015
	(0.079)	(0.070)	(0.059)	(0.014)
34	0.48***	0.52***	0.02	0.016
	(0.044)	(0.025)	(0.058)	(0.043)
35	0.47	0.49***	0.07**	0.021
	(0.054)	(0.033)	(0.023)	(0.017)
36	0.41***	0.50***	0.05*	0.002
	(0.018)	(0.016)	(0.014)	(0.004)

5.6 Conclusion

To summarize, in this chapter I presented a model that lays out conceptually how access to electricity can affect productivity in the industry. I then discuss productivity measurement issues and present solutions from the literature with a special focus on electrification and the production function. Using productivity estimates, I show that electrification increases average productivity in the desa and enhances allocative efficiency at an aggregate level. These results show that electrification, by increasing firm turnover, has productivity-enhancing effects on the aggregate economy. This is evidence that poor infrastructure, no access to electricity in this case, can help explain the productivity gap between developing and developed countries. By lowering barriers to entry, not only do the new firms benefit from production, but this also has an aggregate effect by intensifying competition in the market and weeding out relatively unproductive firms in the market.

Appendix

5.A Limitations of the Model

This model is very simple and abstracts from many features that could be important in determining the effect of electrification on industry productivity.

- Trade: I assume that each location is a separate market and that firms don't sell in other locations. This is obviously an unrealistic assumption as these firms are medium and large manufacturing firms and the desas are too small to constitute their whole market. The model can be extended to allow for trade across location as in Melitz and Ottaviano (2008) and the comparative statics with respect to sunk cost of entry and input price in the location's own cutoff all go through. Therefore, we can still learn something from the simple closed economy model about the effect of electrification on productivity at the location level.

- Spillovers: Given that the true model involves trade across different locations and since most firms in my data produce tradable goods, the presence of spillovers across different locations complicates the interpretation of my results. Electrifying one location can have an effect on firms in other locations, and these effects are likely to be negative. What I estimate as the average difference between electrified and non-electrified locations could be therefore a combination of creation of new economic activity and relocation of economic activity from those who don't get electrified (or are already electrified) to locations that get newly electrified. An important question is whether there is any creation of new economic activity in response to electrification, or does electrification only displace economic activity? I addressed this question in the empirical section where I test for the presence spillovers in chapter 3 and 4. The results show that spillovers are minimal in this particular setting. Theoretically, the size of the spillovers depend the substitutability of the products being traded, transportation costs, and the number of trading partners. If transportation costs are very large, then spillovers will be minimal. Spillovers can also be minimal if there is a very large number of markets: the general equilibrium effects will be small because each market is too small to affect other markets.

5.B Additional Results

Table 5.5: Production Function Coefficients, Olley-Pakes

industry	Returns to Scale	β_k	β_l	β_m	β_e
15	1.07	0.07	0.21	0.50	0.29
16	1.07	0.02	0.38	0.59	0.08
17	1.03	0.03	0.32	0.64	0.04
18	0.99	0.00	0.34	0.59	0.06
19	1.04	0.03	0.32	0.52	0.17
20	1.08	0.03	0.35	0.68	0.02
21	1.46	0.02	0.55	0.73	0.16
22	1.07	0.06	0.35	0.55	0.12
23	1.11	0.24	0.01	0.79	0.06
24	1.05	0.05	0.32	0.59	0.09
25	1.02	0.06	0.26	0.66	0.05
26	1.09	0.03	0.61	0.41	0.04
27	2.50	0.26	1.67	0.15	0.42
28	1.02	0.05	0.23	0.66	0.08
29	1.00	0.09	0.23	0.57	0.11
31	1.30	0.05	0.15	1.04	0.06
32	1.19	0.10	0.03	0.60	0.46
33	1.14	0.02	0.53	0.56	0.03
34	1.07	0.00	0.37	0.56	0.14
35	1.12	0.03	0.53	0.56	0.00
36	1.04	0.04	0.37	0.55	0.07
Average	1.16	0.39	0.06	0.59	0.12

Notes: Estimated coefficient of a Cobb-Douglas production function in capital, labor, materials and electricity following Olley and Pakes (1996) using the investment proxy as detailed in section 5.3.

Table 5.6: Production Function Coefficients, ACF

industry	Returns to Scale	β_k	β_l	β_m	β_e
15	1.40	0.00	1.21	0.14	0.05
16	1.27	0.06	0.30	0.81	0.10
17	1.02	0.04	0.29	0.64	0.06
18	1.01	0.02	0.41	0.55	0.03
19	1.03	0.03	0.34	0.56	0.09
20	0.98	0.04	0.24	0.67	0.03
21	0.99	0.01	0.21	0.75	0.02
22	1.09	0.03	0.39	0.52	0.14
23	1.29	0.17	0.22	0.88	0.01
24	1.02	0.05	0.19	0.60	0.18
25	1.01	0.04	0.26	0.68	0.04
26	1.08	0.03	0.59	0.42	0.04
27	1.03	0.14	0.11	0.73	0.05
28	1.05	0.05	0.34	0.42	0.24
29	1.09	0.07	0.46	0.54	0.02
31	1.05	0.02	0.23	0.62	0.18
32	1.08	0.09	0.29	0.58	0.12
33	1.10	0.02	0.27	0.52	0.30
34	1.10	0.00	0.41	0.48	0.21
35	1.11	0.03	0.49	0.56	0.03
36	1.17	0.06	0.68	0.41	0.03
Average	1.09	0.38	0.05	0.57	0.09

Notes: Estimated coefficient of a Cobb-Douglas production function in capital, labor, materials and electricity following Ackerberg, Caves, and Frazer (2015) using the electricity spending proxy as detailed in section 5.3.

Chapter 6

Electrification, Marginal Cost, and Market Power

6.1 Introduction

In the previous chapter, I evaluated the effect of electrification on average productivity in the industry and reallocation. I relied on revenue-based production function estimation techniques to estimate production function coefficients and ultimately back out revenue productivity estimates, TFPR. Traditional techniques of revenue-based production function estimation (e.g. Olley and Pakes (1996), Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015)) suffer from well-known demand side biases as documented by the literature [1].

The reason why sales revenue is more widely used as the output variable when estimating production functions is that physical quantities are seldom observed in typical firm-level datasets. Using sales revenue instead of physical quantity risks contaminating the estimation by prices in two ways. The first is that the presence of prices in the productivity term creates another sources of endogeneity of inputs and failing to account for that will result in inconsistent production function coefficients. Second, when using TFPR measures to evaluate the effect of a certain treatment, electrification in the case of this paper, on productivity, the estimated effect will be a combination of the effect of the treatment on productivity and the

[1] Foster, Haltiwanger, and Syverson (2008) show that unobserved demand shocks can significantly bias estimates of productivity, or marginal cost.

effect on price. Recall from chapter 5 equation 5.32:

$$TFPR_{it} = \phi_{it} + p_{it}$$

Suppose electrification increases average productivity ϕ_{it} in the market. But higher productivity leads to lower prices. Therefore electrification would lower average prices in the market. Given these two opposing effects, using TFPR measures to evaluate the effect of electrification on industry productivity is not ideal as the demand side effects will confound the resulting estimates. I present a solution in this chapter, and discuss additional issues that would arise from this solution.

I estimate a quantity-based production function at the industry level, using rich product-level price and quantity data. Given that price data are available and mark-ups can be recovered from production function estimates, I will provide structural estimates of marginal cost. Using mark-up and marginal cost estimates, I will analyze the effect of electrification on the firms' cost structure and market power.

6.2 Quantity-Based Production Function Estimation

One way to avoid demand side biases is to estimate a quantity-based, product-level, production function. I acquired a supplement to the manufacturing census where I observe for each product the firm produces in a certain year the sales revenue and the sales volume, or quantity. Using revenue and quantity data allows me to calculate prices.

I therefore take advantage of the price and physical quantity data which I observe (and are most likely set) at the product level to estimate a quantity-based product-level production function. Using the production function estimates, and making some structural assumptions on the data, I am able to estimate the following unobserved measures: (i) physical productivity, (ii) markups, and (iii) marginal cost. I detail the estimation methodology below.

Two additional biases arise when estimating quantity-based production function. The first is an input price bias since input quality is not observed. The second is the input allocation bias as input allocation across products within multi-product firms[2] is unobserved. I closely follow De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) in dealing with these biases with two differences. The first is the choice of inputs in the production function. I use a translog production function in capital, labor, and electricity[3]. The choice of functional form allows for a richer substitution pattern (relative to a Cobb-Douglas) between inputs to understand the role of access to energy in affecting marginal cost. Second, I allow unobservable input prices to depend on access. I describe briefly the procedure below[4].

6.2.1 Empirical Framework

First consider the production function of product j produced by firm i in year t in logs:

$$q_{ijt} = f_j(\mathbf{x}_{ijt}, \beta) + \phi_{it} + \epsilon_{ijt} \tag{6.1}$$

where the vector $\mathbf{x}_{ijt}$ contains $k_{ijt}, l_{ijt}, e_{ijt}$, the product specific physical capital, labor, and energy and β is a vector of production function parameters. In practice, for input x, we observe a deflated version of x_{ijt} at the firm level $\tilde{x}_{it}$ where the following relationship holds in logs:

$$x_{ijt} = \rho_{ijt} + \tilde{x}_{it} - w^x_{ijt} \tag{6.2}$$

In equation 6.2, ρ_{ijt} is the log share of firm input expenditure dedicated to product j and w^x_{ijt} if the log deviation of firm-product specific price of input from the industry average. Substituting 6.2 in 6.1 yields:

$$q_{ijt} = f_j(\tilde{\mathbf{x}}_{ijt}, \beta) + A(\rho_{ijt}, \tilde{\mathbf{x}}_{ijt}, \beta) + B(w^x_{ijt}, \rho_{ijt}, \tilde{\mathbf{x}}_{ijt}, \beta) + \phi_{it} + \epsilon_{ijt} \tag{6.3}$$

The $A(.)$ function represents the bias stemming from unobserved input allocation across products within firm. I deal with this bias first by estimating the production function for single product firms only[5] while correcting for selection into being a single product firm[6]. When estimating equation (6.3) for single product firms,

[2] The median number of products per firm per year is 2.

[3] The translog production function takes the following form

$$q = \beta_k k + \beta_{kk} k^2 + \beta_l l + \beta_{ll} l^2 + \beta_e e + \beta_{ee} e^2 + \beta_{lk} lk + \beta_{ke} ke + \beta_{le} le + \beta_{lek} lek + \phi$$

[4] I refer the reader to De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for a more detailed discussion.

[5] This is a sub-sample of all firms that are producing a single product at any point in time, including firms that become multiproduct firms in later periods (and vice versa) and those who remain single product.

[6] a procedure similar to controlling survival as in Olley and Pakes (1996).

the $A(.)$ term drops out. The $B(.)$ term represents the input price bias. De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) show that input prices are a function of output prices p_{it}[7] and other variables proxying for product quality such as market share ms_{it}, location dummies G_i and product dummies K_i. In addition to these variable, I allow input prices to depend on access C_{it}. This gives rise to the following input price control function[8]:

$$w_{it}^{x} = w_t(p_{it}, ms_{it}, K_i, G_i, C_{it}) \tag{6.4}$$

In practice, w_{it}^{x} is estimated by a polynomial in the terms of the $w_t(.)$ function.

This leaves one bias remaining, which is the classical bias from unobserved productivity ϕ_{it}. I follow the literature as in Olley and Pakes (1996), Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015) by using the first order condition of a variable input, in my case electricity spending, as a proxy for productivity[9].

The final step is to deal with the input allocation bias represented by the $A(.)$ term for multi-product firms. Given the estimated production function coefficients and the input price control, ϕ_{it} and the ρ_{ijt}'s can be solved for using the residual from 6.3, as the only unknown is the ρ_{ijt}'s from $A(.)$ function[10] and ϕ_{it} is the constant. This is done by solving a simultaneous system of equations where the left hand side is the residual from 6.3 and the unknowns are the ρ_{ijt}'s and ϕ_{it}.

6.2.2 Production Function Estimates

Following the procedure above, I estimate a production function for each two-digit industry using product level quantity data. I assume $f_j(\tilde{\mathbf{x}}_{ijt}, \beta)$ to be a translog function in capital, labor and electricity spending. Below, I present the average output elasticities and discuss how electrification affects the different inputs differently.

Table 6.1 presents the average output elasticities with respect to the three inputs: capital, labor, and energy across all firms from all industries in the data. The break-

[7] Vertical differentiation model

[8] Coefficients of the input price control function are not separately identified by input, so they have to be firm specific instead of product specific.

[9] I implement the one step estimator as suggested by Wooldridge (2009)

[10] We know the functional form

down of average elasticities by industry is presented in table 6.11 in Appendix 6.5. The average output elasticity with respect to capital is 3%, with respect to labor 40%, and with respect to energy is 24%. The second row of table 6.1 presents the standard deviation of these elasticities: the translog production function allows different elasticities across firms within an industry. The low elasticity of output with respect to capital is expected. With constant returns to scale, the capital share is expected to be around 30% and labor share around 70%. Once electricity or materials are included in the production function, the capital share decreases substantially. Since production in developing countries is more labor intensive, the capital share is expected to be low, especially when including electricity. Looking again at the average elasticity of output with respect to energy of 24%, it is evident that energy, electricity in particular, is an important factor of production. The detailed industry-level average output elasticities can be found in tables 6.11 in Appendix 6.5.

Table 6.1: Average Output Elasticities

	Capital	Labor	Energy
Mean	0.03	0.40	0.24
Std Dev	(0.09)	(0.23)	(0.16)

In order to understand the substitution patterns between the different inputs, table 6.2 shows the average translog production function coefficients across all industries. Table 6.10 in appendix 6.5 presents the industry-level estimated production function coefficients including those on the interaction between capital, labor, and energy across different industries. β_{ke} is consistently larger that β_{le} indicating that energy and labor are stronger substitutes than energy and capital. On average, β_{ke} is 2.3%. Not surprisingly, the sign of β_{ke} indicates a certain degree of complementarity between capital and electricity. On the other hand, β_{le} is on average equal to -4.6%, showing that energy and labor are substitutes. These estimates are in line with the firm-level IV estimates on input ratios in table 4.2. Capital and energy are more complementary relative to labor and energy, creating differential re-optimization responses in their respective ratio to energy. This leads to a more capital-intensive production process.

Table 6.2: Average Production Function Coefficients

Coefficient	Mean	Std Dev
β_k	-0.076	(1.034)
β_l	0.232	(1.492)
β_e	0.085	(1.623)
β_{kk}	-0.001	(0.013)
β_{ll}	0.032	(0.050)
β_{ee}	0.028	(0.030)
β_{lk}	0.005	(0.112)
β_{ke}	0.023	(0.097)
β_{le}	-0.046	(0.150)
β_{lek}	-0.001	(0.009)

A useful exercise is to see whether these production function estimates, coupled with the reduced from estimates of the effect of access on inputs from table 4.1 predict well the effect of access on sales.

$$\begin{aligned}\frac{\partial q}{\partial C} &= \frac{\partial q}{\partial k} * \frac{\partial k}{\partial C} + \frac{\partial q}{\partial l} * \frac{\partial l}{\partial C} + \frac{\partial q}{\partial e} * \frac{\partial e}{\partial C} \\ &= 3\% * 3.42 + 39\% * 1.79 + 24\% * 4.01 \\ &= 1.76\end{aligned} \tag{6.5}$$

This exercise predicts, given the estimates output elasticities with respect to inputs and the effect of access on inputs, that access would increase output by 1.76 times. This number is comparable to and is around 70% of the reduced form estimate of access on firm deflated sales in table 4.1, chapter 4. Note that this number includes a total effect of access on average firm deflated sales as it combines the direct effect of access on incumbents as well as changes induced by entry and exit.

6.2.3 Illustrating Demand-Side Bias in TFPR measures

I argued above that using TFPR measures to evaluate the effect of electrification on productivity is not ideal. This is because of the demand-side bias resulting from the presence of price in TFPR measures. I now illustrate this bias by comparing the effect of electrification on TFPR estimates from chapter 5 to the effect of electrification on physical productivity TFPQ measure estimated in this chapter.

Table 6.3 shows the OLS, IV, and reduced-form results of regressing TFPR on access, using the distance to the hypothetical least cost grid as an instrument. The OLS estimates are again smaller in magnitude than the IV estimate. Focusing on column (1) panel B, I find that on average electrifying a desa increases revenue productivity in the desa. To explore heterogeneity in the effect of access average revenue productivity across entrants and incumbents, proxied by firm age, I estimate the same equation for young and old firms separately. A young firm is a firm whose age is below the median age. IV regressions in panel B show that this increase in average revenue productivity is driven by an increase in the revenue productivity of younger firms. This evidence is not necessarily consistent with a turnover channel where electrification induces the inefficient incumbents to exit. We would expect that in that case the average productivity of older firms is also higher. I replicate this analysis using different productivity estimates following Ackerberg, Caves, and Frazer (2015) and using electricity spending as a proxy for robustness. The results are presented in table 6.12 in appendix 6.5 and yield the same conclusion.

A potential reason could be demand side biases that arise when estimating a revenue based production function. The above method fails to account for the biases caused by the presence of price in $TFPR$. The goal is to check if connected firms have on average higher physical productivity, ϕ_{it}. Testing this channel with regressions of $TFPR_{it}$ on access is not ideal. To see why, consider equation (5.32). Suppose that access increases the average productivity ϕ_{it}. But price and productivity ϕ_{it} are negatively correlated: more productive firms have lower marginal costs and therefore lower prices. This means that if access increases the average ϕ_{it} in the market and decreases the average price, the two effects can potentially cancel out.

Therefore, given that TFPR estimates are a combination of productivity and prices, the results in columns (2) and (3) of table 6.3 could be driven by a differential effect of access on prices for younger and older firms.

Table 6.3: Effect of electrification on TFPR by Age Group.

Sample: Firm-Level			
	All	**Young**	**Old**
	(1)	(2)	(3)
Dependent Variable	log(TFPR)	log(TFPR)	log(TFPR)
Panel A: OLS			
$Access_{vpt}$	0.0184*	0.0179	0.0169
	(0.0100)	(0.0148)	(0.0105)
Panel B: IV			
$Access_{vpt}$	0.177**	0.369***	0.060
	(0.089)	(0.003)	(0.096)
First Stage F	43.76	36.81	33.08
Panel C: Reduced Form IV			
Z (KM)	-0.000486**	-0.0010***	-0.00016
	(0.000236)	(0.00032)	(0.00025)
Observations	134,391	47,921	86,439
IndustryxYear FE	✓	✓	✓
Province FE	✓	✓	✓
Geo Controls	✓	✓	✓
Firm Controls	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV and Reduced-Form regressions of TFPR on access defined at the desa level. TFPR is measured following Olley and Pakes (1996). Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

To separate the effects, I use the product-level price data and physical productivity estimates. I estimate the following equation for product j (which is a subset of industry s) produced by firm i in desa v, province p, industry s and year t is:

$$y_{jivpst} = \alpha + \beta Access_{vpst} + \nu \mathbf{X}_{ivpst} + \eta \mathbf{V}_{vpt} + \gamma_p + \delta_t + \delta_j + \epsilon_{jivpst} \tag{6.6}$$

where δ_j are product-level fixed effects.

Table 6.4 shows the results from regressing log price and ϕ_{it} on access for all, young and old firms. The OLS estimates in panel A are smaller in magnitude than the IV estimates as before. The IV estimates of the effect of access on ϕ_{it} in panel B columns (2), (4) and (6) are all positive, significant and of the same magnitude, indicating that the difference in the average physical productivity of electrified and non-electrified firms is the same across firm cohorts. The coefficient in column (3) panel B shows that the difference in price between products produced by young connected firms and young unconnected firms is not statistically different from zero. However, there is a negative effect of access on the average price of products produced by older connected firms.

Together with the results on TFPR from table 6.3, the results from this table are consistent with a significant demand side bias. There are two forces in play: electrification increases average productivity in the market through selection and decreases average prices at the same time. Therefore the effect of electrification on revenue productivity is ambiguous, but unambiguously positive on "true" or physical productivity. Going back to the model in section 5.2, these results show that the effect of electrification on ϕ^* is positive, in other words, electrification increases selection in the market. Using the sharp theoretical prediction in equation (5.15), this results is consistent with electrification reducing barriers to entry, attracting more entrepreneurs to the market, which in turn intensifies competition and forces the less productive firms to exit. This evidence does not rule out an effect of electrification on input prices p_x, but it does indicates that first effect through a reduction in barriers dominates the second.

Table 6.4: Impact of access on Price and ϕ_{it} by Age Group.

	All		Young		Old	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	log(Price)	ϕ_{it}	log(Price)	ϕ_{it}	log(Price)	ϕ_{it}
			Panel A: OLS			
Access_{vpt}	-0.0125	0.108***	-0.0191	0.208***	-0.0129	0.0633
	(0.0261)	(0.0340)	(0.0414)	(0.0532)	(0.0291)	(0.0388)
Observations	127,427	127,427	40,406	40,406	86,226	86,226
			Panel B: IV			
Access_{vpt}	-0.375	0.932***	0.0845	0.931*	-0.576*	0.804*
	(0.245)	(0.355)	(0.397)	(0.532)	(0.319)	(0.427)
Observations	127,427	127,427	40,406	40,406	86,226	86,226
First Stage F	25.23	25.23	17	17	16.27	16.27
			Panel C: Reduced Form IV			
Z (KM)	0.000803	-0.00199***	-0.000193	-0.00213*	0.00109**	-0.00152**
	(0.000500)	(0.000678)	(0.000901)	(0.00120)	(0.000521)	(0.000713)
Observations	127,427	127,427	40,406	40,406	86,226	86,226
IndustryxYear FE	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV and Reduced-Form regressions of two different measures of TFPR on access defined at the desa level using equation 6.6. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

6.3 Prices, Marginal Costs, and Mark-ups

In this section, I use the product-level production function estimates to understand how electrification affects the cost of production of the firm as well as its market power. Before turning to the marginal cost estimates, I start by looking at what happens to product-level output when the grid arrives.

6.3.1 Electrification and Product-Level Output

Table 6.5 shows the effect of electrification on product-level sales revenue and sales quantity using specification (6.6). Panel A presents OLS estimates, which are positive and significant. This indicates that access and output are postively correlated at the product level, confirming the firm level results from chapter 4. Again, these OLS estimates are smaller in magnitude than the IV estimates in panel B. The IV estimates show that the causal effect of access on compliers is around a 1.5 times increase in average revenues and 2.5 times increase in average quantity sold relative to non-compliers. These results are comparable to the firm-level estimates in table 4.1 from chapter 4, although a bit smaller. This is possibly because this product-level starts from 1994 instead of 1990 like the firm-level panel, and is missing a few years from the early nineties where the effects of electrification might have been the largest. The fact that quantity increase by more than revenues is an indication that prices might be going down as a result of electrification. Panel C presents the reduced form regression results which state that the product-level output measures are decreasing in the distance to the hypothetical grid.

Table 6.5: Impact of access on the product-level Sales and Quantity.

Dependent Variable (Log)	(1) Sales	(2) Quantity
Panel A: OLS		
Access$_{vpt}$	0.145***	0.164***
	(0.0485)	(0.0562)
Panel B: IV		
Access$_{vpt}$	0.942**	1.317**
	(0.462)	(0.557)
First Stage F	28.68	28.68
Panel C: Reduced Form IV		
Z (KM)	-0.00213**	-0.00298**
	(0.00102)	(0.00120)
Observations	127,818	127,818
Product FE	✓	✓
Year FE	✓	✓
Province FE	✓	✓
Geo Controls	✓	✓
Firm Controls	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS , IV, and reduced form regressions of firm sales revenue and quantity on access, using specification (6.6). Access is defined at the desa level. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

6.3.2 Structural Estimates of Marginal Cost and Mark-up

In this section, I explore how access affects prices, marginal cost and mark-ups. Given production function estimates, it is possible to estimate markups using the elasticity of sales with respect to a variable input[11] and the share of expenditure on that input in the total sale of the product. To illustrate, consider firm i's cost minimization problem in time t using production technology $F_{it}(X^1_{it},...,X^V_{it},K_{it})e^{\phi_{it}}$ where X^v_{it} if variable input v, and K_{it} is the dynamic capital input and ϕ_{it} is (log) productivity. The Lagrangian function is therefore:

$$\mathscr{L}(X^1_{it},...,X^V_{it},K_{it},\lambda_{it}) = \sum_{v=1}^{V} W^v_{it}X^v_{it} + r_{it}K_{it} + \lambda_{it}(Q_{it} - F_{it}(X^1_{it},...,X^V_{it},K_{it})e^{\phi_{it}}) \quad (6.7)$$

Realizing that the Lagrangian multiplier is the shadow price of increasing output by one unit, λ_{it} is in fact the marginal cost of firm i in year t. Taking the first order condition with respect to variable input X^v_{it} and rearranging gives the following expression for markup $\mu_{it} \equiv \frac{P_{it}}{\lambda_{it}}$

$$\mu_{it} = \theta^v_{it}(\alpha^x_{it})^{-1} \quad (6.8)$$

where $\theta^v_{it} \equiv \frac{\partial F_{it}(.)\phi_{it}}{\partial X^v_{it}} \frac{X^v_{it}}{Q_{it}}$ is the elasticity of output with respect to input X and $\alpha^x_{it} \equiv \frac{W^v_{it}X^v_{it}}{P_{it}Q_{it}}$ is the share of expenditure on X in the total sales which is observable in the data. If price data is available, then marginal cost can be estimated as the ratio of price to markup:

$$MC_{it} = \lambda_{it} = \frac{P_{it}}{\mu_{it}}$$

6.3.3 Results

As I describe above, production function estimates, coupled with observed product level prices, allow me to obtain structural estimates of marginal cost. Table 6.6 investigates the effect of access on prices, marginal cost, and mark-ups using specification (6.6), which includes product-specific fixed effect (9 digit product code) to ensure that the same type of products are compared.

The IV results are in panel B. Column (1) shows that effect of access on prices is negative, suggesting a competitive effect of electrification, although the difference be-

[11] Typically materials or electricity.

tween price of electrified and non-electrified desas is not significant. This however could be masked by selection effects and I explore that in the next section. Column (2) shows that access causes the average marginal cost of firms in electrified desas to decrease by 55% relative to non-electrified desas. Column (3) panel B shows that products produced by connected firm have higher markups. Going back to equation (5.17) from the model predictions in chapter 5 section 5.2, average markups conditional on survival are equal to $\bar{\mu} = E[\mu(c)|c < c^*] = \frac{1}{2}(c^* - \bar{c})$. The model gives an ambiguous prediction on what would happen to electrification in response to electrification. This is because there are two forces in play: a competitive force that forces firms to decrease their mark-ups, and a selection force that attracts relatively more efficient firms (or products) to the market. The result in panel B column (3) therefore shows that competition effect of electrification is outweighed by the selection effects.

Table 6.6: Impact of access on the product-level price, marginal cost, and markup.

Dependent Variable (Log)	(1) Price	(2) Marginal Cost	(3) Markup
	Panel A: OLS		
Access_{vpt}	-0.0199	-0.0591*	0.0392
	(0.0259)	(0.0347)	(0.0272)
	Panel B: IV		
Access_{vpt}	-0.318	-0.798***	0.480*
	(0.221)	(0.304)	(0.253)
First Stage F	28.68	28.68	28.68
	Panel C: Reduced Form IV		
Z (KM)	0.000718	0.00180***	-0.00109**
	(0.000485)	(0.000624)	(0.000547)
Observations	127,818	127,818	127,818
Product FE	✓	✓	✓
Year FE	✓	✓	✓
Province FE	✓	✓	✓
Geo Controls	✓	✓	✓
Firm Controls	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS , IV, and reduced form regressions of firm price, marginal cost, and markup on access using specification (6.6). Access is defined at the desa level. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

As before, the estimated causal difference between electrified and non-electrified desas is a combination of the effect of electrification on incumbents and a selection effect where electrification attracts a different type of firms or products (or both).

The product-level results from this section are consistent with the firm-level results on sales and input demands. These results also confirm that one contributing mechanism to the positive effect of electrification on firm performance is cost-based. Electrification decreases average marginal cost of production in the industry, leading to an increase in average sales. Firms connected to the grid are therefore more productive and electrification promotes manufacturing efficiency.

6.3.4 Effect of Access on Incumbent Products

I now turn to evaluating the effect of access on *within* product-firm pair production cost and pricing, or incumbent product-firm pairs. This exercise is useful because it nets out the selection effects of electrification; both the selection of new firms and the selection of new products produced by existing firms as a response to electrification.

I estimate the below equation where outcomes y_{jivpst} include price, marginal cost, and mark-up, all in log at the product level. I include a product-by-firm fixed effect to compare what happens within each firm-product pair when the grid arrives. Again, I instrument $Access_{vpst}$ with the distance to the hypothetical grid interacted with time dummies.

$$y_{jivpst} = \alpha + \beta Access_{vpst} + \nu \mathbf{X}_{ivpst} + \eta \mathbf{V}_{vpt} + \gamma_{ji} + \delta_t + \epsilon_{jivpst} \tag{6.9}$$

where γ_{ji} is a product-by-firm fixed effect.

Table 6.7 presents the OLS, IV and reduced-form regression results of equation (6.9). Panel A presents the OLS results for the price, marginal cost and mark-up. The OLS is smaller in magnitude and biased towards zero relative to the IV estimates. Panel B presents the IV results. Column (1) shows that for the same product a particular firm produces, electrification reduces the price of that product by 54%. Compared to the effect of access on average prices in table 6.6 panel B column (1), this effect is large and significant. This suggests that electrification selects prices with higher prices into the market.

Column (2) shows that marginal cost falls by a similar magnitude as the fall in price. Consistently, the effect on mark-up in column (3) is zero. This indicate that the efficiency savings within a product-firm resulting from electrification are completely passed through to the consumer. In addition, this effect is similar to the estimated effect in table 6.6 panel B column (2), indicating that selection effect is similar to the effect of electrification on incumbents.

Focusing again on the effect on mark-up in column (3), the fact that within product-firm markup doesn't fall and that across firms the mark-up increases is evidence for selection where electrification selects a different type of firms or products that are able to charge higher markups. I now investigate which firms charge higher mark-ups.

Table 6.7: Product-level price, marginal cost, and mark-up

Sample: Product-Level			
	(1)	(2)	(3)
Dependent Variable (Log)	Price	Marginal Cost	Markup
	Panel A: OLS		
Access_{vpt}	0.0476	0.0377	0.00997
	(0.0347)	(0.0343)	(0.0212)
	Panel B: IV		
Access_{vpt}	-0.782***	-0.701**	-0.0753
	(0.277)	(0.289)	(0.139)
First Stage F	7.53	7.53	7.53
	Panel C: Reduced Form IV		
Zx1995	-0.000358	4.16e-05	-0.000399
	(0.000422)	(0.000563)	(0.000358)
Zx1996	0.00517***	0.00495***	0.000222
	(0.000636)	(0.000774)	(0.000433)
Zx1997	0.000917	0.000524	0.000393
	(0.000578)	(0.000724)	(0.000473)
Zx1998	0.000128	0.000894	-0.000766
	(0.000676)	(0.000800)	(0.000525)
Zx1999	-7.72e-05	-1.28e-05	-6.45e-05
	(0.000735)	(0.000889)	(0.000575)
Zx2000	0.000877	0.000956	-7.93e-05
	(0.000734)	(0.000887)	(0.000555)
Observations	111,029	111,029	111,029
FirmxProduct FE	✓	✓	✓
Year FE	✓	✓	✓
Geo Controls	✓	✓	✓
Firm Controls	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS, IV, and reduced-form regressions of product price, marginal cost, and markup on access defined at the desa level from specification (6.9). Mark-up is estimated following De Loecker, Goldberg, Khandelwal, and Pavcnik (2016). Marginal cost is calculated from observable prices and estimated mark-ups. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Table 6.8 presents the results from a regression of log mark-ups on access, but splitting the sample by young and old firms. Splitting the sample this way will allow us to understand which type of firms, i.e entrants or incumbents (proxied by age) are the ones who are charging the higher mark-ups. Given that the results from table 6.7 show that incumbents pass-through all their cost savings due to electricity to consumers, it is possible that the selection effect driving the increase in average mark-up in electrified areas is driven by new firms coming in and charging higher mark-ups. Results in table 6.8 confirm this intuition. Column (1) presents the effect of access on average mark-ups charged by young firms and column (2) the effect of access on average mark-ups charged by old firms. The estimated IV coefficients in panel B show that the average mark-up charged by young firms in electrified desas is higher than the average mark-up charged by the same age group of firms in non-electrified areas. In contrast, there is no statistically significant difference between the average mark-ups charged by old firms in electrified and non-electrified desas. These results are consistent with complete pass-through estimated in table 6.7. This suggests that electrification is selecting a different type of firms into the market, and these firms are charging higher mark-ups.

Table 6.8: Impact of access on the product-level price, marginal cost, and markup, by age group.

	Young	Old
	(1)	(2)
Dependent Variable (Log)	Markup	Markup
Panel A: OLS		
Access_{vpt}	0.00149	0.0628*
	(0.0369)	(0.0330)
Panel B: IV		
Access_{vpt}	0.718*	0.389
	(0.407)	(0.303)
First Stage F	19.14	18.89
Panel C: Reduced Form IV		
Z (KM)	-0.00170**	-0.000787
	(0.000821)	(0.000614)
Observations	40,305	86,715
Product FE	✓	✓
Year FE	✓	✓
Province FE	✓	✓
Geo Controls	✓	✓
Firm Controls	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS , IV, and reduced form regressions of markup on access following equation (6.6). Access is defined at the desa level. Robust standard errors in parentheses clustered at the desa level. The sample in column (1) is that of young firm where young is define as age below median. The sample in column (2) is the sample of old firms with age above median. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

6.4 Product Differentiation

The results from table 6.7 suggest that when the grid arrives to a desa, the price and marginal cost of incumbent firms that keep producing the same products decrease substantially. Together with the results from table 6.6, this indicates that the new firms who enter post access are producing the *same* products (since I include 9-digit product fixed effects) but are able to charge higher mark-ups.

Why are the new firms or new products selected into the market by electrification able to charge higher mark-ups? If these firms were producing exactly the same products as those produced by incumbents then they wouldn't be able to charge higher mark-ups. Note that in all the above specification I include a fixed effect for each product (9 digit level). A demand side effect where consumers in electrified desas are willing to pay more for products in these desas can be ruled out since this would mean that also incumbents can now increase their mark-ups.

For these firms to be able to charge higher mark-ups, it must be that they are producing some differentiated products within the narrow definition of products I use in the analysis. Product differentiation could be horizontal (different characteristics) or vertical (better quality). Distinguishing between these two types of product differentiation is not directly possible due to the fact that product characteristics and product quality are not observable and cannot be estimated, at least without imposing some restrictive structural assumptions on the data. Instead, I aim to provide some suggestive evidence that firms with higher mark-ups are systematically different in their product mix and are more likely to provide different products or innovate.

Table 6.9 evaluates the effect of access to the grid on the likelihood that a product that was produced in the last two years - by the firm, in the desa, or in the district - is still being produced in year t. Each of the outcome variables in table 6.9 is meant to capture how much innovation or change in the product mix of the firm is taking place, relative to the firm itself (columns (1) and (4)), relative to itself and products produced by other firms in the desa (columns (2) and (4)) and relative to itself and the other products available in the district (columns (3) and (6)). I take these three dummies to be a measure of innovation in the product mix.

Columns (1)-(3) present the OLS, IV and reduced form regressions corresponding to specification (6.6). Panel A shows the OLS estimates, which show that there is a negative correlation between access and innovation. The IV estimates in panel B show the causal effect of access on the difference in innovation between firms in electrified and non-electrified desas. The coefficient in column (1) panel B shows that products in electrified desas are 15% less likely to have been produced by the same firm in the last two years. In other words, electrification leads on average to more innovation in the product mix within the same firm in the cross section. In column (2) panel B the IV coefficient is also negative and significant. It indicates that products produced in electrified desas in period t are 15.5% less likely to have been produced in the previous two years in that desa. Similarly, the coefficient in column (3) panel B shows that products produced in electrified desas in period t are 12% less likely to have been produced in the previous two years in that same district. Together, these results suggest that electrification increases innovation or change in the product mix that firm produce relative to the products produced by that same firm in the past, but also relative to other products produced in the same location. This is evidence for some degree of product differentiation as a result of electrification, and could potentially explain why mark-ups in electrified desas are higher.

Columns (4)-(6) in table 6.9 replicate the same analysis but with the addition of firm-by-product fixed effects as in specification 6.9 in order to estimate the effect of access on incumbents' product mix. The coefficient in column (4) panel B means that a product is 30% more likely to be produced by the same firm if it has been produced by that same firm after it gets access to the grid. Given that electrification leads to substantial cost savings within the same firm-product pair (table 6.7), it is not surprising that firms will keep producing the same product post access.

Relative to the coefficient in column (1), this estimate only captures the effect on incumbents as opposed to a combination of the effect on incumbents and the selection effect of firms that enter after the grid arrives. The fact that the effect on the incumbents' likelihood of keeping the same product is positive whereas in the cross section the effect is negative means that electrification is selecting firms that innovate or experiment more in the products they produce into the market.

The coefficients in columns (5) and (6) panel B show that incumbents are also more likely to produce products that were produced in the desa or the district the last two years when the grid arrived, although these effects are not statistically different than zero. Given that the coefficients in columns (2) and (3) are negative and significant, this again means that electrification selects firms that produce products that are different than what is offered in the desa. All this evidence suggests that electrification leads to more product differentiation. Table 6.13 in appendix 6.B to this chapter repeats the analysis in table 6.9 but going back 5 years instead of 2 for robustness. The conclusions of this section remain unchanged.

Table 6.9: Impact of access on product mix

	All			Incumbents		
	(1)	(2)	(3)	(4)	(5)	(6)
Produced in last 2 years	by firm	in desa	in district	by firm	in desa	in district
Panel A: OLS						
Access_{vpt}	-0.0101*	-0.00539	-0.00272	-0.00568	-0.0258	-0.0211
	(0.00565)	(0.00711)	(0.00664)	(0.0193)	(0.0181)	(0.0172)
Panel B: IV						
Access_{vpt}	-0.148***	-0.155***	-0.119**	0.301**	0.182	0.200
	(0.0513)	(0.0573)	(0.0570)	(0.150)	(0.130)	(0.128)
First Stage F	28.68	28.68	28.68	7.53	7.53	7.53
Panel C: Reduced Form IV						
Z (KM)	0.000334***	0.000351***	0.000270**			
	(0.000101)	(0.000108)	(0.000120)			
Zx1995				-0.00128***	-0.00105***	-0.00106***
				(0.000320)	(0.000288)	(0.000283)
Zx1996				-0.000537**	-0.000181	-0.000191
				(0.000268)	(0.000274)	(0.000264)
Zx1997				0.00201***	0.00107***	0.000953***
				(0.000416)	(0.000402)	(0.000369)
Zx1998				0.00111***	0.000712**	0.000696**
				(0.000350)	(0.000305)	(0.000285)
Zx1999				-0.000465	0.000150	0.000306
				(0.000319)	(0.000266)	(0.000257)
Zx2000				-0.000475	-0.000454*	-0.000309
				(0.000320)	(0.000276)	(0.000260)
Observations	127,818	127,818	127,818	111,029	111,029	111,029
Product FE	✓	✓	✓			
FirmxProduct FE				✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV and Reduced-Form regressions of product mix dummies on access using equation (6.6) in colums (1)-(3) and equation (6.9) in columns (4)-(6). The dependent variable in columns (1) and (4) is a dummy variable equal to one if product j is being produced by firm i in year t and has been produced by firm i any time since $t-2$. The dependent variable in colums (2) and (5) ((3) and (6)) is a dummy equal to one if product j is being produced by firm i in desa v (district d) in year t and has been produced in desa v (district d) any time since $t-2$. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

6.5 Conclusion

In summary, I find that electrification increases average physical productivity. Electrification leads to substantial cost savings in incumbent products, and these savings are completely passed through to consumers. However, electrification on average increases average mark-ups in the industry. This is due to the selection of possibly differentiated products that enter the market post electrification. I show that electrification selects products that are less likely to be produced before either by the firm itself or in the geographic vicinity.

Electrification has some welfare enhancing effects on the market in two ways: first it reduced the price of existing products that survive. Second, it increases the variety of products offered, which under some assumptions on demand is welfare improving.

Appendix

6.A Production Function Estimates

Table 6.10: Production Function Coefficients

industry	β_k	β_k	β_e	β_{kk}	β_{ll}	β_{ee}	β_{lk}	β_{ke}	β_{le}	β_{lek}
15	0.280	0.632	0.399	-0.014	0.016	0.035	-0.004	-0.009	-0.088	0.002
16	-0.082	0.012	0.291	-0.005	0.062	0.035	0.023	0.017	-0.062	-0.002
17	0.750	0.465	0.213	-0.004	0.067	0.039	-0.071	0.006	-0.088	0.001
18	-0.737	-1.564	-0.770	0.003	0.068	0.016	0.071	0.084	0.067	-0.009
19	0.658	1.002	1.290	0.008	0.118	0.049	-0.118	-0.006	-0.228	0.005
20	-1.086	-2.216	-1.821	-0.001	0.063	-0.013	0.093	0.177	0.176	-0.014
21	2.401	3.239	2.779	-0.019	0.105	0.087	-0.218	-0.108	-0.458	0.015
22	1.993	2.931	3.994	0.004	-0.008	-0.049	-0.179	-0.223	-0.228	0.018
23	6.391	13.630	10.023	0.146	0.188	-0.010	-1.074	-0.461	-1.005	0.055
24	-0.048	2.323	0.904	0.011	-0.019	0.045	-0.022	-0.017	-0.140	0.002
25	-1.660	-1.374	-2.248	-0.013	-0.034	0.050	0.202	0.144	0.163	-0.015
26	-0.193	0.583	0.078	0.005	0.028	0.048	-0.005	0.016	-0.084	0.000
27	-2.752	-4.106	-6.890	-0.013	0.076	0.064	0.202	0.413	0.368	-0.028
28	0.121	0.290	0.497	0.008	0.014	0.022	-0.015	-0.040	-0.048	0.002
29	-0.538	1.568	0.870	-0.017	-0.071	-0.017	0.061	-0.001	-0.064	0.003
31	2.556	2.693	2.254	0.019	0.188	-0.097	-0.463	0.039	-0.294	0.016
32	-7.002	-3.840	-11.889	-0.007	-0.145	0.115	0.568	0.664	0.735	-0.051
33	-4.390	-5.584	-5.044	0.075	0.207	0.068	0.156	0.403	0.269	-0.029
34	1.090	-0.862	4.668	0.049	0.310	0.118	-0.166	-0.160	-0.528	0.010
35	2.002	2.190	4.505	0.050	0.147	0.094	-0.270	-0.306	-0.497	0.024
36	-0.893	-0.373	-0.523	0.005	0.001	-0.009	0.084	0.060	0.081	-0.007
37	-0.926	-1.149	-0.915	0.000	-0.025	-0.002	0.117	0.054	0.133	-0.008
Average	-0.082	0.228	0.076	-0.001	0.033	0.029	0.005	0.023	-0.048	-0.001

Notes: Estimated coefficient of a translog production function in capital, labor, and electricity following the procedure in section 6.2.1.

Table 6.11: Average Output Elasticities

industry	N	nrobs	Capital	Labor	Energy
15 Food and Beverages	29555	12520	0.03 (0.04)	0.40 (0.07)	0.28 (0.09)
16 Tobacco Products	4197	3435	0.00 (0.02)	0.69 (0.19)	0.26 (0.13)
17 Textiles	15517	3796	0.06 (0.09)	0.37 (0.14)	0.29 (0.15)
18 Wearing Apparel , Fur	14614	3581	0.02 (0.04)	0.41 (0.19)	0.10 (0.04)
19 Leather, leather products and footwear	5036	1691	0.02 (0.06)	0.67 (0.22)	0.27 (0.13)
20 Wood Products (excl. furniture)	7128	2312	0.07 (0.06)	0.15 (0.18)	0.11 (0.06)
21 Paper and paper products	2584	1013	0.17 (0.14)	0.20 (0.49)	0.32 (0.22)
22 Printing and Publishing	3846	740	-0.02 (0.12)	0.52 (0.13)	0.41 (0.24)
23 Coke, refine petroleum products, nuclear fuel	260	140	0.07 (0.53)	0.44 (1.03)	0.69 (0.30)
24 Chemicals and chemical products	9386	1761	0.04 (0.06)	0.53 (0.34)	0.27 (0.11)
25 Rubber and plastic products	9312	3226	0.04 (0.07)	0.20 (0.13)	0.29 (0.12)
26 Non-metallic mineral products	14797	3290	0.02 (0.06)	0.34 (0.12)	0.31 (0.14)
27 Basic metals	1065	503	0.11 (0.20)	0.20 (0.27)	0.31 (0.19)
28 Fabricated metal products	5829	2198	0.02 (0.05)	0.24 (0.06)	0.20 (0.06)
29 Machinery and equipment n.e.c.	3410	1158	-0.04 (0.18)	0.51 (0.21)	0.24 (0.09)
31 Electrical Machinery and apparatus	1095	633	0.04 (0.34)	-0.00 (0.55)	0.32 (0.31)
32 Radio, television and communication equipment	498	336	0.13 (0.13)	0.47 (0.54)	0.33 (0.21)
33 Medical, precision and optical instruments	457	310	0.10 (0.23)	0.03 (0.39)	0.32 (0.19)
34 Motor vehicles, trailers, semi-trailers	934	691	0.04 (0.15)	-0.02 (0.38)	0.47 (0.28)
35 Other Transport Equipment	1693	790	0.03 (0.19)	0.35 (0.30)	0.40 (0.19)
36 Furniture, manufacturing n.e.c	11543	3393	-0.07 (0.07)	0.55 (0.09)	0.06 (0.02)
Average	14142.96	4762.09	0.03	0.39	0.24

Notes: Average output elasticities with respect to inputs from a translog production function in capital, labor, and electricity estimated following the procedure in section 6.2.1.

6.B Additional Results

Table 6.12: Effect of electrification on TFPR by Age Group.

Sample: Firm-Level			
	All	**Young**	**Old**
	(1)	(2)	(3)
Dependent Variable	log(TFPR$_{acf}$)	log(TFPR$_{acf}$)	log(TFPR$_{acf}$)
Panel A: OLS			
Access$_{vpt}$	-0.008	0.0167	-0.0208
	(0.0138)	(0.0174)	(0.0161)
Panel B: IV			
Access$_{vpt}$	0.353**	0.632***	0.214
	(0.138)	(0.002)	(0.152)
First Stage F	43.76	36.81	33.08
Panel C: Reduced Form IV			
Z (KM)	-0.001***	-0.0017***	-0.00057
	(0.00034)	(0.00047)	(0.00038)
Observations	134,391	47,921	86,439
IndustryxYear FE	✓	✓	✓
Province FE	✓	✓	✓
Geo Controls	✓	✓	✓
Firm Controls	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV and Reduced-Form regressions of TFPR on access defined at the desa level. TFPR is measured following Ackerberg, Caves, and Frazer (2015) using electricity spending as a proxy. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Table 6.13: Impact of access on product mix

	All			Incumbents		
	(1)	(2)	(3)	(4)	(5)	(6)
Produced in last 5 years	by firm	in desa	in district	by firm	in desa	in district
Panel A: OLS						
Access_{vpt}	0.00185	0.00386	0.00652	-0.00286	-0.00877	-0.00467
	(0.00554)	(0.00558)	(0.00572)	(0.0183)	(0.0165)	(0.0164)
Panel B: IV						
Access_{vpt}	-0.0999**	-0.121**	-0.100**	0.266**	0.195*	0.182*
	(0.0451)	(0.0493)	(0.0508)	(0.126)	(0.114)	(0.108)
First Stage F	28.68	28.68	28.68	7.53	7.53	7.53
Panel C: Reduced Form IV						
Z (KM)	0.000226**	0.000273***	0.000226**			
	(9.88e-05)	(0.000103)	(0.000111)			
Zx1995				-0.00116***	-0.00107***	-0.00106***
				(0.000279)	(0.000265)	(0.000263)
Zx1996				-0.000281	-0.000175	-0.000156
				(0.000222)	(0.000222)	(0.000220)
Zx1997				0.000718***	0.000805***	0.000817***
				(0.000212)	(0.000220)	(0.000218)
Zx1998				0.00108***	0.00109***	0.00106***
				(0.000300)	(0.000303)	(0.000303)
Zx1999				0.000304	0.000211	0.000137
				(0.000321)	(0.000307)	(0.000304)
Zx2000				0.000368	-0.000372	-0.000364
				(0.000615)	(0.000703)	(0.000664)
Observations	127,818	127,818	127,818	111,029	111,029	111,029
Product FE	✓	✓	✓			
FirmxProduct FE				✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Geo Controls	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓

*** p<0.01, ** p<0.05, * p<0.1

Notes: Results from OLS and IV and Reduced-Form regressions of product mix dummies on access using equation (6.6) in colums (1)-(3) and equation (6.9) in columns (4)-(6). The dependent variable in columns (1) and (4) is a dummy variable equal to one if product j is being produced by firm i in year t and has been produced by firm i any time since $t-5$. The dependent variable in colums (2) and (5) ((3) and (6)) is a dummy equal to one if product j is being produced by firm i in desa v (district d) in year t and has been produced in desa v (district d) any time since $t-5$. Robust standard errors in parentheses clustered at the desa level. Geographic controls are defined at the desa level and include distance to coast, distance to nearest city, elevation, distance to road, desa political and legal status, and infrastructure controls. Firm Controls include export, cohort, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Chapter 7

Conclusion

In this thesis, I show that electrification has a substantial causal impact on the industrial sector. To make this possible, I collected a new and comprehensive dataset on the electrification infrastructure in Indonesia and combined it with rich manufacturing census micro data at various levels. Access to electricity attracts more firms, increases competition, and increases industry productivity by weeding out the less productive firms more often. Access to electricity can therefore help narrowing the productivity gap between firms in developing and developed countries. Electrification also changes the composition of firms: firms in electrified areas are larger, more efficient, and sell more differentiated products that were not produced before. All this evidence suggests that electrification is welfare improving.

I highlight a new economic mechanism through which electrification causes industrial development. This mechanism, firm turnover, is unlikely to operate in response to short-run improvements in electricity supply. The extensive margin of electrification induces extensive margin responses in firm decisions, which affects the composition of firms in the industry. Electrification attracts more firms into a market. This creates more competition and makes it more difficult for unproductive firms to survive. By increasing firm turnover, electrification increases average productivity in the market. This mechanism is similar to selection induced by trade liberalization where exposing domestic firms to international competition forces the least productive firms to exit as in Pavcnik (2002) and Melitz (2003). Electrification therefore promotes industrial development by increasing the efficiency with which markets allocate resources from unproductive firms towards more productive firms.

While the infrastructure literature has made substantial progress in understanding the effect of transportation (roads, railways) on development, we are at the very beginning of understanding how access to energy affects economic development. This thesis has taken a small step towards a better understanding of the relationship between energy infrastructure and development. However, there is still a lot to be learned. Electrification projects are typically large-scale costly investments and it is important to quantify their benefits. In some instances, like in Lee, Miguel, and Wolfram (2016) and Burlig and Preonas (2016), benefits from electrification do not necessarily justify the investment and are not as large as we expect them to be. Large investments in electrification have been made in various African countries over the last decades, but Africa is yet to industrialize.

It is therefore important to understand how electrification and other institutional features might interact. For instance, other large institutional barriers to entry or to market access might prevent electrification from triggering entry and allowing for productivity gains. In the presence of credit constraints, the effect of electrification could be even larger, because it can lower the cost of entry for constrained entrepreneurs and reduce the extent of misallocation. These are a few of the open questions that remain to be answered in future work on electrification and development.

Once we have a better understanding of how and when access to energy leads to growth, it is then important to think about how we can provide energy and use it to grow the economy without harming the environment. Energy is potentially essential to bring people out of poverty, but it is also important to provide it in a cheap and sustainable way. This provides us with a new set of challenges and research opportunities that we have not thought about previously in the experience of electrification and industrialization in the developed world.

Bibliography

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Allcott, H., A. Collard-Wexler, and S. D. O'Connell (2016). How do electricity shortages affect industry? evidence from india. *American Economic Review 106*(3), 587–624.

Atkeson, A. and P. J. Kehoe (2007). Modeling the transition to a new economy: Lessons from two technological revolutions. *The American Economic Review 97*(1), 64–88.

Banerjee, A., E. Duflo, and N. Qian (2012). On the road: Access to transportation infrastructure and economic growth in china. Working Paper 17897, National Bureau of Economic Research.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013). Cross-country differences in productivity: The role of allocation and selection. *The American Economic Review 103*(1), 305–334.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics 122*(2), 775–805.

Bloom, N., A. Mahajan, D. McKenzie, and J. Roberts (2010, May). Why do firms in developing countries have low productivity? *American Economic Review 100*(2), 619–23.

Burlig, F. and L. Preonas (2016). Out of the darkness and into the light? development effects of rural electrification. *Working Paper.*

Chandra, A. and E. Thompson (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics 30*(4), 457–490.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica 80*(6), 2543–2594.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform. *Econometrica 84*(2), 445–510.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from south africa. *American Economic Review 101*(7), 3078–3108.

Donaldson, D. (2010). Railroads of the raj: Estimating the impact of transportation infrastructure. *Working Paper*.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics 131*(2), 799–858.

Duflo, E. and R. Pande (2007). Dams. *122*(2), 601–646.

Duranton, G., P. M. Morrow, and M. A. Turner (2014). Roads and trade: Evidence from the us. *The Review of Economic Studies 81*(2), 681–724.

Duranton, G. and M. A. Turner (2012). Urban growth and transportation. *Review of Economic Studies 79*(4), 1407–1440.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *The Review of Economic Studies*.

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

Foster, L., J. Haltiwanger, and C. Syverson (2008, March). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425.

Gertler, P. J., M. Gonzalez-Navarro, T. Gracner, and A. D. Rothenberg (2014). The role of road quality investments on economic activity and welfare: Evidence from indonesia's highways. *Unpublished manuscript*.

Griliches, Z. and J. Mairesse (1995). Production functions: the search for identification.

Hoch, I. (1962). Estimation of production function parameters combining time-series and cross-section data. *Econometrica 30*(1), 34–53.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Lee, K., E. Miguel, and C. Wolfram (2016). Experimental evidence on the demand for and costs of rural electrification. *Working Paper*.

Lenin, V. I. (1920). Collected works. *Moscow: Progress Publishers XXXI*, 513–518.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Lipscomb, M., A. M. Mobarak, and T. Barham (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, Q. and L. D. Qiu (2016). Intermediate input imports and innovations: Evidence from chinese firms' patent filings. *Journal of International Economics 103*, 166–183.

McCawley, P. (1978). Rural electrification in indonesiaâĂŞis it time? *Bulletin of Indonesian Economic Studies 14*(2), 34–69.

McCawley, P. a. (1971). *The Indonesian electric supply industry*. Ph. D. thesis, The Australian National University.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *The review of economic studies 75*(1), 295–316.

Mundlak, Y. (1961). Empirical production function free of management bias. *American Journal of Agricultural Economics 43*(1), 44–56.

Nguyen, K. T. (2014). Privatization and productivity of upstream industries: Firm and industry level evidence from vietnam. *Working Paper*.

Olley, G. S. and A. Pakes (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.

Redding, S. J. and M. A. Turner (2014). Transportation costs and the spatial organization of economic activity. Working Paper 20235, National Bureau of Economic Research.

Reinikka, R. and J. Svensson (1999). How inadequate provision of public infrastructure and services affects private investment. Policy research working paper series, The World Bank.

Rud, J. P. (2012). Electricity provision and industrial development: Evidence from india. *Journal of Development Economics 97*(2), 352 – 367.

Ryan, N. (2017). The competitive effects of transmission infrastructure in the indian electricity market. *National Bureau of Economic Research.*

Ryan, N. (2018, May). Energy productivity and energy demand: Experimental evidence from indian manufacturing plants. *National Bureau of Economic Research* (24619).

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy 112*(6), 1181–1222.

Syverson, C. (2007). Prices, spatial competition and heterogeneous producers: an empirical test. *The Journal of Industrial Economics 55*(2), 197–222.

Tybout, J. R. (2000, March). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature 38*(1), 11–44.

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