Does Electrification Cause Industrial Development? Grid Expansion and Firm Turnover in Indonesia

Dana Kassem*

July 2019

Abstract

I ask whether electrification causes industrial development. I combine newly digitized data from the Indonesian state electricity company with rich manufacturing census data. To understand when and how electrification can cause industrial development, I shed light on an important economic mechanism - firm turnover. In particular, I study the effect of the extensive margin of electrification (grid expansion) on the extensive margin of industrial development (firm entry and exit). To deal with endogenous grid placement, I use an instrumental variable approach exploiting the location of colonial electric infrastructure and the need for an interconnected grid in the island of Java. I find that electrification causes industrial development, represented by an increase in the number of manufacturing firms, manufacturing workers, and manufacturing output. Electrification increases firm entry rates, but also exit rates. Empirical tests show that electrification creates new industrial activity, as opposed to only reorganizing industrial activity across space. 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. Therefore, without the possibility of entry or competitive effects of entry, the effects of electrification are likely to be smaller.

(*JEL* D24, O13, O14, O18, Q41, R11, R12)

* Department of Economics, University of Mannheim, d.kassem@uni-mannheim.de. I thank Robin Burgess and Oriana Bandiera for helpful advice and continuous support. I also thank Gharad Bryan, Michel Azulai, Matteo Benetton, Thomas Drechsel, Greg Fischer, Alessandro Gavazza, Maitreesh Ghatak, Hanwei Huang, Rachael Meager, Panos Mavrokonstantis, Marco Gonzalez-Navarro, Kieu-Trang Nguyen, Michael Peters, Alex Rothenberg, Mark Schankerman, Pasquale Schiraldi, Arthur Seibold, John Sutton, Junichi Yamasaki, and various seminar and conference participants. I am grateful for the support of Suroso Isnandar, Muhammad Ikbal Nur, Musa Partahi Marbun, Ahmed Yusuf Salile and various others at the Indonesian State Electricity Company Perusahaan Listrik Negara (PLN). Financial support from the IGC, STICERD, and the DFG-CRC TR 224 is gratefully acknowledged.

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 in rural electrification, with the expectation that this investment will "enhance industrialization and emergence of [...] industries[3]". There is consensus among policymakers that access to electricity is an essential ingredient for industrial development, which is considered a fundamental driver of growth. Despite the policy relevance of this issue, there is surprisingly little causal evidence on whether grid electrification truly drives industrial development or whether this relationship is a correlation along the path of development.

To answer this question, I use a rapid, government-led grid expansion during a period of rapid industrialization in Indonesia. I 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 using firm-level manufacturing census data from Java, where 80% of Indonesian manufacturing firms are located. These data allow me to understand *when* and *how* electrification affects industrial development.

The answer to this question is ex-ante ambiguous. On the one hand, electrification can cause industrial development by attracting new firms into the market, for example by lowering barriers to entry. On the other hand, electrification could be a *white elephant*; a costly investment but nothing happens when the grid arrives. Recent economic evidence shows that the benefits of *rural* electrification are generally not as large as previously thought[4]. Electrification in various Sub-Saharan African countries has increased substantially over the last decades, but these countries have not witnessed industrial development. So I ask, does electrification cause industrial development? Or do these investments have little impact on the pace of industrial development?

This paper 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

[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] Kenya's Rural Electrification Authority 2017

[4] Examples include Lee, Miguel, and Wolfram (2019) and Lenz, Munyehirwe, Peters, and Sievert (2017) who focus on residential electrification, Bos, Chaplin, and Mamun (2018) who provide a review, and Burlig and Preonas (2016).

grid to new locations, has been studied on employment (Dinkelman (2011)) and general development-level indexes (Lipscomb, Mobarak, and Barham (2013)). 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. Fisher-Vanden, Mansur, and Wang (2015), 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 electrification enhances or decreases manufacturing productivity is 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 and similar institutional characteristics[5]. In 1990, Java, the most developed and densely populated island in Indonesia, was only around 40% electrified[6]. 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%[7]. 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 tries to make progress on this issue by using an instrumental variable approach inspired by the transportation literature and closely related to Faber (2014). I exploit a supply-side natural experiment based on the need of

[5] McCawley (1978)
[6] Statistik PLN 1989/1990, Indonesian Electricity Company.
[7] Statistik PLN 2000, Indonesian Electricity Company.

the state electricity monopoly to have a single interconnected electricity grid in Java. I construct a hypothetical electric transmission interconnection that is a function of incumbent disconnected colonial power plants and construction cost. To underpin instrument credibility, I construct two separate hypothetical interconnections, differing in the way cost is defined. The first hypothetical interconnection takes cost of construction to be a simple linear function of land elevation, a local geography factor. The second hypothetical interconnection takes cost to be the Euclidean distance between incumbent power plants, ignoring local geography completely. I apply Kruskal (1956)'s minimum spanning tree algorithm to finds the combination that minimizes the global construction cost of the transmission interconnection.

The hypothetical interconnections abstract from endogenous demand factors that could be driving the expansion of the grid and focus on cost factors only. The result is two separate instruments, defined as the distances to each of the hypothetical interconnections, used to instrument for endogenous access to electricity. This strategy allows for variation in the instruments while controlling fully for local geography, and more broadly, local determinants of the cost of connection. The identification assumption is that conditional on various controls, including local geography, baseline industrial outcomes and other types of infrastructure, the instruments only affect industrial outcomes through their effect on access to the grid.

The data-sets used in this paper come from various sources. I collected and digitized 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 manufacturing firms, manufacturing output, number of manufacturing workers, and entry and exit rates in any desa in Java. The information on manufacturing activity comes from 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. These variables will allow me to look at the effect of electrification on different measures of firm performance and the composition of firms in the market. I then combine productivity estimates with market share information to study the effect of electrification on reallocation at an aggregate industry level.

The results in this paper show that electrification causes industrial development at a local level by increasing manufacturing activity. Access to the grid increases the number

of manufacturing firms, number of workers in manufacturing, and manufacturing output. Interestingly, electrification increases firm turnover by increasing not only entry rates, but also exit rate. Both the geographic instrument and the Euclidean instrument yield similar estimates, suggesting that when accounting carefully for potential confounders, geographic instruments can be credible.

A second empirical challenge is related to 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. In the context of this paper, this occurs if electrifying one location affects the industrial outcomes of neighboring locations. I carefully examine the extent of spillovers by conducting various empirical tests. Results show that the extent to which electrification causes spillovers is minimal relative to the direct effect of electrification. The evidence for spillovers is mixed, but even the largest estimates of having a connected neighbor on industrial outcomes do not explain the total effect of electrification. Taken together, electrification in Java created new industrial activity.

Finally, electrification causes average firm size to increase, both in terms of how much output the firm produces and how much inputs it demands. Also, electrification increases average productivity, consistent with higher firm turnover and a change in the composition of firms in the market. 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.

Much like various Sub-Saharan African countries today, Indonesia in the 1990s suffered from weak credit institutions, poor infrastructure such as primitive roads in rural areas causing high transportation costs, and favoritism was widely prevalent under the Suharto regime. Needless to say, there are many differences between Java and Sub-Saharan Africa, but the results in this paper indicate that electrification occurring in a weak institutional environment can indeed cause industrial development.

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 (2018), 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 de-

velopment. Ryan (2017) studies the effect of expanding the transmission infrastructure on the competitiveness of 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 on county-level development. Rud (2012) shows that industrial output in Indian states 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. Abeberese (2017) shows that Indian manufacturing firms, when facing higher energy prices, switch to less electricity-intensive production technologies and produce less output.

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) and Hsieh and Klenow (2009), 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 by linking infrastructure to reallocation and turnover in explaining the low productivity of firms in developing countries.

Section 2 presents the institutional background of electrification in Indonesia, summarizing the history of the Indonesian power sector and the objective of the Indonesian government during the period of the study. Section 3 introduces the new data on the Indonesian electrification infrastructure and presents the empirical strategy. Section 4 presents evidence on the effect of electrification on local industrial outcomes and investigates how electrification affects the organization of industrial activity across space. I evaluate how electrification affects firm performance in Section 5. In Section 6, I exam-

ine the implications of electrification on industry productivity and reallocation. Finally, Section 7 concludes.

2 Institutional Background

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. Political unrest, lack of funds, hyperinflation and the lack of qualified management and engineers lead to a period of decline in efficiency, poor operating conditions, and inadequate expansion (McCawley (1971)). This in turn lead 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 and institutional environment. To put things into perspective, in 1975, Indonesian GDP per capita was around $216, higher than the GDP per capita in India of $162[8]. 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 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

[8]Source: World Bank.

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.)[9].

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[10]. 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 lead 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 1 presents the dramatic increase in electrification ratios in Java during the sample period. Figure 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 1b), 1996 (Figure 1c), and finally in the year 2000 (Figure 1d), when most of Java was fully electrified.

3 Data and Empirical Strategy

3.1 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 full electrification as can be seen in Figure 1.

[9]Source: IEA Statistics 2014

[10]Source: Official planning documents.

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[11]. With a considerable amount of time and resources, I collected and digitized 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), and (ii) maps (digital, for example Figure 3, and paper maps, Figure 4 and Figure 5) 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 6 shows the expansion of the grid during the sample period where the yellow bolts represent transmission substations.

The expansion of the transmission grid in Java during that period was rapid and substantial as shown by the summary statistics in Table 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.

3.2 Industrial Outcomes

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[12]. Data on desa level boundaries were acquired from 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 the contiguous land of Java[13], 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.

[11] Source: author's calculations.

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

[13] I exclude the outer islands that are administratively part of Java because the identification strategy proposed later does not apply to these locations.

Table 2 presents desa-level summary statistics. 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 that 1. However, the median is 0. To capture the extensive margin of industrialization, and to avoid sample selection, I include all the desas in Java in the sample regardless of whether the desa 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 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 $24,000$.

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 addition, I use GIS data on land elevation, cities, waterways, coastline and roads in Java. I measure the Euclidean distance from each desa (centroid) to each of these geographic features and to the nearest electric substation and the hypothetical least cost interconnection. 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 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.

3.3 Empirical Strategy

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:

$$Y_{vpt} = \alpha + \beta Access_{vpt} + \boldsymbol{V'}_{vpt}\boldsymbol{\eta} + \gamma_{pt} + \epsilon_{vpt} \tag{1}$$

where $Access_{vpt}$, is an indicator variable equal to one if the desa is within 15 KM[14] of the nearest transmission substation in year t. $\mathbf{V}_{vpst}$ is a vector if desa level controls, which I list in detail in Section 3.3.3, and γ_{pt} are province-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 Indonesia, 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 existing supply. PLN then decides to expand its infrastructure if they believe there will be a gap between supply and demand in the future. I explain this methodology in detail in Appendix E. 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 biased. 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 biased. Another element in the decision of expanding the grid is cost of construction, which potentially creates exogenous variation that is useful for identification.

In order to deal with the endogeneity problem, I propose an instrumental variable approach exploiting a supply-side natural experiment. In 1969, electricity grid in Java consisted of 5 different disconnected grids across the island (Figure 4). 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 5). Various transmission lines were built for the main purpose of interconnecting the grids. 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.

Figure 7 illustrates the empirical strategy in a simplified manner. Consider two disconnected grids Grid 1 and Grid 2. These represent the incumbent infrastructure built by Dutch electricity companies and existed by 1969. During the 1970s and the 1980s, the two grids became interconnected by the green line. Consider two desas A and B that only differ in their distance to the green line. Because desa A is closer to the green line, it is then more likely to get connected to the electricity grid in the 1990s compared to desa B. The blue lines represent the instrument. Because of concerns regarding the placement of the green line, for example transmission lines could be targeted at areas that are different than others, such as non-farming land, I create a hypothetical version of the green line. The hypothetical version, is based solely on cost factors, taking as given the location of

[14]This threshold was chosen based on conversations with electrical engineers at the Indonesian state electricity monopoly and to match average electrification statistics from the period.

incumbent infrastructure. In total, the incumbent infrastructure I consider consists of 15 power plants which I identify from historical maps as the main power plants in the 5 separate grids, built by colonial Dutch utilities[15]. The hypothetical interconnection is essentially an instrument for the actual grid transmission network (which I do not observe) that abstracts from demand factor driving the location of the actual grid and focuses on cost. I consider two alternative definitions of cost, resulting in two different versions of the hypothetical interconnection, and therefore two instruments. I describe below how these instruments are constructed.

3.3.1 Geographic Least Cost Interconnection

The hypothetical geographic least cost interconnection, takes local geography[16], elevation in particular, in its cost function. Elevation is an important cost factor in grid construction as higher terrains require higher transmission towers and generally more expensive transmission equipment[17]. In addition, transmission towers in mountain ridges involve expensive investments in electrical protection equipment, for example, against lightning[18]. I implement the following procedure to construct the hypothetical geographic least cost interconnection:

1. For each location on the map, I assign a cost value based on a digital elevation model[19].

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

3. I use Kruskal's algorithm[20] to find the least cost combination of least cost paths such that all power plants are interconnected. The resulting network is the hypothetical geographic least cost transmission interconnection.

Figure 8 shows the resulting hypothetical geographic least cost interconnection. The straight-line distance to the hypothetical geographic interconnection from the centroid of a desa, which I label G_v and refer to as the geographic instrument, is then used as an instrument for access to the grid.

[15] See Figure 4. List of power plants: Banyuwangi (Diesel), Cilicap (Diesel), Cirebon (Diesel), Jatiluhur (Hydro), Jelok (Hydro), Karangkates (Hydro), Ketenger (Hydro), Kracak (Hydro), Lamajan (Hydro), Madiun (Diesel), Perak (Diesel), Semarang (Thermal), Suralaya (Diesel), Tanjung Priok (Thermal), and Ubrug (Hydro).

[16] As discussed later, I control for local geography in the main specification to address the issue that these can affect industrial outcomes directly

[17] Such as more tower units, and more dead-end towers.

[18] Slope is also an important cost factor in the cost of transmission network, and potentially a more important one than elevation. I choose not to include slope in the the least cost function since it is shown in the transport literature that slope is an important cost for road and railway construction, and therefore a threat to exclusion.

[19] Cost is simply equal to the value of land elevation

[20] 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.

3.3.2 Euclidean Least Cost Interconnection

An alternative to a geography-based least cost interconnection would be a straight-line version of the hypothetical interconnection where cost is defined as the Euclidean distance. I therefore apply the same algorithm above but instead of using elevation, I use Euclidean distance as a measure of cost, ignoring local geography completely. Kruskal's algorithm will then find the least *distance* combination of straight lines such that all power plants are interconnected. Figure 9 shows the resulting hypothetical Euclidean least cost interconnection. The distance to the straight-line least cost interconnection, labeled E_v, is then used as a second instrumental variable.

3.3.3 Controls

To ensure that desas A and B in Figure 7 only differ in their distance to the hypothetical least cost interconnection, I control for various desa-level characteristics. One concern is that the location of the incumbent power plants is endogenous. To alleviate this concern, I exclude desas within a $10KM$ radius of power plants. Power plants are built close to the consumption centers to which they supply electricity in order to minimize transmission losses. Because consumption centers are typically cities, I include distance to nearest city[21] as a control variable to deal with the concern that the instrument could be correlated with the distance to the city closest to the power plant.

Controlling for local geography, especially with the geography-based instrument G_v, is essential. This is to ensure that it is not local geography that is driving the economic outcomes directly, instead, through the instrument's effect on electrification, i.e. to guarantee exclusion. Therefore, controlling for desa elevation is necessary.

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 of Sumatra and Kalimantan, which are easily reachable from the north coast because of proximity and good wave conditions in the Java sea. It is then important to control for distance to coast in any empirical specification to avoid any threats to exclusion.

Typically, different types of infrastructure are correlated, for example, the electricity grid and the road network. In all my specification, I control for the distance to the nearest regional road. I also control for the availability of other non-energy infrastructure facilities. These include railway station, motor station, river pier, sea port, and airport. For political economy concerns that could be correlated with electrification, I control for political and legal status of the desa. Political status is an indicator variable equal to one if the desa is governed by an elected official, and zero if governed by appointed civil

[21] More precisely, the distance to the city that is closest to the nearest power plant.

servant. Legal status of the desa refers to whether the desa is formed under a Government decree, Ministerial decree, Regency decree, or other.

A concern is that some unobserved persistent shocks to industrial desa-level outcomes that are also correlated with distance to the hypothetical least cost interconnections are driving the estimated effects. Therefore, I control for the baseline value of the dependent variable in each regression, where 1990 is taken to be the baseline year and is excluded from the analysis.[22]

Some of these controls might threaten the exclusion of the instrument not only in a static sense, but also dynamically. For example, locations with different elevations or distance to nearest city or nearest road, and with higher baseline outcomes, can have different industrial growth trends, which then affects the correlation of the instrument with access. I therefore flexibly control for these effects by including an interaction of each of these four variables with year effects.

To summarize, geographic desa controls include elevation-by-year effects, distance to coast, distance to nearest city-by-year effects, and distance to nearest road-by-year effects. Other desa level controls include baseline outcome-by-year effects, various infrastructure availability dummies, political status, and legal status. Table 4 presents a summary. The identification assumption is that, conditional on these controls, the distance to the hypothetical geographic least cost interconnection in not correlated with the potential outcomes of desas.

3.3.4 Geographic Instrument Variation and Controls

Given that the geographic instrument G_v is based on geography, what variation is left in the distance to the hypothetical interconnection 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 geographic least cost interconnection? The answer is because what matters for the hypothetical geographic least cost interconnection is *global* geography, not local geography. This is because the algorithm described above has the objective of minimizing the cost of building the transmission interconnection, 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 lowest areas because in some locations choosing a path at higher elevation might lead to a lower path further ahead on route to the next power plant. This creates variation in the distance to the hypothetical interconnection for locations with the same local geographic characteristics.

[22] The final desa sample is from 1991-2000.

3.3.5 First Stage

Table 5 shows the first stage regressions using separately the distances to the hypothetical interconnections G_v and E_v as an instrumental variable. The dependent variable is $Access_{vpt}$, the indicator variable for electrification. Standard errors are clustered at the desa level; the level at which the instrument varies.

The coefficient in Column (1) is negative and significant, indicating that the further away a desa is from the hypothetical geographic interconnection, the less likely it is to have access to electricity. Adding all the geographic controls discussed above lowers the estimate in Column (2), but the relationship remains strong and significant. Column (3) shows the first stage regression of access on the Euclidean instrument E_v. Reassuringly, E_v is significantly negatively correlated with access, suggesting that the location of the incumbent power plants is a key driver of the strong first stage regression in the first two columns. The estimate changes only slightly after adding the geographic controls in Column (4). This is in contrast to the geographic instrument that relies more on local geography. The first stage F-statistic is high enough to guarantee relevance of both instruments, avoiding weak instrument bias. When adding geographic controls, the first stage F-statistic drops proportionately more (comparing the move from Column (1) to (2) to Column (3) to (4)) for the geographic instrument G_v relative to the Euclidean instrument E_v. This observation illustrates the heavier dependence of G_v on geography relative to E_v.

3.3.6 Geographic Instrument Validity

In this section, I discuss the validity of the geographic instrument G_v and present a placebo test. I create a placebo geographic least cost interconnection that connects 15 random points in Java using the same least cost algorithm as the one used in the main instrument (Figure 8). If access to the grid is correlated with the distance to this least cost placebo interconnection, it would mean that local geography, irrespective of the location of the actual electric transmission interconnection, is what is driving the correlation between access and the instrument. Figure 10 illustrates the placebo hypothetical least cost interconnection. The origin points to be connected by the algorithm were randomly chosen by the computer. The same algorithm applied to create the geographic least cost interconnection using the main incumbent power plants was applied to connect these randomly generated points on a single network.

Columns (5) and (6) in Table 5 present the results from the first stage regressions of access on the placebo instrument. While the coefficient is significant in Column (5) without the controls, it is very close to zero. The significance completely disappears when adding the controls in Column (6) and there is no correlation between access to the grid and the distance to the placebo interconnection. The first stage F statistic in Column

(5) is almost at the acceptable conventional level of 10, but it becomes close to zero in Column (6) when the relevant controls are added. This illustrates how failing to control for local geography with geographic instruments poses serious threats to exclusion. The fact that access and distance to the placebo interconnection, conditional on controls, are not correlated, alleviates the concern that correlation between access and the the geographic instrument is purely driven by geography. The incumbent infrastructure and the global geography of the island are therefore what determines the correlation between access and G_v, conditional on controls.

4 Impact on Industrialization

I start by examining the effect of electrification on industrial outcomes at the local level. In order to understand the mechanisms through which electrification affects local industry, I then look at how firm turnover, as measured by the entry and exit rates of firms, is affected by electrification. 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. By focusing on the extensive margin of electrification in this section, the aim is therefore to see whether electrification has any effect on the extensive margin of industrialization.

4.1 Effect of Electrification on Local Industry

I estimate the effect of the arrival of the grid on the number of manufacturing firms, manufacturing employment and manufacturing output at the desa level. To capture the extensive margin of industrialization, and because there are many desas with zero firms, I use the level of these variables instead of the log (See Table C1 in appendix C for results with zero-preserving log transformations.). Table 6 shows the IV and OLS regression results for three desa-level outcomes as in specification (1), with and without the desa-level controls to assess estimates stability.

Panel A and Panel B present the results using the two instruments G_v and E_v, respectively. The IV estimates are positive and significant, and reassuringly, of similar magnitudes across the two instruments. The coefficient in Column (2) in Panel A says that the causal effect of grid access on the number of firms in a desa is an increase of 0.65 firm. Considering that the average number of firms per desa in the sample is 0.91, this effect is large and around 70% of the average (bottom Panel). Similarly for the number of workers and manufacturing output, the IV estimates in columns (3)-(4) and (5)-(6) are positive, large and economically 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 caveat, 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. 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 A and B columns (3) to (6) 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.

Across all outcome variables, the OLS estimates in Panel C are positive and significant, suggesting that there is a positive correlation between access to electricity and industrial outcomes. Compared to the IV estimates in Panels A and B, the OLS estimates are 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 (e.g. Michaels (2008), Atack and Margo (2011), Duranton and Turner (2012)) indicating that infrastructure is allocated to less developed areas. This means that the OLS estimates will underestimate the effect of electrification on manufacturing, as the results show.

While the direction of the OLS bias is common in the infrastructure literature, the difference in magnitudes between the IV estimates and the OLS estimate is rather large, and calls for a discussion. I present and discuss four potential reasons for the magnitude of this difference below.

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. With the geographic instrument G_v, this means that the distance to the hypothetical interconnection, 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 interconnection 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 second section of this paper. In addition to geographic and infrastructure controls, I control for other political and economic characteristics. Finally, the fact that the results using the Euclidean instrument, which relies much less on geography, are similar to the IV estimates with the geographic instrument, and the results from Section3.3.5 with the placebo interconnection, show that local geography does not drive the correlation between access and the distance to the geographic least

cost interconnection[23]. Given this rich set of controls and the evidence from the various empirical tests presented thus far, it is unlikely that a violation of the exclusion restriction is driving the difference in magnitudes between the IV and OLS estimates.

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[24].

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 are potentially 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 Table C.2 in the appendix, where I use a more accurate definition of access at the firm-level, show a large difference between IV and OLS estimates. This indicates 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 6, 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 effects of access on the number of workers in manufacturing and manufacturing output do not appear so large. Compare the estimated IV coefficients for these variables in Panels A and B Columns (4) and (6) to the average outcomes in a desa with a positive number of firms. The effect on man-

[23]In Section 4.4, I estimated a difference-in-difference specification at the desa-level, which time-invariant desa characteristics, and results are still positive and significant.

[24]Source: author's calculation.

ufacturing labor is around 30% of the effect of moving from a non-industrialized desa to an industrialized desa. The effect on manufacturing output is around 60% of the average output in industrial desa. The estimated effect of electrification is therefore comparable to partially moving from a non-industrialized desa to an average industrialized desa, with economically sensible magnitudes.

4.2 Effect of Electrification on 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 4.1, electrification causes the total number of firms in a desa to increase. I take advantage of the richness of the data and investigate the effect of electrification on the extensive margin of industrialization with the goal of uncovering the economic mechanism through which this increase in industrial outcomes is occurring. One such economic mechanism is firm turnover. Firm turnover is producer-level churning represented by the entry and exit rates of firms. Churning is a healthy sign in the market where more productive businesses replace less productive ones. In appendix D, I present a theoretical model of how electrification can affect firm turnover. I now turn to the role of entry and exit as drivers of industrialization.

Table 7 presents the estimates of effect of electrification on the number of entrants, number of exiting firms, and firm turnover represented by the entry and exit rates. Focusing on the IV estimates in Panel A, columns (1) and (2) show that electrification increases the average number of entrants in electrified desas relative to the number of entrants in non-electrified desas. For completeness, columns (3) and (4) show the same estimates for the effect on the number of exiting firms, which are also positive and significant[25]. Columns (5)-(6) and (7)-(8) 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[26]. Panel B presents the results from the same regressions but using the Euclidean IV as an instrument. The results are again similar in magnitudes using both instruments. The IV estimates in Panels A and B columns (5) and (6) show that access to the grid increases firm entry rate by around 3 to 6 percentage points. Interestingly, in Columns (7) and (8), the coefficient on access shows that the exit rate *also* increases due to electrification, although by a smaller amount (1 to 2 percentage points). This is consistent with the increase in the total number of manufacturing firms from columns (1) and (2) in Table 6 in the previous section. Electrification therefore increases firm turnover, leading to more

[25] Of course, this relation is mechanical given that electrification increases the number of firms as shown in the previous section.

[26] This is the sample of industrialized desas corrected for entry and exit.

churning in a given desa. Higher churning is a sign of efficiency where firm selection into and out of the desa is at work. As before, the OLS estimates in Panel C are positive and smaller in magnitude than the IV estimates in Panels A and B, and are therefore downward biased.

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.

4.3 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, or firms that would have entered anyway, but now they choose to enter in 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 smaller or negligible than the estimated effects at the desa level.

Put differently, the stable unit treatment value assumption (SUTVA) might not be satisfied in the identification strategy in this analysis. SUTVA requires that the treatment applied to one unit does not affect the outcome for another unit. Electrifying one desa can have an effect on industrial activity in other desas. 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 changes the interpretation of the results. 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 this section, I investigate the possibility of electrification creating new economic activity versus relocating economic activity. I start by looking at the possibility of firm relocation.

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 it 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.

In the context of Java, even if a desa is far from the grid at a certain point in time, it eventually got connected to the grid. Given that this is a period of rapid expansion of the grid in Java, eventually all desas became connected to it. So unless a firm is really impatient, the benefit of moving to an electrified desa today versus waiting to get access in a few years is unlikely to be a profitable action. Confirming this insight, I observe no firm movements across desas in the dataset[27][28].

To estimate spillovers, I perform three empirical exercises. Given the argument made above and the rapid grid expansion, spillovers are likely to be stronger 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 (1) at the district[29]-level, a higher administrative division than a desa[30]:

$$Y_{dt} = \alpha + \beta Access_{dt} + \boldsymbol{V'}_{dt}\boldsymbol{\eta} + \gamma_{pt} + \epsilon_{dt} \tag{2}$$

where $Access_{dt}$ is equal to one if at least 50% of desas in a district have access to electricity. I include the same controls as before in vector $\mathbf{V}_{dt}$, all defined as the mean. I also include the baseline outcome calculated at the district level in 1990. $Access_{dt}$ is instrumented with the average distance to the hypothetical interconnections in the district G_{dt} and E_{dt}. I exclude all districts that include at least one desa that is within $10KM$ of an incumbent power plant.

If spillovers are significant, then theoretically the effect of electrification on manufacturing outcomes should be smaller at the district-level[31]. For comparability with the desa-level results in Table 6, I use the average number of firms, average number of manufacturing workers, average manufacturing output, and average number of entrants and exiting firms in a district across desas as opposed to the total on the left hand side.

Table 8 presents the results. Comparing the IV estimates of the effect of electrification on the number of firms in Columns (1) and (2) in Panels A and B to their counterparts in Table 6, the estimates at the district level are larger, meaning that relocation of firms

[27]Less that 1% of the firms change desas between 1990-2000. I exclude these firms from the analysis.

[28]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 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.

[29]Kecamatan in Bahasa

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

[31]It's not clear how spillovers would affect the entry and exit rates, at least not without an explicit theoretical model, as the net effect will depend on the effect on the number of entrants and exiting firms in the numerator relative to the effect on the total number of firms in the denominator

within district is unlikely. Compared to the desa-level estimates of the effect on the number of entrants and exiting firms in Table 7 in Columns (1) to (4), Columns (3) to (6) show that the effect of access on these industrial outcomes at the district level is very close to the effect at the desa-level. Together, the results in the first 6 columns of Panels A and B suggest that extensive margin spillovers are unlikely. Moving to the effect on district-level manufacturing labor, the estimated effect in Column (7), Table 6 Columns (3), appears to be smaller. This difference however disappears once the district controls are added in Column (8) relative to the estimate in Table 6 Columns (4). As for manufacturing output in the last 2 Columns of Table 8, although quite comparable to the estimates in Columns (5) and (6) of Table 6, the estimates are smaller. This suggests that there might be some intensive margin spillovers where there are negative spillovers of electrification on output locally. The difference in estimates at the district and desa level are most apparent for manufacturing output, which is not surprising as products are easier to move than workers and firms.

Second, I repeat the desa-level analysis from Equation (1) but jointly estimating the main effect of access $Access_{vpt}$ and the spillover effect N^C_{vpt}, conditional on the total number of neighboring desas N_{vp}.

$$Y_{vpt} = \alpha + \beta Access_{vpt} + \mu N^C_{vpt} + \theta N_{vp} + \boldsymbol{V'}_{vpt}\boldsymbol{\eta} + \gamma_{pt} + \epsilon_{vpt} \tag{3}$$

N^C_{vpt} is defined as the number of connected neighboring desas, defined to be within a 7 kilometer radius measured between the centroids of the desas[32]. On average, the total number of neighbors is 43.5 desas and the number of connected neighbors is 29.28 desas. The coefficient on N^C_{vpt} will therefore measure the effect of having an additional electrified neighboring desa on desa outcome Y_{vpt}. A negative estimate of μ would mean that there are negative spillover effects of having connected neighboring desas. If

$$\hat{\beta} + \hat{\mu} * \bar{N^C_{vpt}} = 0$$

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. I instrument access with the desa's own distance to the hypothetical interconnection, and the number of connected neighbors by the average distance of neighbors to the hypothetical interconnection[33], both interacted with time dummies to aid with power.

Table 9 presents the IV and OLS results of Equation (3). Focusing on the IV results

[32]The results are similar when using a smaller or larger radius and are available upon request.

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

in Panels A and B, the estimated coefficients in the preferred specification with all the controls across all industrial outcomes are comparable to the IV results in Table 6 and Table 7 columns (1) to (4). 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 mixed. Starting with the number of firms, the result in Column (2) with the geographic IV in Panel A indicates that there are small negative spillovers of the neighbors access on the number of firms. However, when evaluated at the mean number of connected neighbors, this effect is small and does not account for the total effect of access on the number of firms. This result disappears with the Euclidean IV in panel B with the controls. As for workers in manufacturing, there is no evidence for negative spillovers in any of the specifications. Moving to manufacturing output, Panel A Column (6) shows that there are negative spillovers in the output market that account for less than 50% of the total effect of access on output. This indicates that spillovers are stronger in the output market, consistent with high relocation costs of firms and workers. This effect however disappears with the Euclidean IV in Panel B. The only effect that is significant across Panels A and B is the spillover on the number of entrants. Since the relocation of entrants is cheaper than the relocation of incumbents, it is possible that the spillovers are stronger for this subset of firms. While the estimates of spillovers are significant in Column (8) Panels A and B, the effects again do not cancel out with the total effect of electrification when evaluated at the mean and account for less than a third of the effect of access in the first rows of Panels A and B. The estimates of the main effect of access on entrants is even larger than the estimated effect without controlling for spillovers as in Table 7. Together, the total effect of access is of a similar magnitude of the effect of access on entrants in Table 7.

The last rows in Panels A and B present 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 all columns. This indicates that indeed electrification does create new economic activity, and the effects are not restricted to relocation of economic activity.

Finally, 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 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}.

$$Y_{vpt} = \alpha + \beta N^S_{vpt} + \theta N_{vp} + \mu G_v + \boldsymbol{V'}_{vpt}\boldsymbol{\eta} + \gamma_{pt} + \epsilon_{vpt} \tag{4}$$

I instrument N^S_{vpt} with the average distance of neighboring desas to the hypothetical

interconnection, conditional on the desa's distance to the least cost hypothetical interconnection G_v or E_v.

Table 10 shows the IV and OLS results from Equation 4. The estimates in Panels A and B across most outcomes are statistically indistinguishable from zero and is small in magnitude, but also positive. Panel A columns (1) and (7) shows a positive effect of an increase in the number of switching neighbors on the number of firms and number of entrants in the desa, but the effect disappears when controlling for the geographic and baseline characteristics. The positive and significant estimates in Column (8) Panel B of the effect of switching neighbors on entry show that there are possible agglomeration effects of electrification. In that case, the effects of electrification would be underestimated.

Given the mixed set of evidence from the three tests above, negative spillovers do not seem to be prominent in this particular setting, and do not explain all the increase in manufacturing activity at the local level. The grid expansion in Java has therefore lead to the creation of new industrial activity, and caused industrialization.

4.4 Fixed Effects IV as an Alternative Strategy

One remaining doubt about the cross-sectional identification strategy is that the rich set of controls used in the main specifications does not account for all unobservables. Adding desa fixed effects, or a fixed effects IV specification, is one suggestion to underpin the credibility of the identification strategy, by controlling for time-invariant unobservables:

$$Y_{vpt} = \alpha + \beta Access_{vpt} + \boldsymbol{V'}_{vpt}\boldsymbol{\eta} + \gamma_d + \delta_{pt} + \epsilon_{vpt} \tag{5}$$

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

A desa fixed effects IV version of specification 1 is possible using the same instruments above interacted with year dummies (since G_v and E_v are time-invariant), exploiting *time* variation in how the instrument explains access. Table 11 presents the results. Estimates of the effect of electrification on industrial outcomes remain positive and significant after including the desa fixed effects. Compared to results in Table 6, the IV estimates with both instruments in Panel A and B are even larger.

While controlling for unobservables is an advantage, using a fixed effects specification comes at a cost. Given that different desas get connected at different times during the sample period, and that many desas are already connected at baseline, a difference-in-difference specification in this setting means comparing desas before and after connection to two sets of control groups at every point in time: connected desas (connected at baseline and year$< t$), and desas that didn't get connected by 2000. In addition to the parallel trends assumption, a condition for the validity of a difference-in-difference specification is that there are no spillovers across units in the cross section, which I investigated in detail above, but also within units over time. The concern is a violation of SUTVA, but

in a dynamic sense: for any given desa, getting access to electricity today will have a "spillover" effect on the outcomes of that same desa in the future. This can be the case if the effect of electrification on industrial outcomes grows over time. A simple example is when the number of firms in year t is a function of access in year t and the number of firms in year $t-1$. A second dynamic spillover is directly related to the effect of electrification. If the effect of access is cumulative over time, where outcome Y is a function of access today and access in previous periods, a difference-in-difference would underestimate the effect of electrification. To test how important these dynamic spillovers are, I estimate the following long-difference specification:

$$\Delta Y_{vp(T-1991)} = \beta \Delta Access_{vp(T-1991)} + \boldsymbol{\Delta V'}_{vp(T-1991)} \boldsymbol{\eta} + \delta_{p(T-1991)} + \nu_{vp(T-1991)} \tag{6}$$

As before, I instrument $\Delta Access_{vp(T-1991)}$ with G_v and E_v separately. I estimate Equation (6) for three different values of T: 1992, 1996 and 2000. If dynamic spillovers matter, we would expect the effect of electrification to grow over time, and the estimates of β would be different across the three models. Results in Table 12 confirm the presence of dynamic effects. Focusing on Panel A Columns (1) to (3), the effect of a change in access on a change in the number of firms at the desa level is positive, and it increases the longer the time difference. The pattern is the same for the effect of $\Delta Access_{vp(T-1991)}$ on the change in the number of workers in manufacturing (Columns (4) to (6)), and on the change in manufacturing output (Columns (7) to (9)). Results in Panel B paint more or less a similar picture, with a less obvious increase of the effect between the 5-year and 9-year difference.

The takeaway from Table 12 is that the effect of electrification on industrial outcomes grows over time. A desa that has been connected for 5 years is thus not a good counterfactual for a desa that just got connected to the grid, and a difference-in-difference specification in this setting is not valid.

5 Electrification and Firm Performance

So far, results show that the expansion of the electricity grid caused an increase in manufacturing activity and increased firm turnover in Java. A change in firm turnover could mean that electrification is changing the composition of firms in the industry by affecting barriers to entry. In this section, I investigate if and how electrification affected the composition of firms. I ask whether electrification increases industrial activity by selecting the same type of firms or are the firms in electrified areas 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. As performance measures, I consider firm sales and inputs, and revenue productivity. I start by explaining how I structurally estimate firm revenue productivity in the following section.

5.1 Estimating Revenue Productivity

Productivity is defined as the efficiency with which a firm transforms inputs into output. Let $F(.)$ be an industry level production technology. Output quantity Q_{it} of firm i in year t if produced according to $Q_{it} = A_{it}F(\mathbf{X}_{it}, \beta)$. Firm productivity is A_{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):

$$y_{it} = p_{it} + q_{it} = f(\mathbf{x}_{it}, \beta) + a_{it} + p_{it} + \epsilon_{it} \tag{7}$$

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

$$\phi_{it} = a_{it} + p_{it} \tag{8}$$

Since ϕ_{it} is unobservable, and it is correlated with inputs, estimating the production function with OLS will give biased estimates of the production function coefficients. Following the literature initiated by Olley and Pakes (1996) and Levinsohn and Petrin (2003), I estimate the production function as in Ackerberg, Caves, and Frazer (2015) using energy spending as a proxy for productivity.

I assume a Cobb-Douglas production function:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \beta_e e_{it} + \phi_{it} + \epsilon_{it} \tag{9}$$

where y_{it} is output, k_{it} is capital, l_{it} is the wage bill, m_{it} is materials spending, 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 2-digit industry.

The classic endogeneity challenge in estimating Equation (9) 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 is the simultaneity bias. In addition, only surviving firms are observed in the data, leading to survival bias. While the recent literature ignores survival bias, this bias might be particularly important in the context of this paper since I show that electrification increases selection in Section 4.2.

I assume that 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}.

I account for the simultaneity bias by using a proxy for the omitted variable, productivity ϕ_{it}. Under the assumption of monotonicity[34], more productive firms will use more

[34]Energy spending is more likely to satisfy the monotonicity assumption than materials as raw materials can be stored.

inputs. Therefore, using a first order condition of the firm optimization problem, energy spending can be inverted to infer productivity:

$$\phi_{it} = h^{-1}(e_{it}, k_{it}, l_{it}, m_{it}) \tag{10}$$

Substituting back in Equation (9):

$$y_{it} = \kappa(k_{it}, l_{it}, m_{it}, e_{it}) + \epsilon_{it} \tag{11}$$

Estimating Equation (11) non-parametrically produces an estimate of predicted output $\hat{\kappa}_{it}$.

The second element of the estimation procedure is the assumption that ϕ_{it} follows a first-order Markov process where productivity today only depends on productivity in the previous period and a random shock:

$$\phi_{it} = g(\phi_{it-1}) + \eta_{it} \tag{12}$$

where $g(.)$ is an unknown function and η_{it} is an i.i.d. shock uncorrelated with k_{it-1}. The estimation proceeds with the following moment conditions:

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

These moment conditions are based on the law of motion of ϕ_{it} as in Equation (12):

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

and

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

I implement the one step GMM estimator as suggested by Wooldridge (2009), controlling for firm survival as in Olley and Pakes (1996).

5.2 Effect of Electrification on Firms

I start by looking at the effect of access on firm output and inputs. I first present the estimation results of specification (13) for different firm-level outcome variables:

$$y_{ivpst} = \alpha + \beta Access_{vpst} + \boldsymbol{X'}_{ivpst}\boldsymbol{\mu} + \boldsymbol{V'}_{vpst}\boldsymbol{\eta} + \gamma_p + \delta_{st} + \epsilon_{ivpst} \tag{13}$$

where δ_{st} are 3-digit industry-by-year fixed effects, $\boldsymbol{X}_{ivpst}$ is a vector of firm controls including ownership dummies[35] and export status. The vector $\mathbf{V}_{vpst}$ include the desa-level controls as before, including desa-level baseline outcomes.

Table 13 shows the IV and OLS versions of specification (13) for the log values of firm-level deflated sales, deflated capital, wage bill, number of workers, energy bill, quantity of electricity consumed in kWh, and revenue productivity TFPR. The treatment variable is $Access_{vpt}$, defined at the desa level and instrumented with G_v and E_v. There is still a relatively strong first stage with both IVs. Panel A 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 A, 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. Finally, Column (7) shows that on average firms in electrified desas are 30% more productive. Results in Panel B with the Euclidean IV are reassuringly similar in magnitude and depict the same pattern. Panel C 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.[36].

Evidence for spillovers in Section 4.3 were mixed, but one of the estimates showed potential negative spillovers in the output market. I therefore estimate Equation (13) including the number of connected neighbors N^C_{vpt} to account for these spillover effects[37] as in Equation (3). Spillovers in the output market could occur because of competition from trade across desas. Spillovers could also occur in the input markets. Electrification in neighboring desas could increase demand for labor and capital from firms located nearby. Negative spillovers in electricity supply could occur if increased access in the region increases the transmission load and causes more outages. I investigate these possibilities empirically. Finally, there could be negative spillovers on revenue productivity through prices if a firm faces more intense price competition from neighboring desas that

[35]There are four types of ownership: domestic, foreign, central government, and local government.

[36]Table C.2 in Appendix C repeats the same analysis but using a different definition for access; $Connected_{it}$. This is a dummy variable defined at the firm-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 13

[37]Neighbors are defined at the desa level and are the neighboring desas to the desa where a firm is located. Neighboring desas are within a 5 Km radius. With 7 km radius the instruments become weak in the selected sample of desas where the firms are located.

are connected.

Table 14 presents the results. The estimated effect of electrification on output, inputs, and TFPR all remain positive, significant, and comparable in magnitude to the estimates in Table 13. The estimated effect of spillovers is only marginally significant for output, capital, number of workers, and electricity. When evaluated at the mean of N^C_{vpt}, the total effects are small relative to the main effect of electrification. The null hypothesis of the joint effect summing up to zero is rejected for all outcomes. As with the results at the desa level, there is no evidence for any spillovers using the Euclidean IV in Panel B, and the evidence is again mixed.

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[38]. Compared to the Allcott, Collard-Wexler, and O'Connell (2016) result, the effect of electrification on average sales in the desa is much larger. Using the most conservative estimate after accounting for spillovers from Table 14 Column (1) Panel A[39], electrification causes average firm sales to increase by 300%. This means that in addition to the within firm effect of electrification on sales, there are large selection effects. The size of the effect highlights that fact that the extensive margin of electricity supply has a greater 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, and O'Connell (2016) also find that shortages do not affect labor input. In contrast, I find a large effect of access on labor. One explanation is that the extensive margin of electrification, through selection, causes the average firm size in the market to increase, and that these selection effects are unlikely to be substantial with shortages.

6 Reallocation

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

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

[39] $exp(2.12 - 0.0263 * 26.48) - 1 = 3.15$

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 revenue productivity index Φ_{st}, defined as the revenue-weighted average of firm revenue productivity $\Phi_{it} = exp(\phi_{it})$ in an industry s in year t, into an unweighted average and a covariance term (Olley and Pakes (1996)):

$$\begin{aligned}\Phi_{st} &= \sum_{i=1}^{N} S_{it}\Phi_{it} \\ &= \frac{1}{N}\sum_{i=1}^{N}\Phi_{it} + \sum_{i=1}^{N}(S_{it} - \frac{1}{N})(\Phi_{it} - \frac{1}{N}\sum_{i=1}^{N}\Phi_{it}) \\ &= \overline{\Phi}_{st} + Ncov(S_{it}, \Phi_{it})\end{aligned} \tag{14}$$

where S_{it} is firm i revenue share in sector s. $\overline{\Phi}_{st}$ 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 the desa level analysis where I instrument access with the average distances of firms in the industry to the hypothetical interconnections G_s and E_s. The estimating equation is:

$$Y_{st} = \alpha + \beta Access_{st} + \boldsymbol{X'_{st}\eta} + \gamma_{pt} + \delta_s + \epsilon_{st} \tag{15}$$

where $\boldsymbol{X_{st}}$ is a vector of controls as before (averaged across firms in the sector), γ_{pt} is province-by-year fixed effect and δ_s is a sector fixed effect. Table 15 presents the results from estimating Equation (15), weighted by number of firms in a sector, and instrumenting for access. When aggregating up to the sector level, the geographic IV G_s becomes weak. I present the results in Panel A for completeness. The IV estimates in Panel B with the Euclidean IV E_s 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.

7 Conclusion

In this paper, I show that electrification has a substantial causal impact on the industrial sector. I highlight a new mechanism through which this effect can occur. This mechanism, firm turnover, is unlikely to operate in response to short-run improvements in electricity supply. 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. It is this turnover mechanism, embodied the increased entry and exit of firms, that is a driver of industrialization. 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 paper 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 (2019) 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.

References

Abeberese, A. B. (2017). Electricity cost and firm performance: Evidence from india. *Review of Economics and Statistics 99*(5), 839–852.

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.

Atack, J. and R. A. Margo (2011). The impact of access to rail transportation on agricultural improvement: The american midwest as a test case, 1850–1860. *Journal of Transport and Land Use 4*(2), 5–18.

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.

Bos, K., D. Chaplin, and A. Mamun (2018). Benefits and challenges of expanding grid electricity in africa: A review of rigorous evidence on household impacts in developing countries. *Energy for Sustainable Development 44*, 64 – 77.

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

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.

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

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review 108*(4-5), 899–934.

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. 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*, 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*.

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.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society 7*(1), 48–50.

Lee, K., E. Miguel, and C. Wolfram (2019). Experimental evidence on the economics of rural electrification. *Journal of Political Economy (Forthcoming)*.

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

Lenz, L., A. Munyehirwe, J. Peters, and M. Sievert (2017). Does large-scale infrastructure investment alleviate poverty? impacts of rwanda's electricity access roll-out program. *World Development 89*(C), 88–110.

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 indonesia–is 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.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics 90*(4), 683–701.

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.

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).

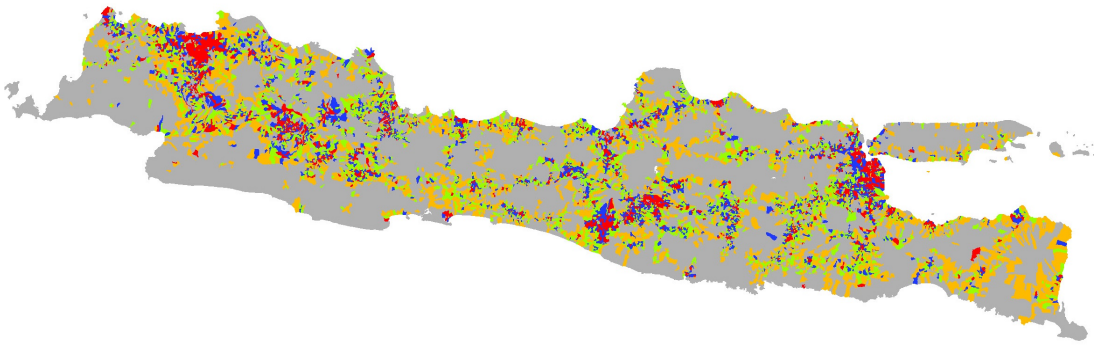
Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy 112*(6), 1181–1222.

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

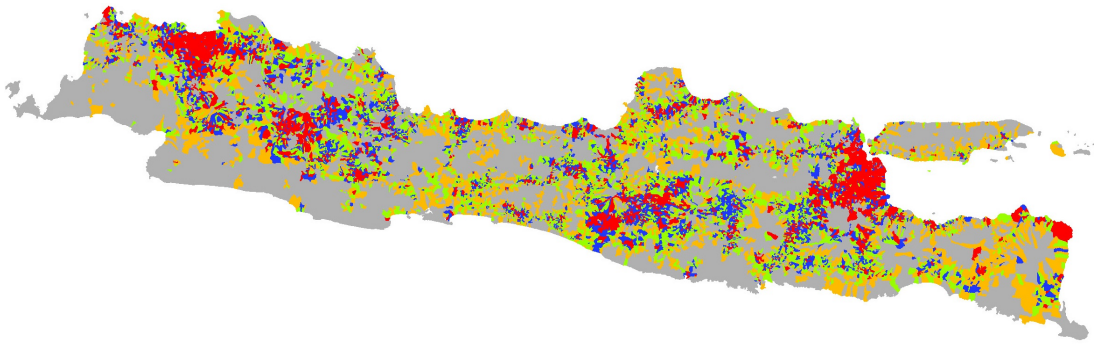
Tybout, J. R. (2000, March). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature 38*(1), 11–44.

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

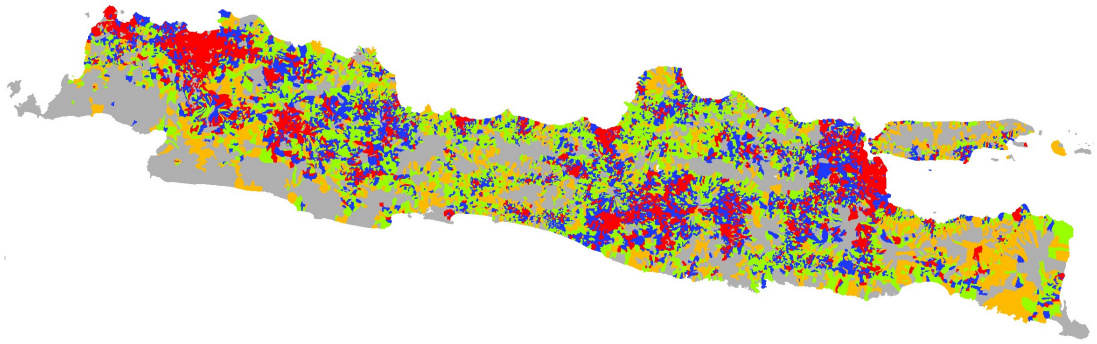
A Figures



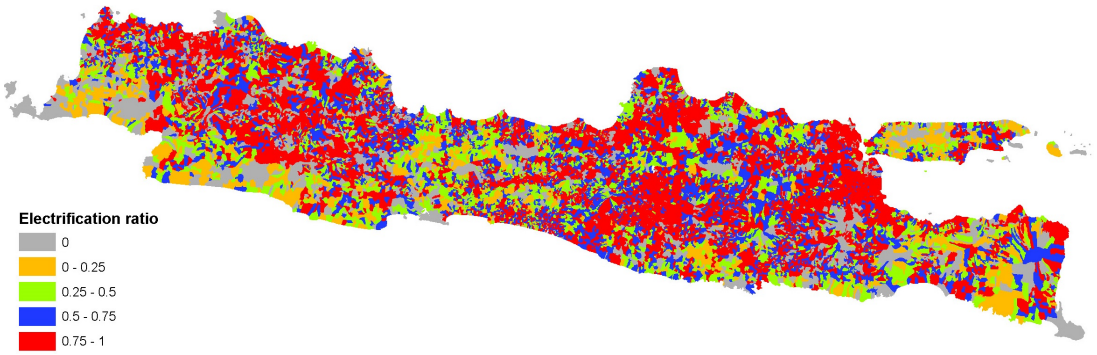
(a) 1990



(b) 1993



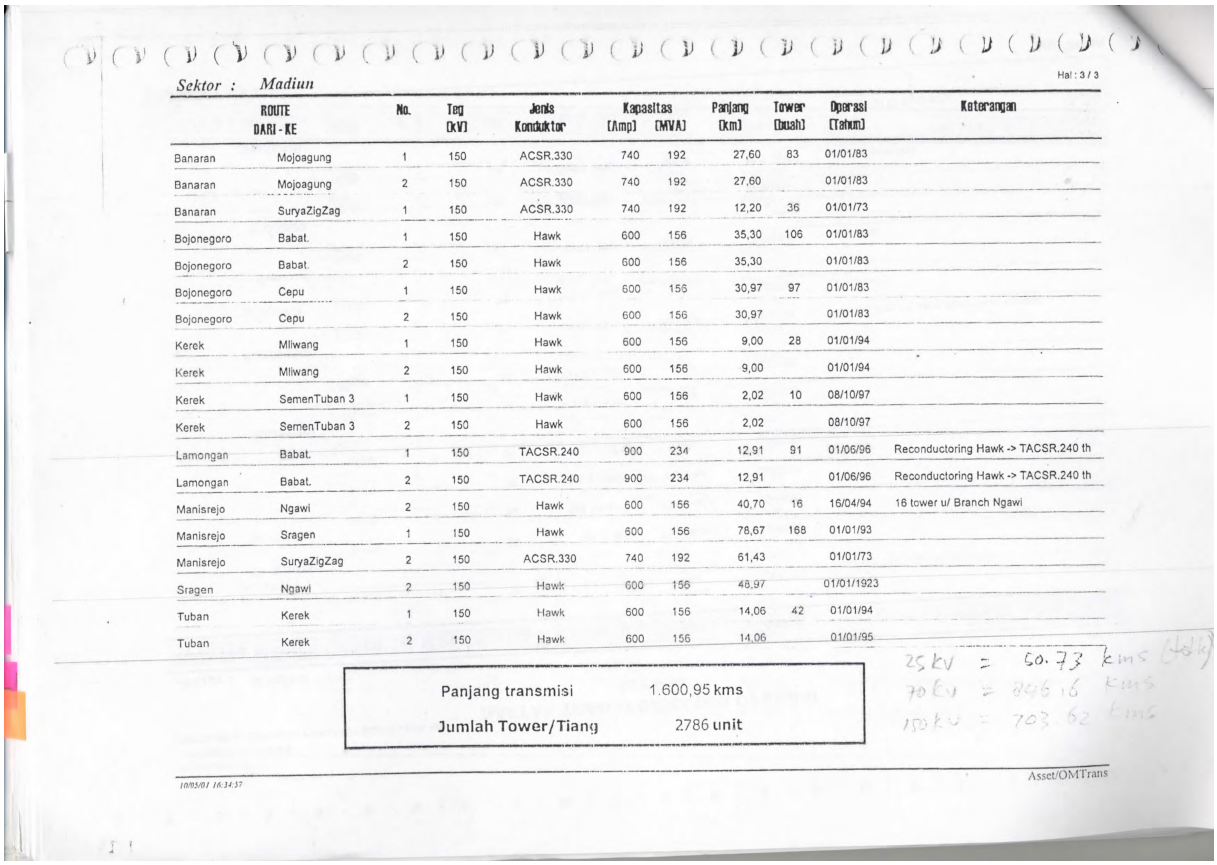
(c) 1996



(d) 2000

Figure 1: Desa-Level Electrification Ratios 1990 to 2000.
Source: PODES, BPS

Figure 2: Example of Inventory Table of Transmission Transformers.



Sektor : Madiun

Hal : 3 / 3

ROUTE DARI	ROUTE 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 (total)
70 kV = 846.16 kms
150 kV = 703.62 kms

10/05/01 16:34:37

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 3: Example of current maps of the transmission network in Java.

(a) Java and Banten



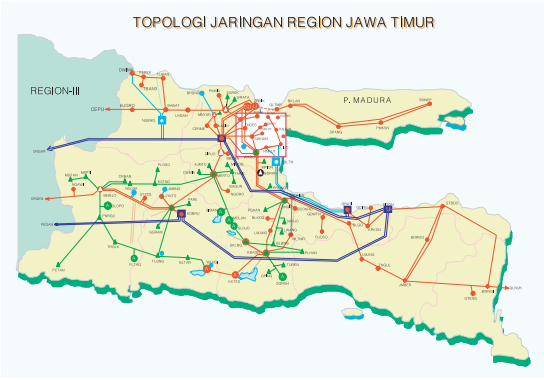
(b) West Java



(c) Central Java and Yogyakarta

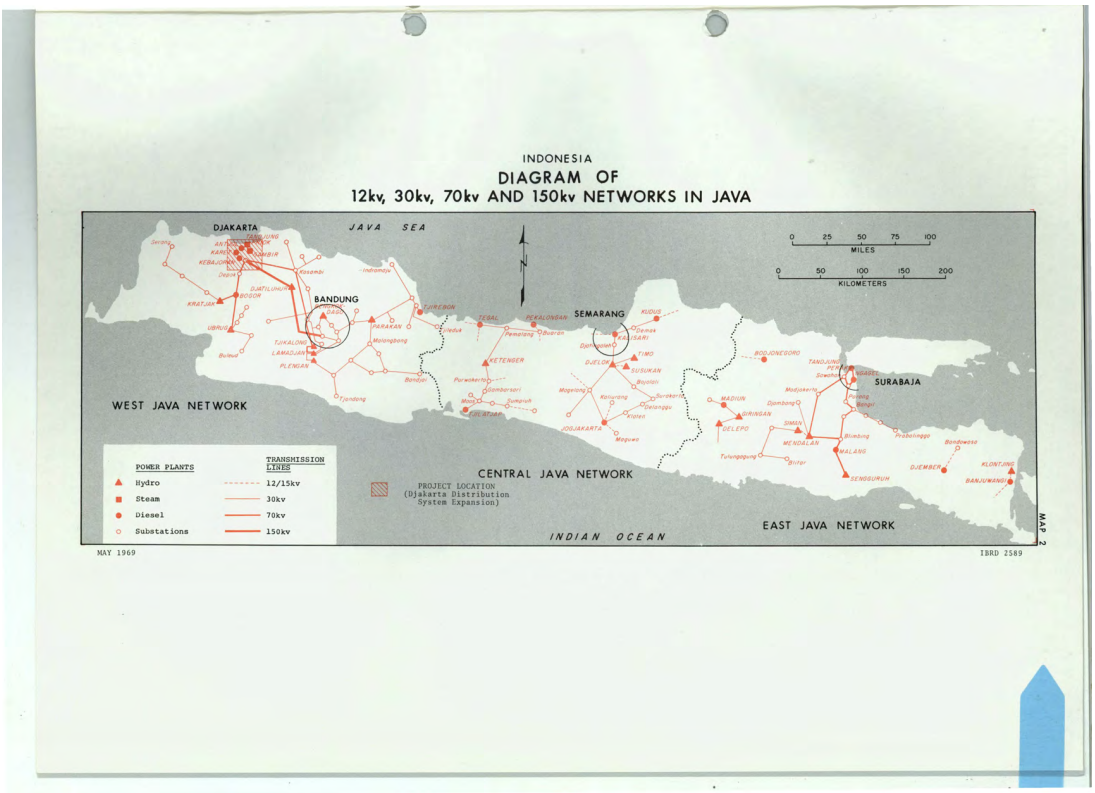


(d) East Java



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

Figure 4: Java Network 1969



Source: World Bank Archives

Figure 5: Java Network 1989



Source: World Bank Archives

Figure 6: Expansion of the Grid 1990-2000

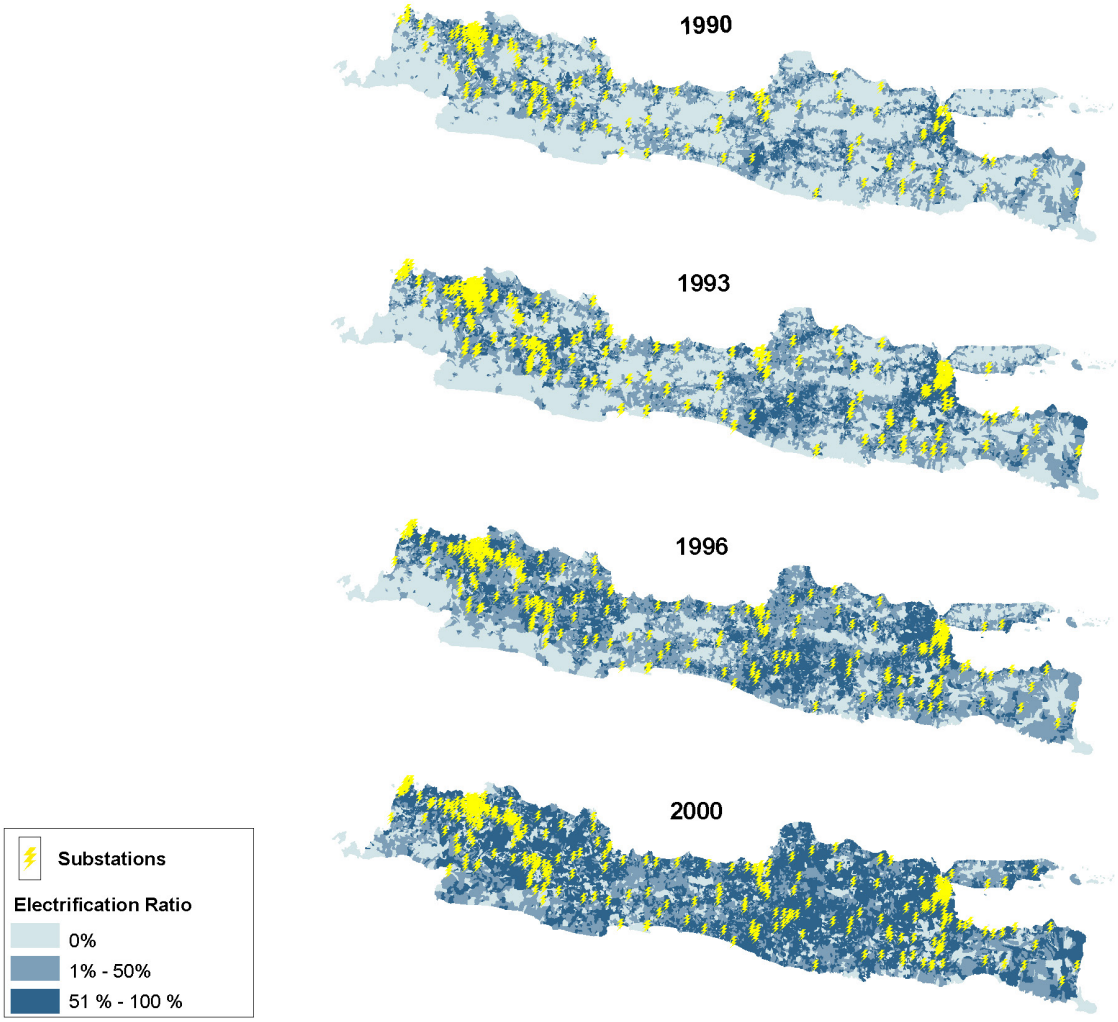


Figure 7: Empirical Strategy

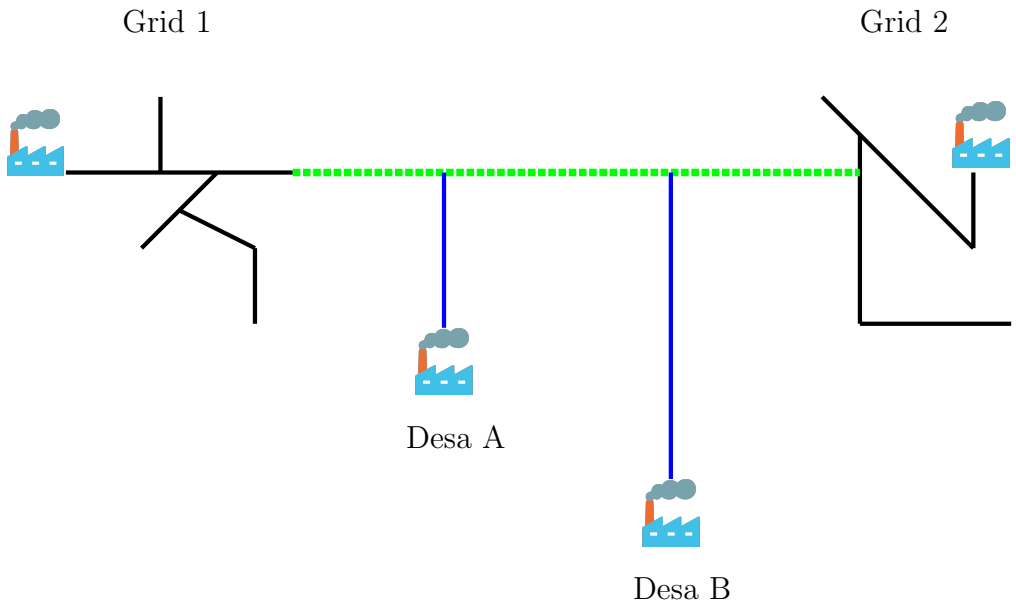


Figure 8: Geographic Least Cost Interconnection

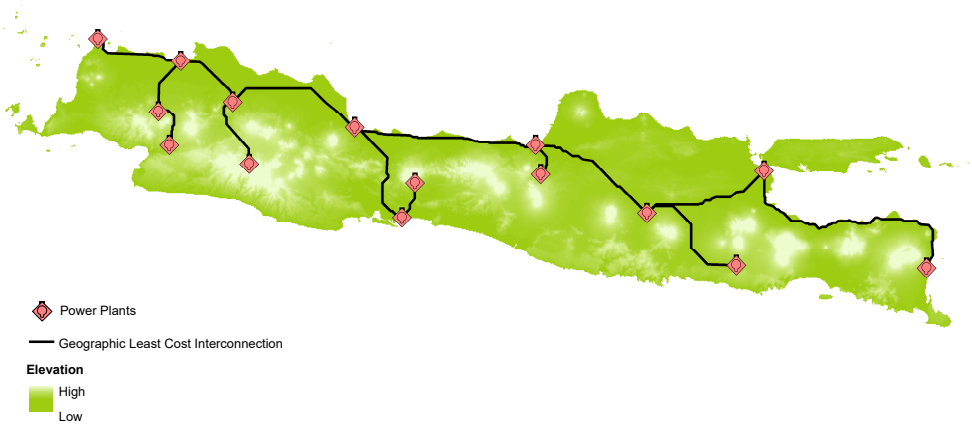


Figure 9: Hypothetical Euclidean Interconnection

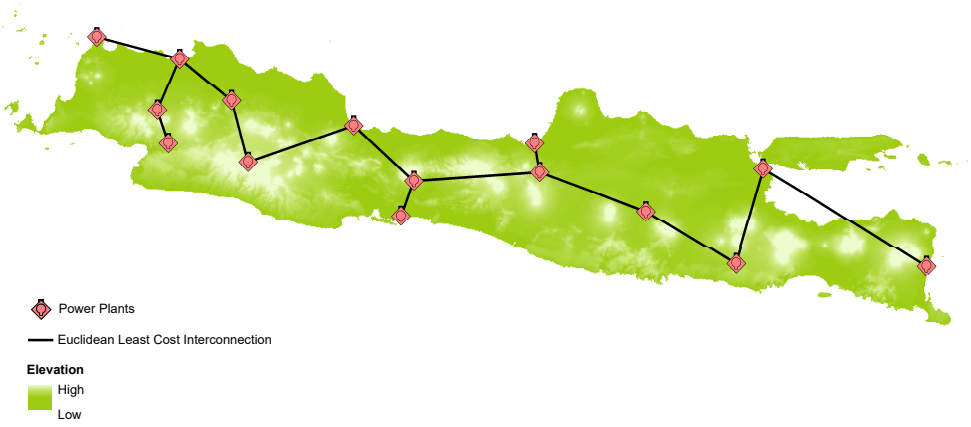
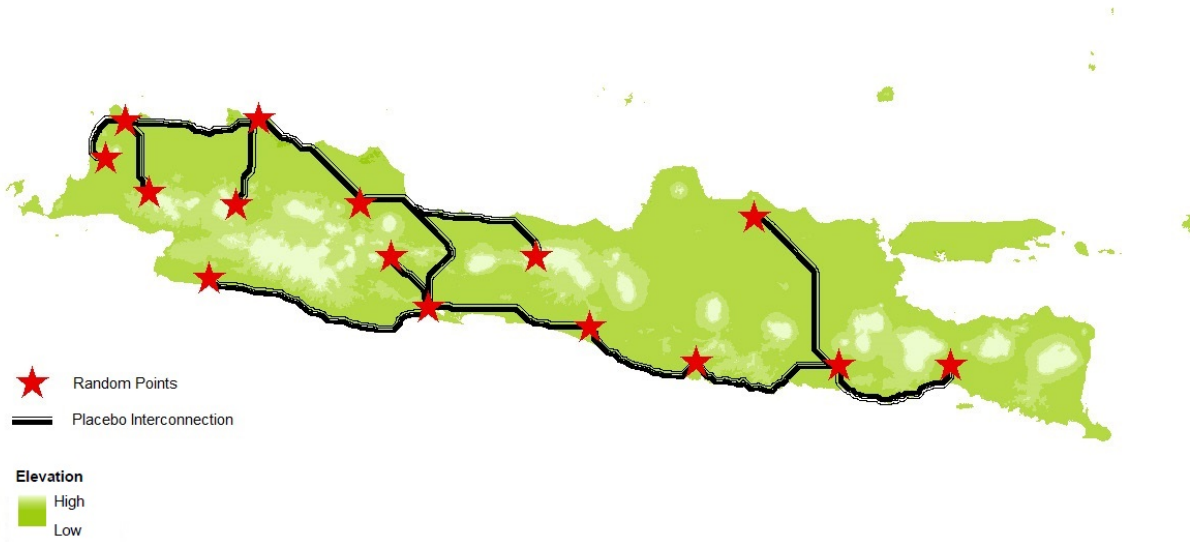


Figure 10: Placebo Least Cost Interconnection



B Tables

Table 1: Summary statistics: Electrification Infrastructure

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

Table 2: Desa-Level Summary Statistics

Variable	Mean	Median	Min	Max
Access	0.59	1	0	1
Number of firms	0.91	0	0	204
Number of firms > 0	4.2	2	1	204
Area (km^2)	5.7	4.3	1	540
Population	4,391	3,328	36	803,732
Pop. Density (per km^2)	2,548	1,451	7.7	36,413
Number of desas in t	22,831			

Table 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

Notes: See Section 3 for discussion. Summary statistics by industry showing the number of unique firm observations in each industry in 1990 and 2000 in Columns (1) and (2). Industries are ordered by the number of observations in each industry in 1990. Access is an indicator variable equal to 1 if a firm is located in a desa with access to electricity. Colums (3) and (4) show the access ratio in each industry between 1990 and 2000.

Table 4: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
Distance to Nearest City (KM)	43.93	32.66	0	227.15
Distance to Coast (KM)	30.63	21.09	0	166.06
Distance to Nearest Road (KM)	3.7	6.92	0	168.59
Elevation (m)	242.3	316.09	0	2686
Political Status	0.93	0.25	0	1
Dummy=1 if desa has a motorstation	0.02	0.14	0	1
Dummy=1 if desa has a railway	0.01	0.09	0	1
Dummy=1 if desa has a seaport	0	0.03	0	1
Dummy= 1 if desa has a riverpier	0	0.04	0	1
Dummy=1 if desa has an airport	0	0.02	0	1
Dummy= 1 if road fits 4 wheeled vehicle	0.95	0.21	0	1
Legal Status: Government Rule	0.77	0.42	0	1
Legal Status: State Ministry	0.17	0.38	0	1
Legal Status: Governor	0.03	0.18	0	1
Legal Status: Other	0.02	0.14	0	1
N	228310			

Notes: See Section 3 for discussion. Summary statistics of the desa-level control variables. Distance variable are all defined as the Euclidean distance measured from the centroid of the desa to the nearest city, cost, or road. Elevation is the average elevation in the desa. Political status is an indicator variable equal to one if the desa is the governed by an elected official, and zero if governed by appointed civil servant. Legal status of the desa refers to whether the desa is formed under a Government decree, Ministerial decree, Regency decree, or other.

Table 5: First Stage Regressions

Instrument	G_v		E_v		Placebo	
Dependent Variable	$Access_{vpt}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Instrument (KM)	-0.00407***	-0.00224***	-0.00459***	-0.00422***	-0.00000037***	0.0000000065
	(0.000131)	(0.000146)	(0.000130)	(0.000150)	(0.000000117)	(0.000000118)
Distance to Coast (KM)		0.000253*		-0.000530***		0.000237
		(0.000147)		(0.000145)		(0.000146)
Dummy=1 if desa has a motorstation		-0.0185		-0.0129		-0.0207
		(0.0138)		(0.0137)		(0.0139)
Dummy=1 if desa has a railway		0.0530***		0.0464**		0.0635***
		(0.0196)		(0.0193)		(0.0194)
Dummy=1 if desa has a seaport		-0.0896		-0.103*		-0.0918
		(0.0582)		(0.0575)		(0.0603)
Dummy=1 if desa has a riverpier		0.00973		0.0241		0.0263
		(0.0456)		(0.0458)		(0.0475)
Dummy=1 if desa has an airport		0.175***		0.156***		0.171***
		(0.0439)		(0.0406)		(0.0444)
Dummy=1 if road fits 4 wheeled vehicle		0.204***		0.194***		0.213***
		(0.0107)		(0.0107)		(0.0109)
First Stage F	968.8	235.5	1242	790.8	9.8	0.003
Observations	228,310	228,310	228,310	228,310	228,310	228,310
Year x Province FE	✓	✓	✓	✓	✓	✓
Baseline Outcome x Year FE	✓	✓	✓	✓	✓	✓
Elevation x Year		✓		✓		✓
Distance to Nearest Road x Year		✓		✓		✓
Distance to Nearest City x Year		✓		✓		✓
Legal and Political Characteristics		✓		✓		✓

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

Notes: See Section3.3.5 for discussion. First stage regressions of access instrumented with distance to hypothetical least cost interconnection in columns (1) and (2), hypothetical Euclidean interconnection in columns (3) and (4), and placebo least cost interconnection in Column (5) and (6). Access is defined at the desa level. A desa has $Access_{vpt} = 1$ if it is within 15 Km of the nearest substation. Baseline outcome is the number of firms in a desa in 1990. Robust standard errors in parentheses clustered at the desa level.

Table 6: Estimates of the Effect of Electrification on Local Manufacturing Outcomes.

	Sample: Desa-Level					
Dependent Variable Y	No. of Firms		No. of Workers in Manufacturing		Output Billion IDR	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Geo IV					
$Access_{vpt}$	0.624***	0.647***	168.7***	161.6***	19.29***	21.03***
	(0.0851)	(0.140)	(27.07)	(40.43)	(2.555)	(3.928)
First Stage F	968.8	235.5	980.3	235.3	991.7	237.4
	Panel B: Euclidean IV					
$Access_{vpt}$	0.418***	0.685***	175.9***	225.7***	20.89***	27.56***
	(0.0828)	(0.112)	(24.60)	(30.28)	(3.412)	(5.074)
First Stage F	1242	790.8	1249	790.4	1265	794.4
	Panel C: OLS					
$Access_{vpt}$	0.119***	0.0872***	36.85***	26.25***	3.699***	2.390***
	(0.0197)	(0.0152)	(5.410)	(3.956)	(0.380)	(0.273)
Observations	228,310	228,310	228,310	228,310	228,310	228,310
Year x Province FE	✓	✓	✓	✓	✓	✓
Baseline Y x Year FE	✓	✓	✓	✓	✓	✓
Desa Controls		✓		✓		✓
Mean Y	0.91		118.65		7.46	
Mean Y|No. Firms> 0	4.11		537.18		33.78	
*** p<0.01, ** p<0.05, * p<0.1						

Notes: See Section 4.1 for discussion. Results from IV and OLS regressions of Equation (1). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Table 7: Estimates of the Effect of Electrification on Entry, Exit, and Turnover.

	Sample: Desa-Level							
Dependent Variable Y	No. of Entrants		No. of Exiting Firms		Entry Rate		Exit Rate	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Geo IV				
$Access_{vpt}$	0.112***	0.121***	0.0624***	0.0464***	0.0593***	0.0570***	0.0241***	0.0236***
	(0.0126)	(0.0248)	(0.00670)	(0.0123)	(0.00899)	(0.00908)	(0.00427)	(0.00829)
First Stage F	976.9	235.6	997.6	238.2	182.6	174.5	184.9	49.81
				Panel B: Euclidean IV				
$Access_{vpt}$	0.0679***	0.0922***	0.0518***	0.0408***	0.0222***	0.0356***	0.0110***	0.00947**
	(0.0125)	(0.0191)	(0.00608)	(0.00797)	(0.00545)	(0.00790)	(0.00293)	(0.00419)
First Stage F	1251	788.5	1270	796	350.8	197.9	354.2	200
				Panel C: OLS				
$Access_{vpt}$	0.0195***	0.0129***	0.00521***	0.00807***	0.00183	0.00301*	0.00450***	0.00200**
	(0.00229)	(0.00215)	(0.00146)	(0.00107)	(0.00179)	(0.00179)	(0.000784)	(0.000826)
Observations	228,310	228,310	228,310	228,310	40,616	40,616	40,616	40,616
Year x Province FE	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Y x Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls		✓		✓		✓		✓
Mean Y	0.059		0.025		0.038		0.015	
Mean Y|No Firms> 0	0.268		0.114		0.038		0.015	

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

Notes: See Section 4.2 for discussion. Results from IV and OLS regressions of Equation (1). Geographic controls are defined at the desa level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Table 8: Estimates of the Effects of Electrification on District Level Outcomes.

	Sample: District-Level									
Dependent Variable Y (District Mean)	No. of Firms		No. of Entrants		No. of Exiting Firms		No. of Workers in Manufacturing		Output Billion IDR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Geo IV										
$Access_{dt}$	0.936***	1.605***	0.0840***	0.120***	0.0658***	0.0588**	130.9***	193.4**	13.00**	16.31**
	(0.277)	(0.509)	(0.0213)	(0.0412)	(0.0148)	(0.0241)	(46.56)	(83.58)	(5.052)	(6.847)
First Stage F	95.44	26.37	95.85	25.91	102.8	26.57	94.81	25.41	98.68	26.06
Panel B: Euclidean IV										
$Access_{dt}$	1.047***	1.110***	0.0653***	0.0805**	0.0644***	0.0590***	189.7***	197.4**	18.29***	26.17**
	(0.270)	(0.388)	(0.0217)	(0.0341)	(0.0141)	(0.0209)	(52.97)	(85.67)	(6.616)	(10.20)
First Stage F	99.43	50.60	100.1	49.46	105.2	51.38	96.60	49.08	101.6	50.40
Panel C: OLS										
$Access_{dt}$	0.122*	0.0100	0.0120**	0.00576	0.0139***	0.00357	13.60	-5.206	1.920**	-0.153
	(0.0712)	(0.0439)	(0.00517)	(0.00433)	(0.00315)	(0.00228)	(9.451)	(8.919)	(0.900)	(0.565)
Observations	15,670	15,670	15,670	15,670	15,670	15,670	15,670	15,670	15,670	15,670
YearxProvince FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Y x Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
District Controls		✓		✓		✓		✓		✓

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

Notes: See Section 4.3 for discussion. Results from OLS, and IV regressions of Equation (2) 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. Geographic controls are defined at the district level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the district level.

Table 9: Estimates of the Joint Effect of Electrification and Spillovers on Desa Level Outcomes.

	Sample: Desa-Level									
Dependent Variable Y	No. of Firms		No. of Workers in Manufacturing		Output Billion IDR		No. of Entrants		No. of Exiting Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Geo IV										
$Access_{vpt}$	0.396*	0.823***	11.33	131.3**	17.63***	28.76***	0.0932**	0.150***	0.0220	0.0390**
	(0.224)	(0.205)	(83.92)	(62.01)	(6.094)	(6.961)	(0.0379)	(0.0303)	(0.0240)	(0.0171)
N^C_{vpt}	0.0172*	-0.0114***	8.954**	0.219	0.294	-0.484***	0.00174	-0.00172**	0.00202*	0.000241
	(0.00898)	(0.00437)	(3.607)	(1.448)	(0.221)	(0.164)	(0.00166)	(0.000874)	(0.00106)	(0.000462)
First Stage F	49.94	66.97	49.85	66.98	49.68	66.70	49.66	66.72	49.58	66.54
H0: $\hat{\beta} + \bar{N^C_{vpt}} * \hat{\mu} = 0$										
Prob>F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Euclidean IV										
$Access_{vpt}$	1.516***	0.935***	247.2***	186.4***	32.74***	33.28***	0.255***	0.159***	0.106***	0.0661***
	(0.217)	(0.216)	(63.86)	(61.84)	(5.855)	(6.609)	(0.0359)	(0.0318)	(0.0188)	(0.0182)
N^C_{vpt}	-0.0327***	-0.00515	-1.087	1.709	-0.232	-0.0480	-0.00585***	-0.00176**	-0.00181***	-0.000761
	(0.00707)	(0.00520)	(1.881)	(1.642)	(0.213)	(0.182)	(0.00130)	(0.000881)	(0.000619)	(0.000497)
First Stage F	67.72	57.52	67.72	57.47	67.41	57.38	67.38	57.36	67.26	57.31
H0: $\hat{\beta} + \bar{N^C_{vpt}} * \hat{\mu} = 0$										
Prob>F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: OLS										
$Access_{vpt}$	0.144***	0.127***	31.89***	25.60***	3.119***	2.617***	0.0193***	0.0158***	0.00211	9.57e-05
	(0.0221)	(0.0213)	(7.345)	(6.801)	(0.751)	(0.686)	(0.00369)	(0.00355)	(0.00208)	(0.00203)
N^C_{vpt}	-0.00116*	-0.00128**	0.0818	0.0848	0.0113	0.000844	-0.000129	-0.000132	0.000263***	0.000220***
	(0.000651)	(0.000627)	(0.160)	(0.146)	(0.0160)	(0.0151)	(0.000110)	(0.000112)	(6.01e-05)	(6.20e-05)
Observations	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310
N_{vp}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year x Province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Y x Year Fe	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls		✓		✓		✓		✓		✓

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

Notes: Results from OLS and IV regressions of Equation (3). Desa controls are defined at the desa level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level. Mean $N^C_{vpt} = 29.28$, mean $N_{vt} = 43.49$.

Table 10: Estimates of the Effect of Connecting Neighboring Desas on Desa Level Outcomes.

	Sample: Desa-Level - Unconnected									
Dependent Variable Y	No. of Firms		No. of Workers in Manufacturing		Output Billion IDR		No. of Entrants		No. of Exiting Firms	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Geo IV										
N^S_{vpt}	0.438*	0.0322	205.0	117.8	0.159	0.208	0.0948**	0.0164	0.0353	0.00408
	(0.242)	(0.206)	(145.8)	(91.93)	(2.144)	(1.452)	(0.0371)	(0.0302)	(0.0242)	(0.0214)
First Stage F	18.79	13.38	18.89	13.39	18.92	13.38	18.93	13.47	18.94	13.30
Panel B: Euclidean IV										
N^S_{vpt}	0.298	0.00428	209.9	137.3	1.954	1.434	0.136**	0.0671*	0.0637*	0.0303
	(0.249)	(0.238)	(130.9)	(90.68)	(1.816)	(1.347)	(0.0555)	(0.0391)	(0.0355)	(0.0278)
First Stage F	24.41	12	24.09	11.95	23.95	11.91	24.25	12.02	24.29	11.87
Panel C: OLS										
N^S_{vpt}	-0.000660	-0.000282	0.947**	1.023**	0.0211	0.0226	-0.000433	-0.000279	-0.000133	-0.000105
	(0.00108)	(0.00110)	(0.455)	(0.455)	(0.0141)	(0.0140)	(0.000333)	(0.000334)	(0.000127)	(0.000127)
Observations	92,939	92,939	92,939	92,939	92,939	92,939	92,939	92,939	92,939	92,939
Own Instrument	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N_{vp}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year x Province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Y x Year Fe	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls		✓		✓		✓		✓		✓

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

Notes: Results from OLS and IV regressions of Equation (4). Desa controls are defined at the desa level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level. Mean $N^S_{vpt} = 0.444$, and mean $N_{vp} = 34.93$.

Table 11: Impact of access on desa level outcomes - fixed effects.

	Sample: Desa-Level					
Dependent Variable Y	No. of Firms		No. of Workers in Manufacturing		Output Billion IDR	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Geo IV						
$Access_{vpt}$	2.456***	1.253***	766.7***	384.0***	209.0***	97.42***
	(0.419)	(0.290)	(119.7)	(91.88)	(31.20)	(16.18)
First Stage F	83.86	96.53	83.86	96.53	83.86	96.53
Panel B: Euclidean IV						
$Access_{vpt}$	1.186***	0.815***	596.5***	206.0*	214.7***	160.5***
	(0.290)	(0.246)	(90.31)	(110.1)	(31.32)	(26.68)
First Stage F	113.7	140.2	113.7	140.2	113.7	140.2
Panel C: OLS						
$Access_{vpt}$	-0.0442***	-0.0592***	-4.659	-8.031*	-2.921***	-3.792***
	(0.0114)	(0.0120)	(3.655)	(4.559)	(0.549)	(0.659)
Observations	228,310	228,310	228,310	228,310	228,310	228,310
Desa FE	✓	✓	✓	✓	✓	✓
Year x Province FE	✓	✓	✓	✓	✓	✓
Baseline Y x Year FE	✓	✓	✓	✓	✓	✓
Desa Controls		✓		✓		✓
Mean Y	0.88		114.33		7.17	
Mean Y|No. Firms> 0	4.08		528.21		33.13	

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

Notes: See Section 4.4 for discussion. Results from IV and OLS regressions of Equation (5). Instruments in Panels A and B are interacted with year dummies. Desa controls are defined at the district level and include distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and time-varying infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Table 12: Impact of access on desa level outcomes -Long Difference

Sample: Desa-Level									
Dependent Variable Y	No. of Firms			No. of Workers in Manufacturing			Output Billion IDR		
Difference	1 Year	5 Years	9 Years	1 Year	5 Years	9 Years	1 Year	5 Years	9 Years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Geo IV									
$\Delta Access_{vpt}$	0.561**	1.571***	4.377***	144.7	300.9***	722.8***	14.66**	37.19***	186.7***
	(0.263)	(0.400)	(1.160)	(113.4)	(106.8)	(236.1)	(5.809)	(11.04)	(56.78)
First Stage F	283.7	68.58	23.97	283.7	69.29	24.27	283.3	69.49	24.17
Panel B: Euclidean IV									
$\Delta Access_{vpt}$	0.894**	4.488***	3.378***	599.3***	1,556***	923.2***	38.95***	126.4***	255.1***
	(0.369)	(1.306)	(0.747)	(180.5)	(412.7)	(194.0)	(9.363)	(36.99)	(57.99)
First Stage F	242.1	21.14	77.32	243	21.75	78.06	242.9	21.85	77.76
Panel C: OLS									
$\Delta Access_{vpt}$	0.0431**	-0.0522***	-0.0900***	-1.936	-10.53**	-10.87**	0.0169	-0.880**	-4.821***
	(0.0170)	(0.0175)	(0.0216)	(5.342)	(4.573)	(5.097)	(0.799)	(0.382)	(1.783)
Observations	22,831	22,831	22,831	22,831	22,831	22,831	22,831	22,831	22,831
Province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Y	✓	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 4.4 for discussion. Results from IV and OLS regressions of Equation (6). Desa controls are defined at the desa level and include distance to nearest city, elevation, distance to road, change in desa political and legal status, and change in time-varying infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Table 13: Impact of connection on the sales and inputs at the firm level.

Sample: Firm-Level							
		Inputs					
Log Dependent Variable Y	Sales	Capital	Wage Bill	No. Workers	Energy Bill	Electricity (kWh)	TFPR (ϕ_{it})
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Geo IV							
$Access_{vpt}$	2.953***	1.437***	1.924***	3.695***	4.391***	6.366***	0.318***
	(0.574)	(0.317)	(0.436)	(0.789)	(0.905)	(1.507)	(0.114)
First Stage F	23.38	23.23	23.13	23.35	22.62	22.53	26.05
Panel B: Euclidean IV							
$Access_{vpt}$	2.926***	1.476***	2.496***	3.315***	3.731***	5.839***	0.191***
	(0.320)	(0.176)	(0.288)	(0.434)	(0.465)	(0.730)	(0.0477)
First Stage F	95.16	94.88	94.83	95.15	95.28	95.11	112.6
Panel C: OLS							
$Access_{vpt}$	0.544***	0.247***	0.367***	0.474***	0.535***	0.699***	0.0234**
	(0.0654)	(0.0287)	(0.0405)	(0.0621)	(0.101)	(0.102)	(0.0105)
Observations	130,016	130,016	130,001	130,016	128,594	112,322	123,620
IndustryxYear FE	✓	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓	✓
Baseline Y_v x Year FE	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 5.2 for discussion. Results from IV and OLS regressions of Equation (13). Desa controls are defined at the desa level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Firm Controls include export, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

Table 14: Impact of connection on the sales and inputs at the firm level - Spillovers.

Sample: Firm-Level							
		Inputs					
Log Dependent Variable Y	Sales	Capital	Wage Bill	No. Workers	Energy Bill	Electricity (kWh)	TFPR (ϕ_{it})
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Geo IV							
$Access_{vpt}$	2.120***	1.134***	1.080**	2.389***	2.746***	3.285***	0.251**
	(0.679)	(0.330)	(0.533)	(0.786)	(1.032)	(1.269)	(0.113)
N^C_{vpt}	-0.0263*	-0.0162**	-0.0112	-0.0323*	-0.0302	-0.0527*	-0.00430
	(0.0158)	(0.00787)	(0.0127)	(0.0196)	(0.0234)	(0.0277)	(0.00281)
First Stage F	11.83	11.69	11.95	12.10	12.14	12.03	12.98
H0:$\hat{\beta} + \bar{N^C_{vpt}} * \hat{\mu} = 0$							
Prob>F	0	0	0.001	0	0	0.002	0.009
Panel B: Euclidean IV							
$Access_{vpt}$	1.683*	0.680*	1.369*	1.663	2.313*	3.627**	0.109
	(0.875)	(0.358)	(0.776)	(1.025)	(1.207)	(1.573)	(0.118)
N^C_{vpt}	0.0279	0.0192	0.0234	0.0350	0.0313	0.0166	0.00327
	(0.0326)	(0.0128)	(0.0286)	(0.0368)	(0.0459)	(0.0520)	(0.00433)
First Stage F	10.71	10.69	10.79	10.74	10.83	9.696	10.96
H0:$\hat{\beta} + \bar{N^C_{vpt}} * \hat{\mu} = 0$							
Prob>F	0	0	0	0	0	0	0
Panel C: OLS							
$Access_{vpt}$	0.544***	0.247***	0.367***	0.474***	0.535***	0.699***	0.0234**
	(0.0654)	(0.0287)	(0.0405)	(0.0621)	(0.101)	(0.102)	(0.0105)
N^C_{vpt}	0.00597**	0.00463***	0.00739***	0.00723**	0.0155***	0.00787	-0.00109**
	(0.00253)	(0.00131)	(0.00185)	(0.00303)	(0.00405)	(0.00514)	(0.000446)
Observations	130,016	130,016	130,001	130,016	128,594	112,322	123,620
IndustryxYear FE	✓	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓	✓
Baseline Y_v	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 5.2 for discussion. Results from IV and OLS regressions of Equation (13). Desa controls are defined at the desa level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Firm Controls include export, and ownership dummies. Robust standard errors in parentheses clustered at the desa level. Mean $N^C_{vpt} = 26.48$, mean $N_{vt} = 30.74$.

Table 15: Estimates of the Effect of Electrification on Reallocation

	Sample: Sector-Level					
Dependent Variable Y	Weighted Average Φ_{st}		Unweighted Average $\overline{\Phi}_{st}$		Covariance $cov(S_{it}, \Phi_{it})$	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Geo IV						
$Access_{st}$	0.570	0.772	0.261	0.284	0.130	0.429
	(0.376)	(0.705)	(0.188)	(0.295)	(0.157)	(0.515)
First Stage F	3.812	1.360	4.168	1.456	10.90	1.364
Panel B: Euclidean IV						
$Access_{st}$	0.337***	0.532***	0.0920*	0.223***	0.244**	0.299**
	(0.119)	(0.170)	(0.0475)	(0.0789)	(0.0909)	(0.126)
First Stage F	27.89	18.19	29.07	18.54	27.54	18.35
Panel C: OLS						
$Access_{st}$	0.0518	0.0588	0.0115	0.0223	0.0409	0.0366
	(0.0567)	(0.0499)	(0.0169)	(0.0163)	(0.0489)	(0.0433)
Observations	7,292	7,292	7,292	7,292	7,292	7,292
Year x Province FE	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓
Baseline Y x Year FE	✓	✓	✓	✓	✓	✓
Sector Controls		✓		✓		✓

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

Notes: See Section 6 for discussion. Results from IV and OLS regressions of Equation (15) weighted by number of firms in a sector. Geographic controls are the average values at the sector level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the sector level.

C Additional Results

Table C1: Impact of electrification on desa level industrial outcomes - Log transformations.

	Sample: Desa-Level											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent Variable Y	Log(1+No Firms)		Log(1+No Workers)		Log(1+Output)		Log(h(No Firms))		Log(h(No Workers))		Log(h(Output))	
Panel A: Geo IV												
$Access_{vpt}$	0.155***	0.173***	0.678***	0.611***	0.274***	0.261***	0.197***	0.219***	0.763***	0.677***	0.321***	0.300***
	(0.0141)	(0.0284)	(0.0688)	(0.134)	(0.0247)	(0.0473)	(0.0180)	(0.0362)	(0.0782)	(0.153)	(0.0294)	(0.0563)
First Stage F	907.9	228.4	920.3	229	929.3	227.1	906.5	228.2	921.5	229.4	927.2	226.6
Panel B: Euclidean IV												
$Access_{vpt}$	0.112***	0.149***	0.672***	0.729***	0.255***	0.287***	0.142***	0.187***	0.756***	0.811***	0.295***	0.329***
	(0.0129)	(0.0169)	(0.0606)	(0.0804)	(0.0222)	(0.0308)	(0.0164)	(0.0215)	(0.0689)	(0.0913)	(0.0262)	(0.0362)
First Stage F	1192	786.3	1200	789.2	1195	773.7	1191	786.4	1202	790.3	1193	773.1
Panel C: OLS												
$Access_{vpt}$	0.0340***	0.0265***	0.204***	0.155***	0.0743***	0.0553***	0.0440***	0.0343***	0.234***	0.178***	0.0891***	0.0663***
	(0.00244)	(0.00242)	(0.0113)	(0.0110)	(0.00380)	(0.00355)	(0.00311)	(0.00310)	(0.0129)	(0.0126)	(0.00452)	(0.00425)
Observations	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310	228,310
Year x Province FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Y x Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls		✓		✓		✓		✓		✓		✓

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

Notes: See Section 4.1 for discussion. Results from IV and OLS regressions of Equation (1). Dependent variables are two different log transformations that preserve zeros for the number of firms, number of workers in manufacturing and total manufacturing output. The first transformation is a $log(1 + X)$. The second and more preferred transformation is hyperbolic inverse sine transformation $log(h(X))$ where $h(X) = X + (X^2 + 1)^{\frac{1}{2}}$ as in for example 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 x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the desa level.

Table C.2: Impact of connection on the sales and inputs at the firm level.

Sample: Firm-Level							
		Inputs					
Log Dependent Variable Y	Sales	Capital	Wage Bill	No. Workers	Energy Bill	Electricity (kWh)	TFPR (ϕ_{it})
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Geo IV							
$Connected_{it}$	4.016***	1.959***	2.610***	5.027***	6.418***	75.94	0.555***
	(0.840)	(0.480)	(0.652)	(1.141)	(1.421)	(53.83)	(0.204)
First Stage F	23.38	23.23	23.13	23.35	22.62	22.53	26.05
Panel B: Euclidean IV							
$Connected_{it}$	5.650***	2.871***	4.799***	6.376***	7.523***	212.5	0.474***
	(0.804)	(0.471)	(0.749)	(1.023)	(1.215)	(275.1)	(0.125)
First Stage F	40.47	39.81	40.87	40.96	34.98	0.603	50.48
Panel C: OLS							
$Connected_{it}$	0.674***	0.226***	0.436***	0.731***	0.755***	-0.689***	0.0570***
	(0.0576)	(0.0256)	(0.0356)	(0.0539)	(0.0846)	(0.0772)	(0.00890)
Observations	130,016	130,016	130,001	130,016	128,594	112,322	123,902
IndustryxYear FE	✓	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓	✓
Baseline Y_v	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 5.2 for discussion. Results from IV and OLS regressions of Equation (13). Desa controls are defined at the desa level and include distance to coast, distance to nearest city x year, elevation x year, distance to road x year, desa political and legal status, and infrastructure controls. Firm Controls include export, and ownership dummies. Robust standard errors in parentheses clustered at the desa level.

D Theoretical Motivation

D.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 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.

D.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:

$$U = H + \int_{i \in I} \alpha q_i di - \frac{1}{2}\eta\Big(\int_{i \in I} q_i di\Big)^2 - \frac{1}{2}\gamma \int_{i \in I} {q_i}^2 di \tag{16}$$

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:

$$q_i = \frac{\alpha}{\eta N + \gamma} + \frac{\eta N}{\gamma(\eta N + \gamma)}\bar{p} - \frac{1}{\gamma}p_i \tag{17}$$

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 (17) to zero:

$$p^{max} = \frac{\gamma\alpha}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} \tag{18}$$

The residual demand for product i from (17) can therefore be written as:

$$q_i = \frac{1}{\gamma}(p^{max} - p_i) \tag{19}$$

D.3 Production

On the production side, consider a single input technology[40] where firm i produces according the the following production function:

$$q_i = \phi_i x_i \tag{20}$$

where ϕ_i is the firm's physical productivity and x_i is the input of production which is supplied inelastically at a constant[41] 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:

$$p(c_i) = \frac{1}{2}(p^{max} + c_i) \tag{21}$$

The equilibrium profit is:

$$\pi(c_i) = \frac{1}{4\gamma}(p^{max} - c_i)^2 \tag{22}$$

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:

$$c^* = p^{max} = \frac{\gamma\alpha}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} \tag{23}$$

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

$$p(c_i) = \frac{1}{2}(c^* + c_i) \tag{24}$$

[40] 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.

[41] 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.

$$\mu(c_i) = \frac{1}{2}(c^* - c_i) \tag{25}$$

$$q(c_i) = \frac{1}{2\gamma}(c^* - c_i) \tag{26}$$

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 on the firm's own marginal cost. The cutoff c^* then implies implies a cutoff level for firm productivity:

$$\phi^* = \frac{w}{c^*} \tag{27}$$

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

D.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:

$$V^e = \frac{w^2}{4\gamma} \int_{\phi^*}^{\infty} \left(\frac{1}{\phi^*} - \frac{1}{\phi}\right)^2 dG(\phi) - s = 0 \tag{28}$$

Equation (28) pins down ϕ^* which summarizes the equilibrium. The equilibrium mass of firms N is determined using equations (21) and (23).

D.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:

$$\frac{d\phi^*}{dw} = -\frac{\partial V^e/\partial w}{\partial V^e/\partial \phi^*} > 0 \tag{29}$$

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:

$$\frac{d\phi^*}{ds} = -\frac{\partial V^e/\partial s}{\partial V^e/\partial \phi^*} < 0 \tag{30}$$

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 (24)-(26) conditional on survival are:

$$\bar{p} = \frac{1}{2}(c^* + \bar{c}) \tag{31}$$

$$\bar{\mu} = \frac{1}{2}(c^* - \bar{c}) \tag{32}$$

$$\bar{q} = \frac{1}{2\gamma}(c^* - \bar{c}) \tag{33}$$

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 (31) predicts that the average observed prices conditional on firm survival is lower when c^* is lower. Equations (32) and (33) 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 (25) and (26), 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 $c^* = p^{max}$:

$$c^* = \bar{p} + \frac{\gamma(\alpha - \bar{p})}{\eta N + \gamma} \tag{34}$$

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

$$N = \frac{2\gamma(\alpha - c^*)}{\eta(c^* - \bar{c})} \tag{35}$$

These equations state that tougher competition (lower c^*) is associated with a higher mass of active firms N and a lower average price[42] $\bar{p}$. To see this[43], suppose N increases, and that surviving firms don't change their prices following entry, keeping $\bar{p}$ constant. From equation (34), c^* will decrease. From equation (31), $\bar{p}$ will decrease as a 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 (31)-(33) 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 (29) and (30) 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 D.1 summarizes the predictions of the model, split by the different channels :

[42] An implicit assumption here is that $\alpha > c^*$ which implies that α is greater than $\bar{p}$ and $\bar{c}$.

[43] The intuition is the same as in Combes, Duranton, Gobillon, Puga, and Roux (2012).

Table D.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	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.

- 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? The results show that spillovers are minimal in this particular setting. Theoretically, the size of the spillovers depend on 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.

E PLN Demand Forecasts

E.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.

E.2 Residential Sector

- Energy Consumed: $E_t^R = E_{t-1}^R * (1 + \epsilon_t^R * g_t) + \Delta Nb_t^R * UK_t^R$ 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[44].
  - $Nb_t^R = Nb_{t-1}^R * (1 + CF_t^R * g_t)$
  - 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.

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

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

E.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:

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

E.5 Disaggregation

Because the Java-Bali system is interconnected, forecast is done at the system level. Figure 11 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 11: 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
Elecir.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	15446.0	16382.5
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	7924.3	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	85905	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	13.95	11.56	12.76	13.85	13.49	11.37	11.20	11.01	10.85	10.69	10.53
Power Contr./Cust.(VA)	187215.1	198881.9	224369.0	2325511.6	260958.2	248515.3	234652.4	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	1769.6	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	39201.6	42110.5	45206.9	48446.1
PLN SALES (GWh)	22401.8	25282.6	27308.1	32491.0	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	132280.8	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 done in 1992 for the Java-Bali system as part of the PLN 10-year business plan. *Source: PLN.*