Discussion Paper Series – CRC TR 224

Discussion Paper No. 052
Project B 07

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

Dana Kassem [1]

July 2021
(First version : November 2018)

[1] Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany,
d.kassem@uni-mannheim.de

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

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

Dana Kassem*

July 2021

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 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 by increasing the number of manufacturing firms, manufacturing workers, and manufacturing output. Electrification increases firm entry rates, but also exit rates, and entry accounts for most of the increase in output. This is consistent with electrification lowering entry costs, increasing competition and forcing unproductive firms to exit more often.

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

* Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany, d.kassem@uni-mannheim.de. I thank Michel Azulai, Oriana Bandiera, Gharad Bryan, Robin Burgess, Rachael Meager, Kieu-Trang Nguyen, Alex Rothenberg, Pasquale Schiraldi, Arthur Seibold, Junichi Yamasaki, and various participants at STICERD, LSE, NBER Summer Institute, Mannheim, and many other conferences. 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 (project B07) 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 high policy relevance of this issue, there is surprisingly little causal evidence on whether grid electrification truly drives industrialization or whether this relationship is merely a correlation along the path of development.

To answer this question, I use a rapid, large-scale government-led grid expansion during a period of rapid industrialization in Indonesia. I put together a comprehensive data set covering a decade 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.

Whether electrification affects industrial development 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 with little causal impact. For instance, recent evidence suggests that the benefits of rural electrification on households are not as large as previously thought[4]. Moreover, 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 has been studied on employment (Dinkelman (2011)) and overall development (Lipscomb, Mobarak, and Barham (2013)),

[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 (2020), Lenz, Munyehirwe, Peters, and Sievert (2017), and Burgess, Greenstone, Ryan, and Sudarshan (2019) who focus on residential electrification, Bos, Chaplin, and Mamun (2018) who provide a review, and Burlig and Preonas (2016).

but the link between electrification and firms has been mostly studied on the intensive margin, namely power shortages (e.g. Fisher-Vanden, Mansur, and Wang (2015), Allcott, Collard-Wexler, and O'Connell (2016)). 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 for long-run development. The main difference between shortages and grid expansion is that shortages are short-run changes in electricity supply, while grid expansion is a long-run change. Therefore, variation in shortages creates short-run firm responses by affecting the input price of electricity and thus intensive margin production decisions. In contrast, the possibility of a firm being connected to the electric grid is more likely to create long-run firm responses such as entry and exit.

Therefore, a mechanism through which electrification potentially affects industrial development is firm turnover - the churning of firms through entry and exit. 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 output and 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 (McCawley (1978)). In 1990, Java, the most developed and densely populated island in Indonesia, was only around 40% electrified[5]. The island has since witnessed a massive and successful government-led effort to expand access to electricity. At the same time, Indonesia experienced fast growth in the manufacturing sector. This allows me to match rich firm-level microdata 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. Large-scale randomized electrification interventions are seldom feasible both in terms of costs and practicality. This poses an empirical challenge in identifying the effect of electrification on industrial outcomes. The empirical strategy I implement exploits exogenous variation from a supply-side natural experiment and is inspired by the transportation literature (Faber (2014)) and the literature on the persistence of historical infrastructure investments. In their paper on the persistence of colonial infrastructure in Indonesia, Dell and Olken (2019) show that investments in transportation infrastructure made by the Dutch persisted to the present. I exploit how similar investments made by the Dutch colonial authorities in the electrification infrastructure have affected the evolution of the

[5]Statistik PLN 1989/1990, Indonesian Electricity Company.

electricity grid in more recent times. This is a supply-side natural experiment based on the need of the state electricity monopoly to have a single interconnected electricity grid in Java. I construct a hypothetical electric transmission interconnection that is a function of incumbent disconnected colonial power plants and the cost-adjusted distance between them. The hypothetical interconnection, therefore, abstracts from endogenous demand factors that could be driving the expansion of the grid and focuses on plausibly exogenous cost factors only. The result is an instrument defined as the Euclidean distance to the hypothetical interconnection. This strategy allows for exogenous variation in the instrument while controlling fully for local determinants of the cost of electrification. The identification assumption is that conditional on various controls, the instrument only affects industrial outcomes through its 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. I fill any gaps 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 (the lowest administrative division) to the nearest transmission substation. 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 all desas 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 allow me to look at the effect of electrification on different measures of firm performance and the composition of firms in the market.

I find that electrification causes industrial development at a local level by increasing manufacturing activity. Access to the grid increases the number of manufacturing firms, the number of workers in manufacturing, and manufacturing output. Interestingly, electrification increases firm turnover by increasing not only entry rates but also exit rates. 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. While the impact on incumbents is comparable to estimates from the literature on shortages (Allcott, Collard-Wexler, and O'Connell (2016)), a decomposition of the effect on total output shows that most of the resulting increase in output comes from entrants. This highlights the role of entry in allowing electrification to have substantial impacts on the aggregate level. These results are theoretically consistent with a decrease in the entry cost in a model à la Hopenhayn (1992), suggesting that electrification increases aggregate productivity by allowing more productive firms in the market, increasing firm turnover, and enhancing 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 Banerjee, Duflo, and Qian (2012), Faber (2014), and Gertler, Gonzalez-Navarro, Gracner, and Rothenberg (2014), Donaldson (2018)), and the role of historical infrastructure in shaping access to future infrastructure and in turn the organization of economic activity (Dell and Olken (2019)).

In terms of electrification infrastructure, a growing literature studies generally the relationship between energy and development. 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 authors look at the effect of electrification in Brazil on county-level development. Rud (2012) shows that industrial output in Indian states increases with electrification, and Fried and Lagakos (2017) examine the impact of rural electrification on migration and structural transformation in Ethiopia. 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. Fisher-Vanden, Mansur, and Wang (2015) find that firms in China respond to shortages by re-optimizing among inputs, which increases their unit cost of production without 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. Other recent examples include Abeberese (2017), Ryan (2018), and Filippini, Geissmann, Karplus, and Zhang (2020). I contribute to this literature by providing novel evidence on the effect of the extensive margin of electrification on industrialization, using comprehensive firm-level microdata.

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 for the lower productivity level of firms in developing countries, in particular, access to electricity. I contribute to this literature by linking infrastructure to selection 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 the main results on the effect of electrification on local industrial outcomes. I evaluate how electrification affects firm performance in Section 5. Finally, Section 6 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 Dutch colonization of Indonesia, access to electricity was unequal and mainly reserved for 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 inefficiency, poor operating conditions, and inadequate expansion (McCawley (1971)). This in turn led to a large electric supply deficit, which meant low household electrification ratios and that businesses and industries had to rely on self-generation. Power supply in Indonesia was poor even relative to other countries with a similar GDP per capita 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[6]. 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

[6]Source: World Bank.

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 consisted 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 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 than the marginal price of electricity from self-generation. In 1989, the level of electricity consumption per capita was still low in Indonesia (137.5 kilowatt-hour (kWh)) relative to other countries at the same development level and its neighbours[7].

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 an 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[8]. 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

[7]Malaysia 1,076 kWh, India 257 kWh, Philippines 361 kWh, and Thailand 614 kWh. Source: IEA Statistics 2014

[8]Source: Official planning documents.

1990. Electricity was mostly concentrated in the capital city of Java, Jakarta, but also the cities Bandung, Yogyakarta, 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

To evaluate the impact of electrification on industrial development in Java, I have constructed a new panel data set on 24,824 Javanese desas[9], the lowest administrative division[10] 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.

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 densely populated 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 E.1, and paper maps, Figure 3a and Figure 3b) 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 the 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 Open Street Maps. The resulting data set is a panel of all transmission substations in Java. Figure 4 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 1a. In 1990, the number of substations was 153. By 2000, there was a total of 320 transmission substations in Java. Total electricity transmission capacity increased from $8,426$ MVA to $25,999$ MVA, more than triple.

[9] Like a county.

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

[11] Source: author's calculations.

To know precisely which desa had access to the grid in a certain year, I would require data on the distribution network, which is not available. Using the digitized electrification data at the transmission substation level, I define access to the grid as a dummy variable equal to 1 if the desa is within 15 km[12] of the nearest transmission substation in year t. I validate this definition of access after introducing the outcome data in the next section.

3.2 Industrial Outcomes

I start the empirical analysis by looking at the effect of access on desa-level manufacturing outcomes. To get information on manufacturing activity in desas, I use the Indonesian Annual Census of Manufacturing, an unbalanced panel of all manufacturing firms in Indonesia with 20 or more employees, where I observe the location of the firm at the desa level[13]. I restrict the analysis to firms located in the contiguous land of Java[14], which constitute around 80% of all Medium and Large firms in Indonesia. This allows me to create variables such as the number of manufacturing firms, number of manufacturing workers, and total manufacturing output in each desa. The resulting data set is a yearly balanced panel of all desas in Java from 1990 to 2000. Table 1b presents desa-level summary statistics. On average, around 60% have access to the grid over the sample period. The average distance to a transmission substation is around 14 km. The average number of medium or large firms per desa is around 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 or rural. Conditional on having a positive number of firms, the average number of firms per desa is around 4.5 firms. The last row of Table 1b shows that there is substantial variation in how large these desas are in terms of population. The final total number of desas per year used in the analysis is around $23,000$.

I use the information on desa-level characteristics from the Desa Potential Statistics (PODES) survey for 1990, 1993, 1996, and 2000. PODES is a survey of Indonesian desas containing information about population, political and legal characteristics, and most importantly, infrastructure availability. These include information on the type of infrastructure available in the desa such as railway, motor station, and airport. In addition, I use GIS data on land elevation and roads in Java. I measure the Euclidean distance from the desa (centroid) to the nearest regional road and the nearest electric transmission substation. This data is used to construct a digital map of desas in Java with various

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

[13]This is from a supplementary data set to the Indonesian Census of Manufacturing acquired from BPS.

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

desa-level characteristics over time.

I then turn to the census of manufacturing and analyze the effect of access to electricity on firm-level outcomes. Table 2 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 Indonesian manufacturing over the sample period. The largest three industries in 2000 are food and beverages, textiles, and furniture. Between 1990 and 2000, the total number of manufacturing firms in Java has increased by more than 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.

Table 3a shows that 85% of firms in the census are located in a connected desa and that the average (median) distance to the nearest substation is 8.5 (4.5) km. Compared to the average connectivity and the average distance to substations at the desa level (Table 1b), average connectivity in the desas where firms are located is significantly higher, and the distance to the nearest substation is smaller. This highlights the fact that firm location is an endogenous outcome, and that the availability of electricity is one important determinant of that outcome.

I use the census data to validate the measure of access defined based on the distance from the desa centroid to the nearest transmission substation. Table 3b shows the average firm size and energy use variables for firms in connected and unconnected desas from the census. $Access_{vpt}$ is an indicator variable equal to 1 if the desa where the firm is located is within 15 km of the nearest transmission substation. On average, firms in connected desas produce output 5.7 times the output of firms in unconnected desas and employ more workers. Since connected firms are larger, they consume more electricity from all sources. In terms of electricity use, firms in connected areas consume more than 7 times as much grid electricity and spend 9 times as much on grid electricity. This indicates that $Access_{vpt}$ measures access to the true grid well. Both connected and unconnected firms own generators. While the share of firms owning a generator in connected desas is higher (32% and 21%) and these firms generate more electricity (perhaps due to the larger scale of these firms), the capacity of generators owned by firms in unconnected desas is 3 times larger. In addition, firms in unconnected desas generate a larger share of their electricity, have lower total electricity to output ratio, but a higher ratio of electricity generated to output. Firms connected to the grid do own generators, but these generators are used proportionately less and are significantly smaller in capacity, suggesting that they are mainly for backup purposes.

I then look at how electricity use changes at the desa level around the time the desa gets access according to this definition. I estimate the following event study-style specification for outcome Y of a desa v in province p and year t:

$$Y_{vpt} = \alpha + \sum_{s=-6}^{s=6} \beta_s \mathbb{1}(T_{vpt}^{Access} = s) + \theta_v + \boldsymbol{V'}_{vpt}\boldsymbol{\eta} + \gamma_{pt} + \epsilon_{vpt} \tag{1}$$

where T_{vpt}^{Access} is the normalized year of access, θ_v is a desa fixed-effect, $\mathbf{V}_{vpst}$ is a vector of desa-level controls, which I list in detail in Section 3.3.1, and γ_{pt} are province-by-year fixed effects. Since I also control for year effects, I choose the reference category to be the period before access ($s = -1$ on the x-axis). I restrict the sample to the switching desas that received a transmission substation within 15 km between 1991 and 1999, relative to the control desas that remained unconnected in 2000. Figure 5 presents the corresponding event study graphs. The variables on the Y-axes are electricity use variables aggregated from the firm-level census to the desa level, to ensure a balanced panel as described above. Figure 5a shows the total quantity of grid electricity consumed by all firms in the desa in kWh. Before access, there is no difference in the average quantity of grid electricity consumed in total between the two groups of desas. Post access, the total quantity of electricity consumed increases in electrified desas. Similarly, Figure 5b shows total grid electricity spending in IDR at the desa level, which becomes higher in switching desas after access. On the other hand, Figures 5c and 5d show respectively that the quantity of electricity in kWh from self-generation and total spending on fuels for the generator at the desa level decrease after access. Therefore, the definition of access based on the distance to transmission substations corresponds well to changes in patterns of electricity use. Specifically, firms consume more grid electricity and generate electricity less often in the post periods, validating the definition of access.

3.3 Empirical Strategy

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

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

where $\Delta x_{vpt} \equiv x_{vpt} - x_{vpt_0}$, and $Access_{vpt}$ is an indicator variable equal to one if desa i is connected to the grid in year t. I choose $t_0 = 1991$ as the baseline year to maximize sample size. The final sample includes desas already connected in 1991, desas that switch between 1992 and 2000, and desas that remain unconnected by 2000. Standard errors are clustered two-way at the desa level and the province-by-year level. Clustering at the desa level, the level at which the treatment [15] varies, is to account for serial correlation over time. Clustering at the province level accounts for spatial correlation in access. In appendix Table E2, I present results with alternative correction for spatial inference

[15] and instrument, defined later.

following Conley (1999).

The differenced specification nets out all location-specific characteristics that could simultaneously drive industrial outcomes and electrification. In addition, β captures the average effect of access including any dynamic effects if for example, the impact of electrification grows over time (as we can see in Figure 5), since different desas get connected in different years. Equivalently, Figure 6 plots the event study specification (1) with log industrial output on the left-hand side for the sample of switching desas that received a connection between 1991 and 1999 relative to desas unconnected in 2000. First, there is no statistically significant difference in average industrial output between treated and control desas before the time of access, alleviating concerns of differential trends. Second, the impact of access on average industrial output is positive, and it appears to grow over time.

However, interpreting the OLS estimation of equation (2) as causal would require the additional strong assumption that the grid was randomly assigned to desas. Electricity grids are placed endogenously to industrial outcomes. In Indonesia, the expansion of the grid is demand-driven. Even in the absence of differential trends ex-ante, it is possible that PLN connected desas that are expected to grow differentially in response to electrification for reasons that are unobservable to the econometrician. In addition, the timing of connection might be correlated with potential outcomes. 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. Importantly, this methodology implies that the bias in OLS 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 the cost of construction, which potentially creates exogenous variation that is useful for identification.

To answer the question of how large-scale electrification impacts the industrial sector, the ideal experiment would be to randomize electrification at the regional level across the whole economy, for example, the desa level. This design would be suitable to capture extensive margin effects of electrification such as the entry and exit of firms. However, it is often infeasible and extremely costly to get utilities to randomize the placement of substations and the placement of the grid across regions, especially at a large scale. An alternative design that is more feasible would be an RCT at the firm level, analogous to the design in Lee, Miguel, and Wolfram (2020). However, a firm-level RCT can miss important populations of interest and important margins of industrial development. First, it is possible that in locations where connections are feasible and assignable in an RCT, some firms would be always takers and the RCT would not capture effects for these firms, yet, these are the firms that are likely to benefit most from electricity. Second, an important part of the impact concerns firms that will enter after electrification and that would not be part of the experimental sample. Third, electrification is likely to have effects on competition and growth that are hard to capture with a firm-level RCT.

To deal with endogeneity, I propose to estimate equation (2) by two-stage least squares (2SLS) exploiting a supply-side natural experiment. In 1969, the electricity grid in Java consisted of 5 different disconnected grids across the island (Figure 3a). Having disconnected grids is inefficient, prevents load sharing across regions, and increases the price of supplying electricity. Therefore, the 1970s and the 1980s witnessed a huge and successful effort by PLN to connect the various grids on the island (Figure 3b). Various transmission lines were built for the main purpose of interconnecting the grids (Figure 3). As a result, desas nearby the lines connecting the grids faced a positive shock to the probability of receiving access to electricity 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 from others, such as non-farming land, I create a hypothetical version of the green line. This hypothetical version abstracts from demand factors, only taking as given the location of incumbent infrastructure and the cost-adjusted distance between them. 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[16]. The hypothetical interconnection is essentially an instrument for the actual transmission interconnection (which I do not observe) that abstracts from demand factors driving the location of the actual grid. I describe below how this instrument is constructed.

The hypothetical least cost interconnection takes local characteristics[17] in its cost function and has the exclusive objective of minimizing the total cost of connecting the

[16]See Figure 3a. 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).

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

grid. I construct it as follows:

1. For each location on the map, I assign a cost value based on a digital elevation model[18]
2. I calculate the least cost path for each pair of power plants based on the cost data.
3. I use Kruskal's algorithm[19] 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 interconnection. The straight line distance to this interconnection labeled Z_v, is then used as the instrumental variable.

3.3.1 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. Recall that specification (2) nets out time-invariant location-specific characteristics such as local geography[20] (e.g. slope, elevation, proximity to the coast, agricultural suitability, availability of natural resources, etc.), including the cost value taken as an input in the hypothetical least cost interconnection.

One concern is that the location of the incumbent power plants is endogenous. To alleviate this concern, I exclude desas within a 10 km radius of power plants. In the appendix (Table E2), I exclude a 20 km radius as a robustness check. I also control for proximity to the nearest origin power plant to deal with the mechanical correlation between the instrument and the nearest origin power plant.

Typically, different types of infrastructure are correlated, for example, the electricity grid and the road network. In all specifications, 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 stations, motor stations, and airports. For political economy concerns that could be correlated with electrification, I control for the 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 an appointed civil servant. The legal status of the desa (such as transmigration settlements) refers to whether the desa is

[18]Cost is simply equal to the value of the land slope. The slope is an important cost factor in grid construction as steeper terrains require generally more expensive transmission equipment (such as more tower units and more dead-end towers). In addition, transmission towers in mountain ridges involve expensive investments in electrical protection equipment, for example, against lightning.

[19]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.

[20]In appendix Table E2, I control for the slope value used as the cost input in the least cost path calculation.

formed under a Government decree, Ministerial decree, Regency decree, or other. I also control for the desa's level of development as classified by the Government of Indonesia[21].

A concern is that some unobserved past persistent shocks to industrial desa-level outcomes (see for example Dell and Olken (2019)) that are also correlated with distance to the hypothetical least cost interconnection are driving the estimated effects. Therefore, I control for the pre-sample period number of firms in the desa in 1990.[22]

To summarize, desa controls include distance to the nearest origin power plant, and distance to the nearest road, various infrastructure availability dummies, political status, legal status, and the number of firms in 1990. Table 1c presents a summary. The identification assumption is that conditional on desa fixed effects, province-by-year fixed effects, and the various desa controls, the distance to the hypothetical least-cost interconnection is not correlated with the potential outcomes of desas.

It is important to note that this strategy does not use the variation in the distance to the nearest origin power plant, instead, it exploits the variation in the distance to the interconnection lines *between* the origin power plants. But what variation is left in the distance to the hypothetical interconnection after adding all the above controls? In other words, conditional on local characteristics, why is it possible to still have two desas with different distances to the hypothetical least cost interconnection? The answer is that a crucial determinant of the hypothetical least cost interconnection is *global* geography, as opposed to the local geography. This is because the algorithm described above has the objective of minimizing the cost of the transmission interconnection, taking the location of the incumbent power plants as given. This creates variation at a relatively low geographic level in the distance to the hypothetical interconnection for locations within the same province and with similar local geographic characteristics.

3.3.2 First Stage

Table 4 shows the first stage regressions. The dependent variable in the first two columns is $\Delta Access_{vpt}$. The coefficient in Column (1) is negative and significant, indicating that the further away a desa is from the hypothetical interconnection, the less likely it is to receive access to electricity between 1992 and 2000. The estimate is stable and changes only slightly after adding the controls in Column (2). Z_v is significantly negatively correlated with access, suggesting that even after controlling for the location of the incumbent power plants and other local characteristics, and conditional on desa effects, the global configuration of historical infrastructure across space has predictive power over which desas get connected between 1992 and 2000. The first stage F-statistic is stable across columns (1) and (2) and is high enough to guarantee relevance, avoiding weak

[21]There are three categories: Swadaya (traditional), Swakarya (transitional), and Swasembada (developed).

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

instrument bias. In Column (3), I show the correlation between Z_v and the change in the distance to the nearest substation between 1991 and 2000, the variable upon which the access definition is based. The coefficient in Column (3) indicates that being further away from the hypothetical interconnection is correlated with a smaller decrease in the distance to the nearest transmission substation.

4 Impact on Industrialization

I start by examining the effect of electrification on industrial outcomes at the desa level. 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, I relate and reconcile the results with findings from the literature.

4.1 Effect of Electrification on Local Industry

I estimate the effect of the grid arrival on the number of manufacturing firms, manufacturing employment, and manufacturing output (IDR) 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 the first two outcomes instead of the log, and a logarithmic transformation of manufacturing output as it is in nominal values for ease of interpretation (See Table E2 in appendix E for results with zero-preserving log transformations, and in levels for output.). Table 5 shows the results for three desa-level outcomes following specification (2).

Panel A presents the IV results. The IV estimates are positive and statistically significant at conventional levels. The coefficient in Column (1) in Panel A says that the causal effect of grid access on the number of firms in a desa is an increase of around 1 firm. Considering that the average number of firms per desa in the sample is 1, this effect is significantly large and around 100% of the sample average (bottom Panel). Similarly, for the number of workers and manufacturing output, the IV estimates in columns (2) and (3) 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, 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 columns (2) and (3) should be interpreted as the causal impact of electrification on the number of workers and output in medium and large manufacturing.

Across all outcome variables, the OLS estimates in Panel B are small in magnitude and are not statistically significantly positive. Recall that this is the full sample of desas, including desas already connected in 1991, desas unconnected in 2000, and desas that switched in between. This OLS estimate is the average effect of access on the change in outcomes for the switchers, relative to desas already connected at baseline and desas unconnected by the end of the sample. It is clear from Figure 6 (which excludes the sample of desas connected at baseline) that the impact of electrification is dynamic and grows over time. Therefore this OLS estimate is biased downwards. 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.

The reduced-form results are presented in Panel C and show that being closer to the hypothetical interconnection causes an increase in industrial activity. Next, I discuss the difference in magnitudes between the IV estimates and the OLS estimates.

First, the validity of a 2SLS 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. This means that the distance to the hypothetical interconnection, conditional on desa effects and other controls, only affects industrial outcomes through its effect on access to the actual grid. A violation of this assumption could lead to a large difference between the OLS and the IV estimates. 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. Recall that the differenced specification nets out all time-invariant desa characteristics such as local geography or other determinants of industrial activity (e.g. historical trade routes). In addition, I include an extensive set of controls for potential time-varying confounders as outlined in Section 3.3. These include infrastructure controls and other political and economic characteristics. In addition, when adding these controls to the first stage regression (Table 4 columns (1) and (2)), the point estimate of the relationship between the instrument and the treatment variable and the corresponding first stage F-statistic remain stable. Given this rich set of controls, it is unlikely that a violation of the exclusion restriction is driving the difference in magnitudes between the IV and OLS estimates.

The second reason, which is the most likely reason, is a compliers' issue. Since I estimate a LATE of access on industrial outcomes, this difference in magnitudes is potentially driven by a sub-population of desas that would benefit *more* from electrification. Table E1 in the appendix provides evidence in favor of this possibility and suggests that complier desas have higher growth potential in response to electrification. In addition, 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 be that these desas will experience higher returns to electrification.

The third 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[23].

Now that I have discussed reasons for the difference between OLS and IV estimates, I ask whether the IV estimates are sensible. Are the IV estimates too large, irrespective of how they compare to the OLS estimates? Looking at the bottom two rows of Table 5, 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 Panel A Columns (2) and (3) to the average outcomes in a desa with a positive number of firms. The effect on manufacturing labor is around 65% of the effect of moving from a non-industrialized desa to an industrialized desa. The effect on manufacturing output is around 33% 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 location to entrepreneurs who are considering opening a factory. As shown in Section 4.1, electrification causes the total number of firms in a desa to increase. I next investigate the effect of electrification on the extensive margin of industrialization to uncover 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 sign of efficiency in the market where more productive businesses replace less productive ones. In appendix C, I present a theoretical model à la Hopenhayn (1992) of how electrification can affect firm turnover. I now turn to the role of entry and exit as drivers of industrialization.

Table 6 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. Con-

[23] There is a fourth possible reason, which is a technical one that is common in two-stage least square (2SLS) strategies with a binary endogenous variable. If the first stage of the 2SLS estimation predicts values of 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 (Source: author's calculation.).

sistent with specification (2), I use the *cumulative* number of entrants E_{vpt} and exiting firms X_{vpt} between year t and 1990. We therefore have:

$$N_{vpt} = N_{vp1990} + E_{vpt} - X_{vpt}$$
$$\Rightarrow \Delta N_{vpt} = \Delta E_{vpt} - \Delta X_{vpt} \quad (3)$$

where $\Delta x_{vpt} \equiv x_{vpt} - x_{vp1991}$ as before. Focusing on the IV estimates in Panel A, Column (1) shows that electrification increases the average number of entrants in the newly electrified desas. Column (2) shows the effect on the number of exiting firms, which is also positive and significant. Subtracting the coefficient in Column (2) from the coefficient in Column (1) gives exactly the effect on the total number of firms in Table 5 Column (1) as in equation (3). Columns (3) and (4) look at the effect of access on firm turnover. Consistently with the definitions of entry and exit, I define the rates as the cumulative entry and exit rates. The first outcome is the cumulative entry rate, defined as the cumulative number of entrants between 1990 and t divided by the total number of firms in t. The second outcome variable is the cumulative exit rate, defined analogously for the cumulative number of exiting firms. As before, the dependent variables in (3) and (4) and the difference in rates between t and 1991. These outcomes are only defined for desas with a positive number of firms in t. Since these are cumulative rates, the rate can be larger than one, and the sample averages are 28% and 22% respectively. The IV estimates in Panels A Column (3) show that access to the grid increases firm entry rate by around 38 percentage points. Interestingly, in Column (4), the exit rate *also* increases due to electrification. Therefore, electrification increases firm turnover, leading to more churning in a given desa. Higher churning is a sign of efficiency where firm selection into and out of the desa is at work.

The results on exit in Table 6 could be driven purely by the increased entry of firms in response to electrification. It is a well-documented fact in the literature (e.g. Evans (1987)) that younger firms have a higher probability of exit. In other words, the increase in the exit of firms could be merely driven by the natural churning of new firms that are trying things out, instead of electrification driving unproductive firms out of the market. To address this issue, I estimate the effect of electrification on the number of exiting firms in multiple age groups. I divide the age of firms into deciles and regress the number of exiting firms in age bin a on access, instrumented with Z_v as before in specification (2). Figure 9 shows the results. Each estimate, along with the 95% confidence interval, corresponds to a separate regression as in Table 6 Column (2), with the change in the number of exiting firms in age bin a on the left-hand side. The effect of electrification on exit is largest for the younger firms, indicating that part of the effect of electrification on exit is a result of the higher entry rates, as young firms have a higher probability of exit. However, electrification increases exit across the whole age distribution to varying degrees and there are firms of age 10 years and older exiting as a result of electrification. This confirms that the increase in exit in response to electrification is not purely driven by the increase in the entry of firms, but electrification is also causing older firms to exit more often.

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 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 Extensive Margin Effects and Estimates from the Literature

Previous estimates from the literature on the impact of the intensive margin of electrification on manufacturing are modest. Specifically, Allcott, Collard-Wexler, and O'Connell (2016) find that electricity shortages affect negatively the revenue of incumbents, but not so much their productivity. On the other hand, estimates of the impact of the extensive margin of electrification on manufacturing are quite large. For instance, Rud (2012) finds that a 1 standard deviation increase in the measure of electrification for agriculture increases state-level manufacturing output by 14% in India. These estimates are large and economically substantial. Papers studying the impact of the extensive margin on various development outcomes such as Lipscomb, Mobarak, and Barham (2013) also find large positive effects.

In this section, I discuss how the extensive margin, entry in specific, leads to larger effects of electrification. Allcott, Collard-Wexler, and O'Connell (2016) analyze the impact of shortage in the *short run* on incumbents and show that shortages have a significant economic cost on these firms. In their paper, the authors explicitly state that the year-to-year variation in shortages they study is unlikely to affect extensive margin outcomes such as entry and exit. On the other hand, in this paper, I study the impact of a longer-term change in electrification, access to the grid. Access to the grid has the potential of affecting extensive margin decisions of firms such as entry and exit, which can impact the manufacturing sector substantially.

First I decompose the total impact of electrification on manufacturing output due to changes in entry, exit, and incumbents' output. The change in total manufacturing output in desa d and year t can be decomposed into the output of incumbent firms that existed in 1991 and survived until year t, the output of entrants between 1992 and t, minus the output of exiting firms that existed in 1991 but did not survive until year t:

$$
\begin{aligned}
Y_{d1991} &\equiv \sum_{i\in d} y_{i1991} = \sum_{i\in I_t} y_{i1991} + \sum_{i\in X_t^C} y_{i1991} \qquad \forall t \\
Y_{dt} &\equiv \sum_{i\in d} y_{it} = \sum_{i\in I_t} y_{it} + \sum_{i\in E_t^C} y_{it} \\
\Rightarrow \Delta Y_{dt} &\equiv Y_{dt} - Y_{d1991} = \Delta Y_{dt}^{I} + Y_{dt}^{E^C} - Y_{dt}^{X} \qquad (4)
\end{aligned}
$$

where I_t is the group of incumbent firms that existed in 1991 and still existed in year t, E_t^C is the set of firms that entered between 1992 and t, and X_t^C is the set of firms that exited the market between 1991 and $t-1$.

I estimate the impact of electrification on each of the components of output (in levels) in (4) by 2SLS using equation (2). The results are in Figure 10. The effect on total output in levels is positive and significant as can be seen in the blue bar. The impact on each of the components of the change in output is presented in the green bars. The point estimates are normalized by the effect on total output. 88% of the increase in output is due to entry highlighting the importance of the extensive margin of industrialization. Almost 4% of the effect is driven by the exit of firms, which is unsurprising as exiting firms are typically smaller towards the end of their life. Finally, 8% of the change in output is due to an increase in the output of incumbents, although the effect is not precisely estimated and is statistically indistinguishable from zero. Without the possibility of entry, the impact of electrification on industrialization would be modest. This highlights the importance of the extensive margin of industrialization in realizing the benefits from large-scale infrastructure investments such as electrification.

Next, I examine whether the impact on industrial activity is indeed a result of *new* industrial activity. The results in the previous section indicate that electrification increases industrial activity at the desa level by attracting more firms. 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 the creation of new economic activity; meaning that the aggregate effect of electrification is smaller than the estimated effects at the desa level.

In the context of Java, even if a desa was far from the grid in a certain year, 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 overly 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. Most importantly, the results from Section 4.2 show that in fact, exit in electrified locations is larger than exit in locations without electricity. This goes in the opposite direction of electrification inducing incumbent firms to relocate from non-electrified areas to electrified areas.

In addition, if negative spillovers are significant, then theoretically the effect of electrification on manufacturing outcomes should be smaller at the district level[24]. I look at how electrification affects the district-level outcomes and compare them to the desa-level outcomes.

I estimate Equation (2) at the district[25] level, one administrative division higher than a desa:

$$\Delta Y_{dt} = \alpha + \beta \Delta Access_{dt} + \boldsymbol{V'}_{dt}\boldsymbol{\eta} + \gamma_{pt} + \epsilon_{dt} \tag{5}$$

where $Access_{dt}$ is equal to one if the district centroid is within 15 km of the nearest transmission substation. I include the same controls as before in vector $\mathbf{V}_{dt}$. I also include the baseline number of firms calculated at the district level in 1990. $\Delta Access_{dt}$ is instrumented with the distance from the centroid of the district to the hypothetical interconnection Z_d. I exclude districts within 10 km of an incumbent power plant. Standard errors are clustered two-way at the province-year level and at the district level.

Figure 11 shows the effect of electrification at the district level compared to the estimates at the desa level from Table 5. The district-level outcomes are the mean value of Y across all desas in the district for ease of comparison. There is no statistically significant difference in the effect of access on the number of firms, number of manufacturing workers, or manufacturing output at district and desa levels. The effects are comparable across the two levels of aggregation and are slightly larger at the district level than those at the desa level. This suggests that there is a significant aggregate effect of electrification in districts.

5 Electrification and Firm Performance

So far, results show that electrification caused an increase in manufacturing activity and firm turnover in Java. A change in firm turnover could mean a change in the composition of firms in the industry. In this section, I make use of the firm-level manufacturing census and I analyze the effect of access at the desa level on firm outcomes. As for performance measures, I consider firm sales and inputs, and revenue productivity. I structurally estimate firm revenue productivity following Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015) using energy as a proxy. I explain this methodology in detail in the Appendix D.

Estimating equation (2) requires that all units exists in the baseline year. The census of manufacturing is an unbalanced panel where firms enter and exit in different years.

[24]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.

[25]Kecamatan in Indonesian

Since it is not possible to observe all firms in the baseline year, an equivalent specification to equation (2) that captures the effect of electrification on all firms, including entrants and incumbents that did not exist in 1991, is not suitable. I therefore resort to a cross-sectional specification as close as possible the identification strategy at the desa level. First I focus the analysis on the firms in switching desas by excluding desas that were already connected in 1990 to capture the effect on switchers as in the desa-level analysis. The results with the full sample is in the appendix Table E3. I estimate equation (6) for different firm level outcome variables of firm i located in desa v, province p, sector s in year t:

$$y_{ivpst} = \alpha + \beta Access_{vpst} + \boldsymbol{X'}_{ivpst}\boldsymbol{\mu} + \boldsymbol{V'}_{vpst}\boldsymbol{\eta} + \gamma_{pt} + \delta_{st} + \epsilon_{ivpst} \tag{6}$$

where δ_{st} are 3-digit industry-by-year fixed effects, $\boldsymbol{X}_{ivpst}$ is a vector of firm controls including export, generator and ownership dummies[26]. The vector $\mathbf{V}_{vpst}$ include the desa level controls as before, including desa-level baseline number of firms. In the appendix Table E2, I present the results from an analogous cross-sectional specification at the desa level. Results are are similar to the differenced specification in (2), suggesting that not controlling for time-invariant characteristics does not change the conclusions substantially.

Table 7 shows the IV, OLS, and reduced-form estimates of specification (6) for the log values of firm-level inputs[27]. The treatment variable is $Access_{vpt}$, defined at the desa level and instrumented with Z_v. The IV estimates in Panel A show that electrification causes an increase in average firm inputs. Columns (1) and (2) present the impact on labor, measured by the wage bill and the number of workers employed at the firm respectively. The impact on labor is not statistically significant at the 5% level but it is positive and around 35%. On the other hand, the estimates for the effect on capital and materials in Columns (3) and (4) are positive, large, and statistically significant at the 5% level. These results imply that the capital-labor ratio of firms in connected desas is on average higher. If capital and energy are more complementary than labor and energy, a decrease in the price of electricity can cause the firm to re-optimize its input choices by substituting away from labor towards capital. The fact that the IV estimates are larger for capital than for labor is reassuring: this is an indirect test for the exclusion restriction of the instrument. If the instrument affects access through for example local geography or access to roads, there's no obvious theoretical reason why the effect on capital should be larger than the effect on labor. However, complementarity between capital and electricity would cause this response in firms if the price of electricity goes down as a result of having access to grid electricity (relative to self-generation), indicating that indeed the instrument is affecting outcomes through its effect on access.

Columns (5) to (8) show the impact of access on energy use. Firms in connected desas spend more on energy (includes spending on electricity and fuels), less on fuels for the

[26]There are four types of ownership: domestic, foreign, central government, and local government.
[27]Nominal values are deflated

generator, consume a higher quantity of electricity (grid and self-generation) in kWh and have a lower self-generation share out of the total quantity of electricity consumed. The impact on the energy bill in Column (5) is smaller than the impact on the electricity quantity in Column (7), indicating that the unit price of electricity is lower in connected desas. This is confirmed by the estimate in column (8), where connected firms generated less of the electricity they consume[28]. These patterns confirm two things. First, the desa-level definition of access, although it might be measured with error (and hence the attenuation bias in the OLS estimates in Panel B), corresponds to changes in how firms use electricity in their production when they get connected. Second, the results increase the confidence in the instrument as the relative impacts suggest that the distance to the hypothetical interconnected affects outcomes through its effect on access, as can be seen in the reduced form in Panel C.

Table 8 shows the impact of electrification on output (deflated sales) and revenue productivity (TFPR). The first 4 columns show the impact on output for the sample as in Table 7 in Column (1), for entrants in Column (2), exiting firms in Column (3), and incumbents in Column (4). I define an entrant to be a firm that is 3 years old or younger. Exit is defined as the set of firms in the last three years they appear in the census[29]. Focusing on Column (1), firms in connected desas produce double the output as those firms in unconnected desas. Relative to the existing literature, the most readily comparable result is from Allcott, Collard-Wexler, and O'Connell (2016). In their paper, the authors look at the effect of shortages on firm-level outcomes for incumbents only. 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 110% increase in sales revenue. In comparison, the point estimate in Column (4), although not significant due to the weak reduced form, implies a 105% increase in the incumbents' output, very similar to the estimated effect by Allcott, Collard-Wexler, and O'Connell (2016). The estimate in Column (1) implies a 160% increase in average output in connected desas, meaning that in addition to the within-firm effect of electrification on sales, there are large extensive margin effects. This can be seen in Column (2) where the largest increase in output can be seen for entrants. In addition, exiting firms are larger. Taken together, these results confirm the desa-level decomposition results in Figure 10.

Columns (5) to (8) show that the average TFPR or profitability is higher in connected desas (Column (5)). The tougher selection environment, including the increased exit of

[28]Connected firms might still use generators as back-up since supply is unlikely to have been 100% reliable during the sample period.

[29]I use this wide definition of entry and exit to maximize the sample size. Restricting the definition of entry to age zero and exit to the year the firm exists results in a small sample size for each group and potentially insufficient statistical power. Results are qualitatively similar and are presented in the appendix (Table E.4) for completeness. Alternatively, results in Table 8 can be interpreted as the impact on younger firms and on firms closer to their exit.

unproductive firms, leads to higher average productivity in the market. The increase in productivity is mostly driven by entrants (Column (6)) who are more productive in connected desas relative to entrants in unconnected desas. Exiting firms are also more productive although the point estimate is not significant. Incumbents in connected desas have higher revenue productivity than in unconnected desas. Recall that TFPR measures a combination of physical productivity and price. The increase in average revenue productivity could be an increase in profitability driven by a decrease in price in response to electrification through the lower input price of electricity. The estimates in Columns (2), (3), (6), and (7) show that on average, entrants and exiting firms in connected desas are larger and more productive. This is consistent with a tougher selection environment. The model in appendix C shows that the marginal firm is more productive (and therefore larger) when entry costs decrease, because of tougher competition. This makes it more difficult to survive and improves average outcomes in the industry. The extensive margin of electrification, through selection, causes the average firm size and profitability in the market to increase.

6 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 (Allcott, Collard-Wexler, and O'Connell (2016)). 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 by the increased entry and exit of firms, that drives 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 (2020) 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 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.

Burgess, R., M. Greenstone, N. Ryan, and A. Sudarshan (2019). Demand for electricity in a poor economy.

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.

Conley, T. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Dell, M. and B. A. Olken (2019, 03). The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java. *The Review of Economic Studies 87*(1), 164–203.

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.

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.

Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The journal of industrial economics*, 567–581.

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

Filippini, M., T. Geissmann, V. J. Karplus, and D. Zhang (2020). The productivity impacts of energy efficiency programs in developing countries: Evidence from iron and steel firms in china. *China Economic Review 59*, 101364.

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.

Fried, S. and D. Lagakos (2017). Rural electrification, migration and structural transformation: Evidence from ethiopia. *Unpublished manuscript, University of San Diego*.

Fried, S. and D. Lagakos (2019). Electricity and firm productivity: A general-equilibrium approach. *Unpublished manuscript, University of San Diego*.

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*.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*(5), 1127–1150.

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

Lee, K., E. Miguel, and C. Wolfram (2020). Experimental evidence on the economics of rural electrification. *Journal of Political Economy 128*(4), 1523–1565.

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.

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.

Rovigatti, G. and V. Mollisi (2016, September). PRODEST: Stata module for production function estimation based on the control function approach. Statistical Software Components, Boston College Department of Economics.

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

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.

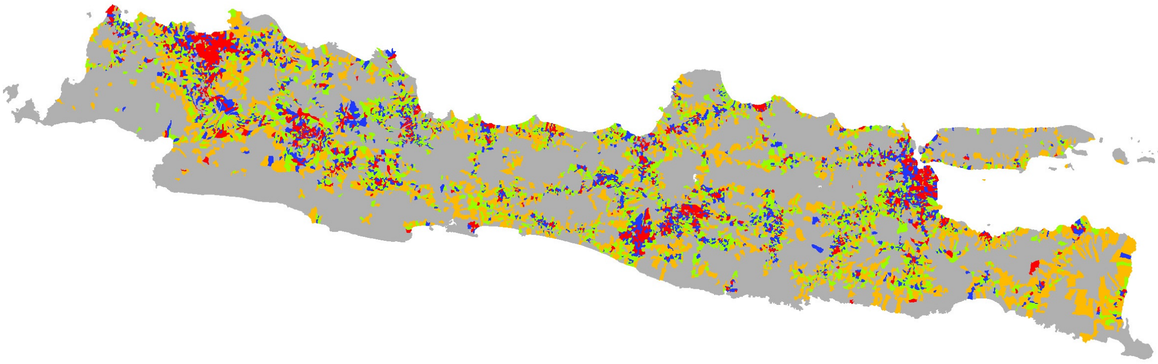
Tybout, J. R. (2000, March). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature 38*(1), 11–44.

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

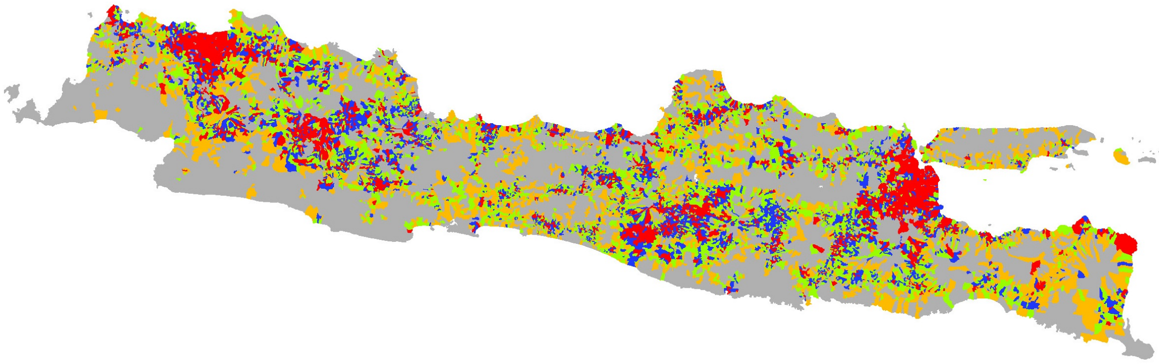
A Figures

Figure 1: Desa-Level Electrification Ratios 1990 to 2000.

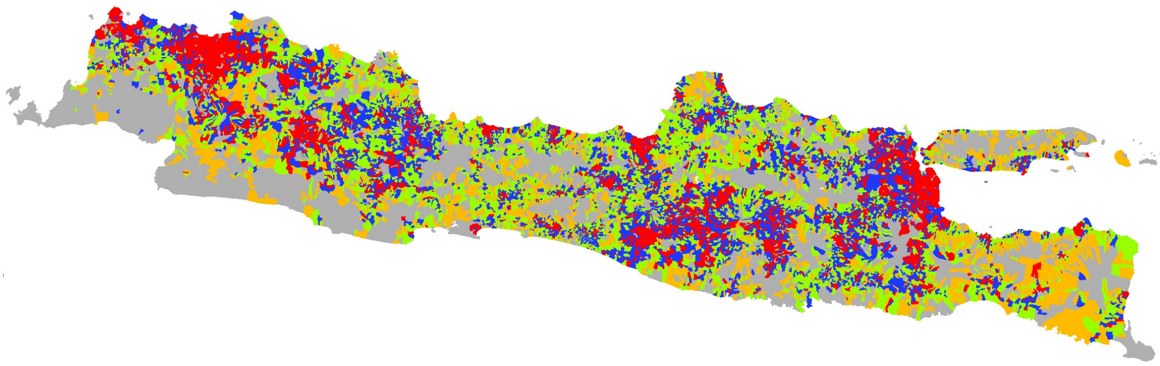
(a) 1990



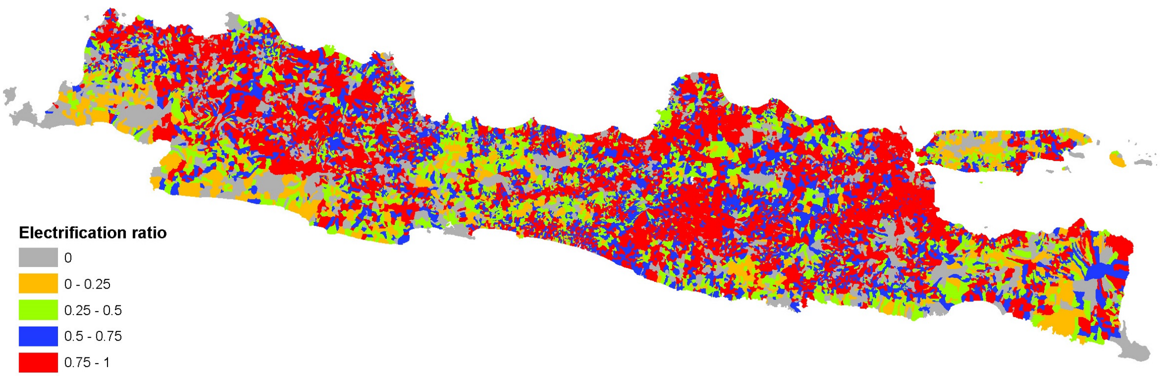
(b) 1993



(c) 1996



(d) 2000



Source: PODES, BPS

Figure 2: Example of Inventory Table of Transmission Transformers.

Sektor : *Madiun*

Hal : 3 / 3

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

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

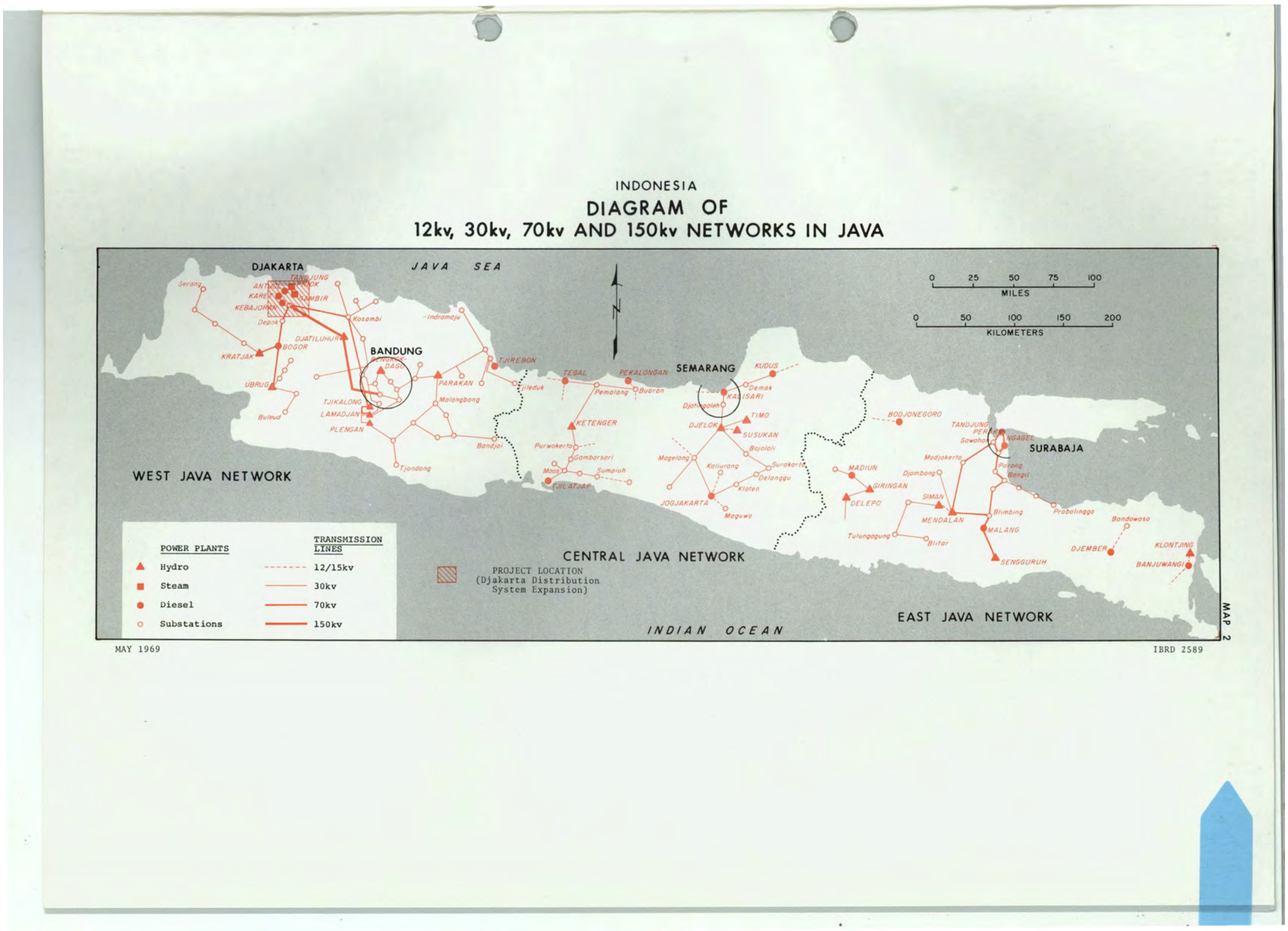
25 kV = 50.73 kms
70 kV = 846.16 kms
150 kV = 703.62 kms

10/05/01 16:34:57

Asset/OMTrans

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

Figure 3: Evolution of Electric Transmission Network in Java. Source: World Bank



(a) Java Network 1969

(b) Java Network 1989

Figure 4: Expansion of the Grid 1990-2000

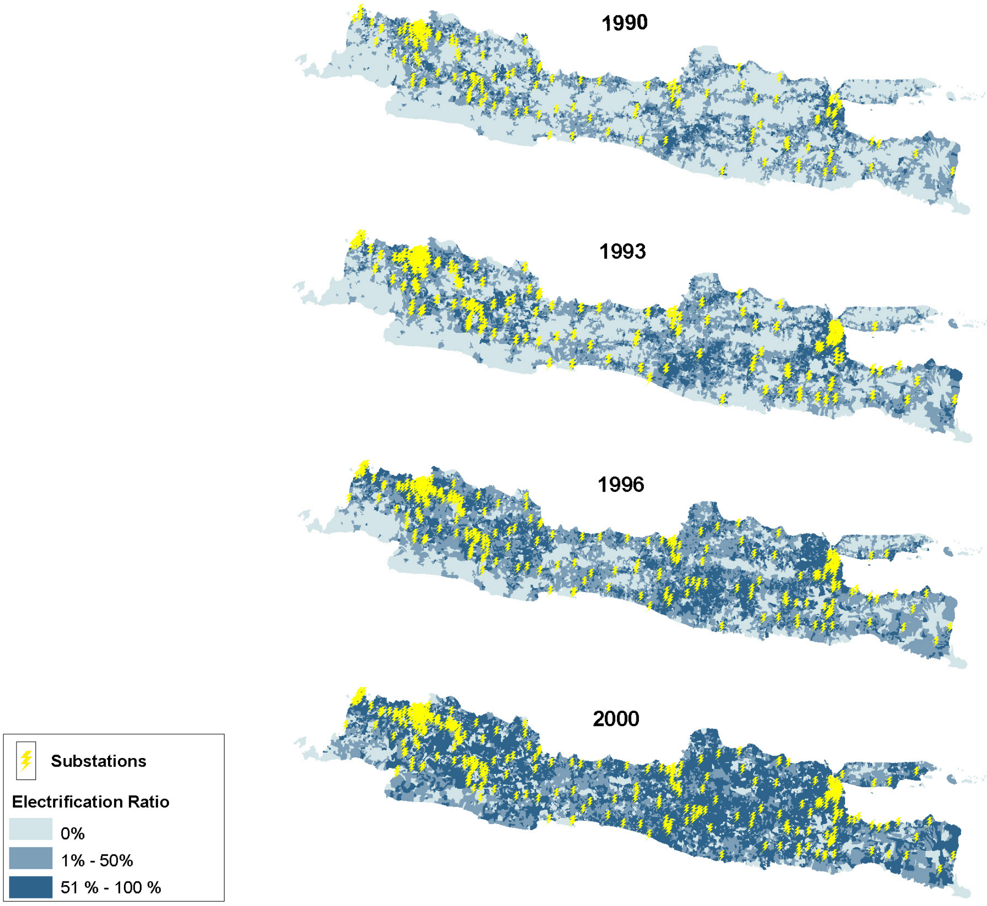
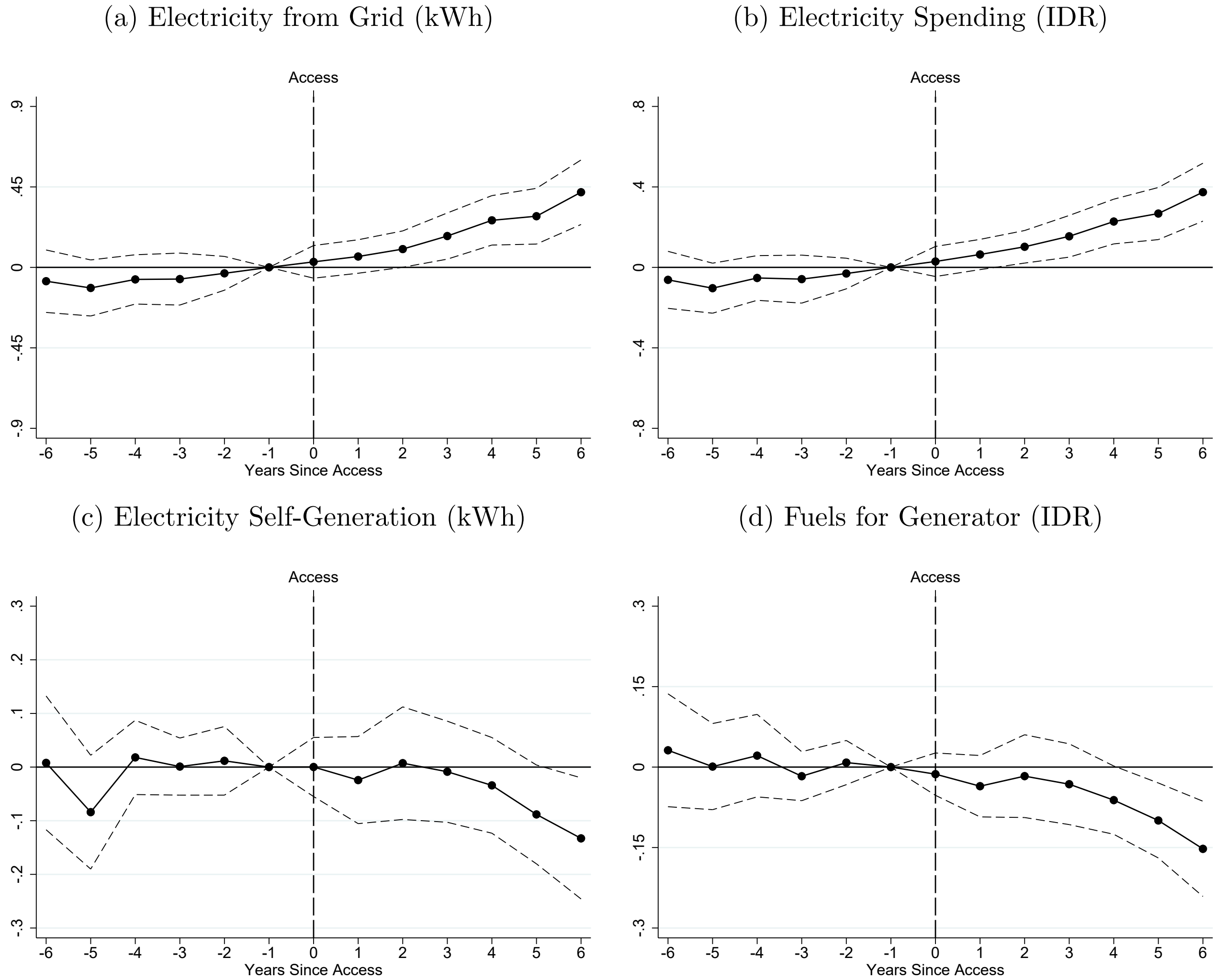


Figure 5: Event Study: Electricity Consumption and Generation



See section 3 for discussion. This figure presents event study style graphs of $log(1 + Y_{vpt})$ on the Y-axis, where Y is the total outcome of interest at the desa level in year t, and normalized time of access on the X-axis. The sample includes switchers (desas connected between 1991 and 1999) and unconnected desas by the end of the sample. The outcome variables are: (a) total grid electricity quantity in kWh consumed by all firms in the desa, (b) total grid electricity spending by all firms in the desa in IDR, (c) total electricity quantity self-generated (kWh), and (d) total spending on fuels for the generator by all firms in the desa. Each dot represents an estimated coefficient from equation (1). The dashed lines represent the 95% confidence intervals. Standard errors are clustered two-way at the province-year level and desa level.

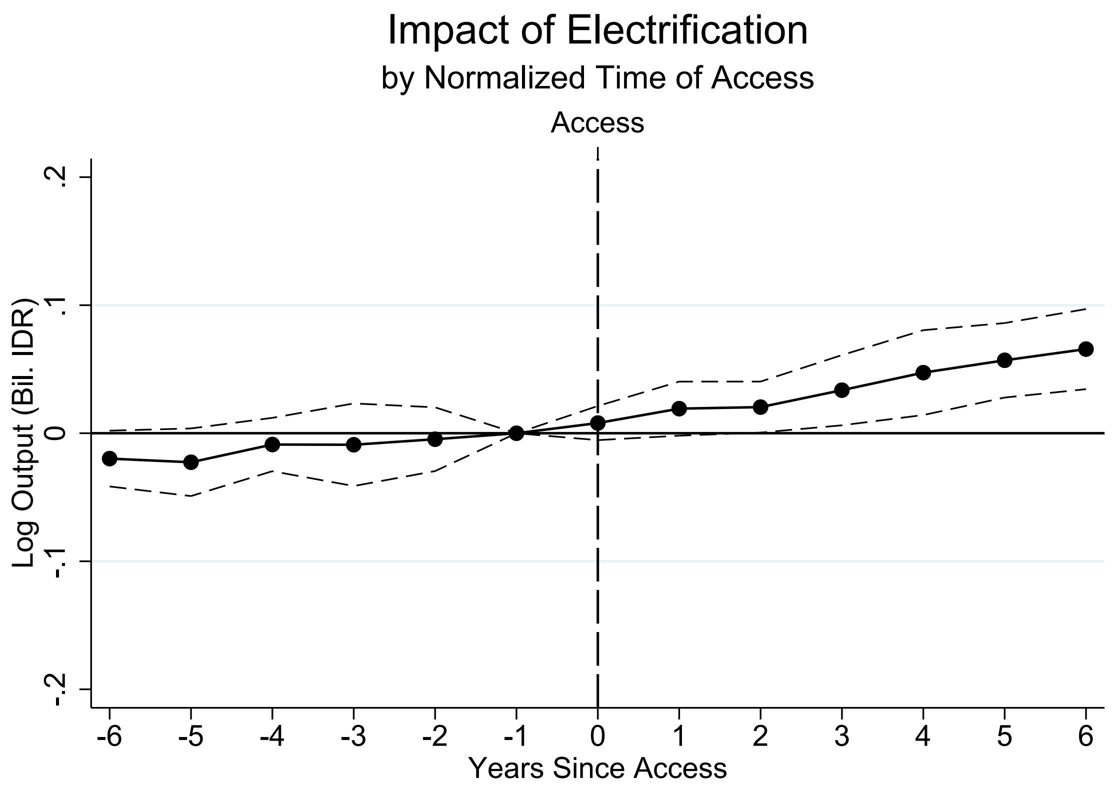
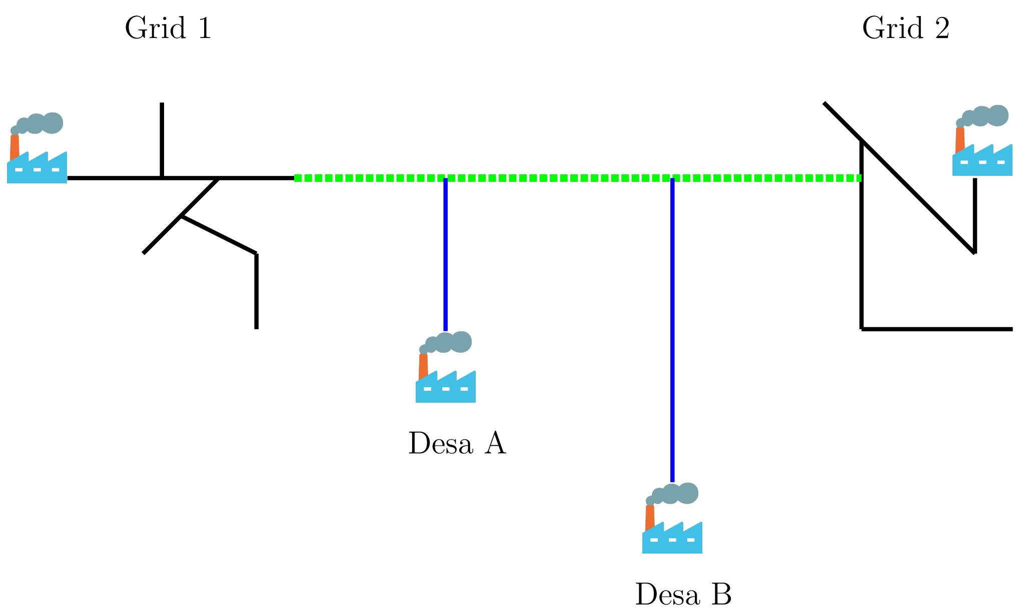


Figure 6: Event study graph based on equation (1). The sample includes switchers (desas connected between 1991 and 1999) and unconnected desas by the end of the sample. The variable on the Y-axis is the logarithm of 1 plus industrial output in billions of Indonesian Rupiahs. Controls include province-by-year fixed effects, desa controls, and desa fixed effects. Standard errors are clustered two-way at the province-by-year level and desa level. See Section 3.3 for a detailed discussion.

Figure 7: Empirical Strategy



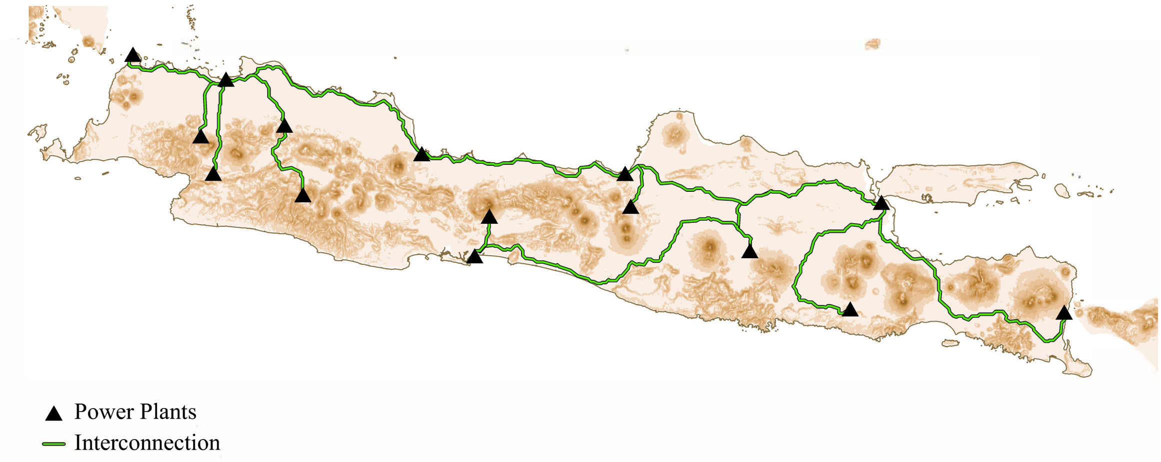


Figure 8: Hypothetical Transmission Interconnection. See Section 3.3 for discussion. The network of green lines represents the least cost transmission interconnection. Euclidean distance from the centroid of desas to the interconnection is used as an instrument for access to electrification.

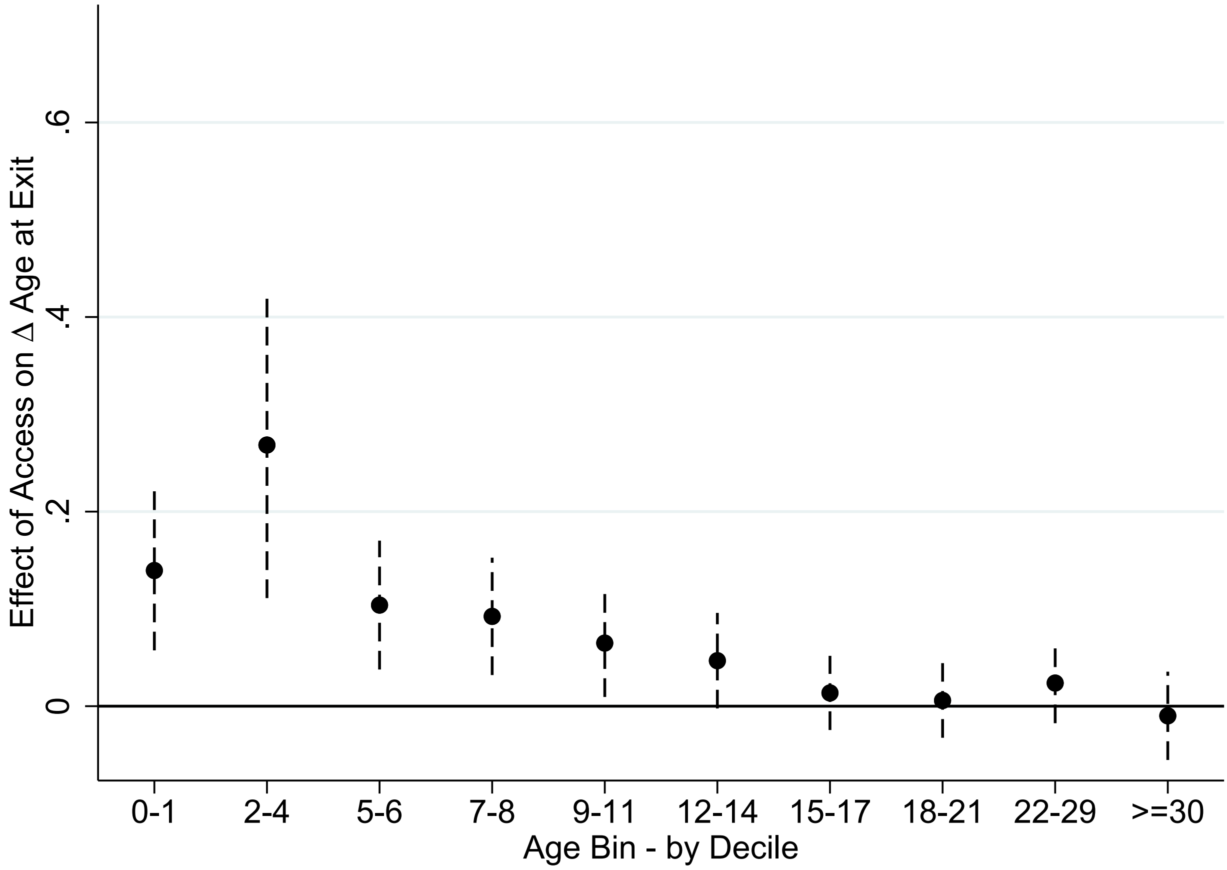
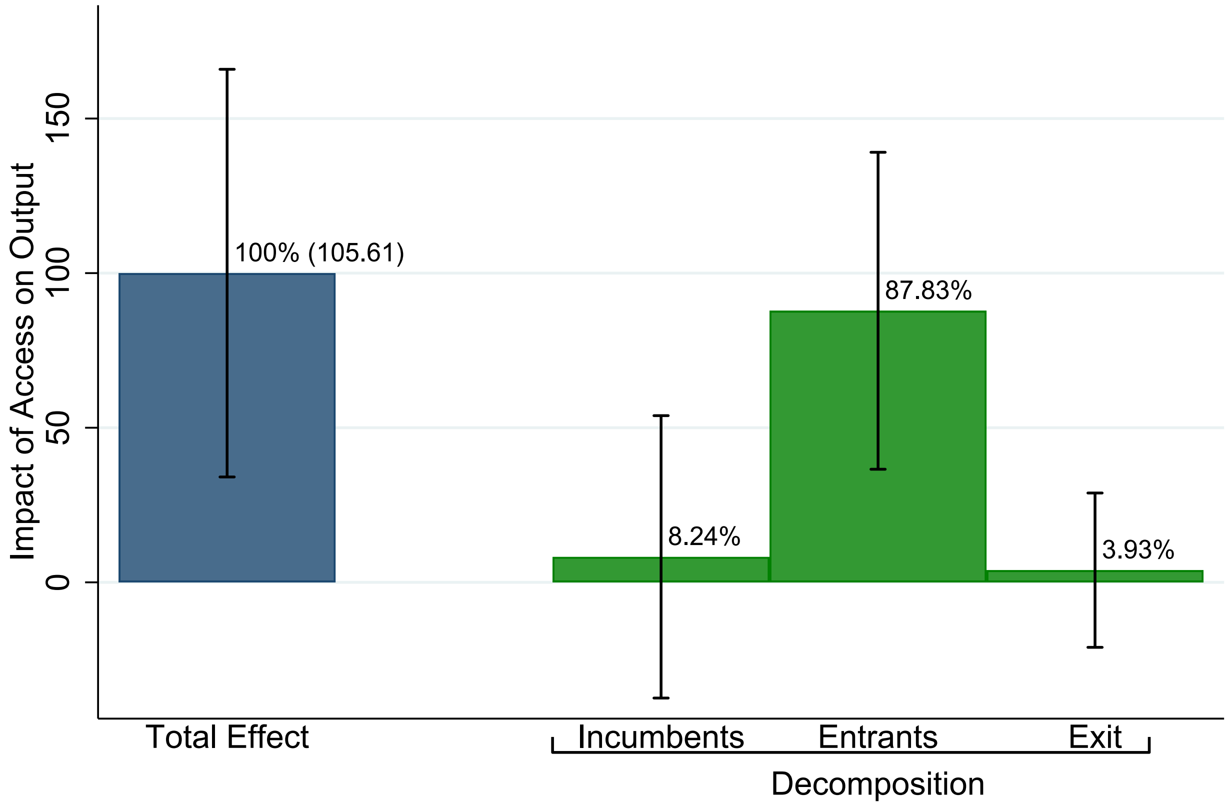


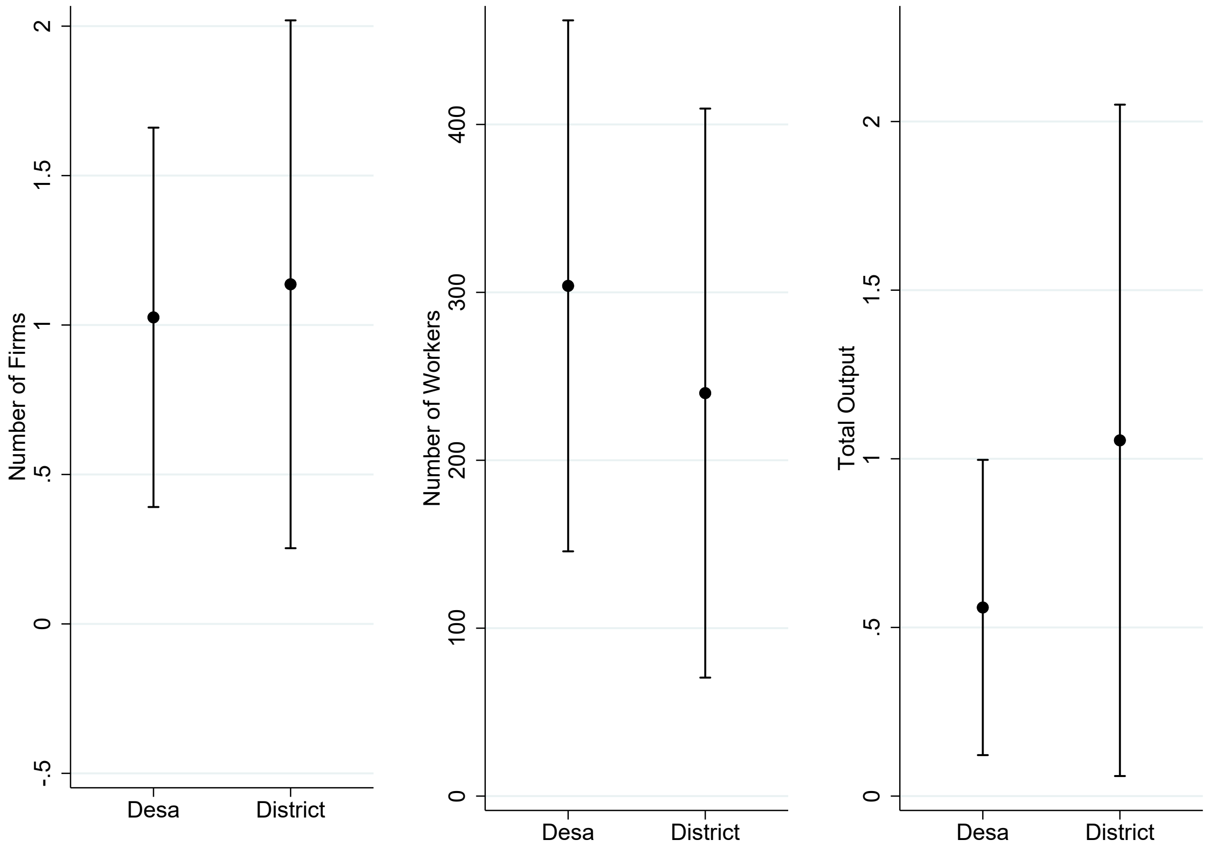
Figure 9: Estimates of the effect of Electrification on Exit by age decile. See Section 4.2 for discussion. Each dot corresponds to an estimate from a separate 2SLS regression with the number of exiting firms in age bin a on access, as in Equation 2. The dashed-lines represent the 95% confidence interval. Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.

Figure 10: Output Decomposition



See section 4.3 for discussion. The graph presents the decomposition of the effect of access on total output into the contribution of incumbents, entrants, and exiting firms. The blue bar corresponds to the IV estimate (specification (2) of the impact of access on the *level* of output, where the point estimate (in billion IDR) is in parenthesis on top of the blue bar. Similarly, the green bars correspond to the IV estimates of the impact of access on each of the components of output. The bars represent the 95% confidence interval. The estimates are normalized by the effect on total output. Standard errors are clustered two-way at the province-year level and desa level.

Figure 11: Comparing District-Level Estimates to Desa-Level Estimates



See section 4.3 for discussion. Each graph corresponds to a different outcome variable. Each dot and 95% confidence interval is from a separate regression using equation (2) at the desa or district level. District outcomes are the district mean value of Y across desas. The desa outcomes correspond to Table 5, and the district outcomes are from the equivalent specifications at the district level. Standard errors are clustered two-way at the province-year level and desa or district level.

B Tables

Table 1: Summary Statistics

(a) Electrification Infrastructure

Variable	1990	2000
Number of Substations	153	320
Total Capacity (MVA)	8426	25999

(b) Desa-Level Summary Statistics

Variable	Mean	Median	Min	Max
Access	0.62	1	0	1
Distance to Substation (km)	14.24	12.26	0.09	105.57
Number of firms	0.99	0	0	191
Number of firms > 0	4.45	2	1	191
Population	4,522.5	3,385	36	803,732
Number of desas in t		22,978		

(c) Desa-Level Controls

Variable	Mean	Std. Dev.	Min.	Max.
No. Firms 1990 (baseline)	0.78	4.02	0	182
Infrastructure				
Distance to Road (km)	3.54	3.36	0	31.25
Motorstation	0.06	0.23	0	1
Railway	0.02	0.15	0	1
Airport	0	0.05	0	1
Road fits 4 wheeled vehicle	0.96	0.2	0	1
Distance to Origin (km) Power Plant	47.26	21.61	10.01	124.57
Legal Status				
Government Rule	0.77	0.42	0	1
State Ministry	0.17	0.38	0	1
Governor	0.02	0.14	0	1
Other	0.02	0.14	0	1
N		228,310		

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. 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 2: Industry-Level Summary Statistics

	Observations		Access	
	(1)	(2)	(3)	(4)
Industry	1990	2000	1990	2000
Food and beverages	1814	2467	0.66	0.87
Textiles	1113	1567	0.71	0.91
Furniture	967	1386	0.74	0.76
Non-metallic products	817	1325	0.77	0.90
Recycling	467	1323	0.75	0.93
Wearing Apparel, fur	703	1124	0.76	0.92
Tobacco products	813	692	0.19	0.82
Rubber and plastic	617	850	0.86	0.96
Chemicals	427	553	0.91	0.97
Machinery and equipment	144	238	0.80	1.00
Fabricated metals	259	526	0.88	0.98
Wood products	303	640	0.78	0.88
Printing and publishing	190	260	0.82	0.99
Leather and footwear	251	426	0.82	0.98
Paper products	179	287	0.77	1.00
Electrical machinery	115	168	0.98	1.00
Motor Vehicles	111	143	0.94	1.00
Basic metals	48	110	0.96	1.00
Other Transport	98	121	0.88	0.98
Medical equipment	33	29	0.88	1.00
Radio, TV equipment	51	95	0.96	1.00
Coke, petroleum, fuel	2	20	1.00	0.90
Total	9,910	15,433		

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 2000. Access is an indicator variable equal to 1 if a firm is located in a desa with access to electricity. Columns (3) and (4) show the access ratio in each industry between 1990 and 2000.

Table 3: Summary Statistics-Firms

(a) Access

Variable	Mean	Median	Min	Max
Access	0.85	1	0	1
Distance to Substation (km)	8.45	4.5	0.09	66.28

(b) Average Firm Size and Energy Use

Variable	$Access_{vpt}$	
	0	1
Firm Size		
Output	1,850,000	10,627,420
Number of Workers	105.5	179.5
Energy Use		
Grid Electricity (kWh)	151,952	1,098,406
Grid Electricity (000 IDR)	19,009	179,396
Generator	0.21	0.32
Electricity Generated (kWh)	151,284	450,913
Geneartor Capacity (kVA/KW)	1,213.14	443.56
Generation Share (%)	21	12
Total Electricity (kWh)/Output	0.12	0.18
Electricity Genearted(kWh)/Output	0.07	0.04
Grid Electricity(kWh)/Output	0.05	0.14

Notes: See Section 3.2 for discussion. Summary statistics presenting the mean of firm-level size and energy use outcomes by access. $Access_{vpt}$ is an indicator variable equal to 1 if the desa where the firm is located is within 15 km of the nearest transmission substation. Output is in 2010 (1 USD is approximately $9,000$ IDR in 2010) Indonesian Rupiahs (billion). Generator is a dummy variable equal to 1 if a firm owns a generator. Generation share (%) is the share of generated electricity out of total electricity. Total electricity is the sum of grid electricity, generated electricity, and electricity purchased from private sellers. The variables in the last three rows are the ratio of total electricity, electricity generated, and grid electricity to output respectively.

Table 4: First Stage Regressions

Dependent Variable	$\Delta Access_{vpt}$		Δ^t_{1991}Distance to Nearest Substation (km)
	(1)	(2)	(3)
Instrument Z_v (km)	-0.000974***	-0.00136***	0.0223**
	(0.000200)	(0.000296)	(0.0103)
Distance to Origin Power plants (km)		0.000944***	-0.0466***
		(0.000256)	(0.00622)
Distance to Regional Road (km)		-0.00285**	0.0524*
		(0.00129)	(0.0268)
Dummy = 1 if road fits 4 wheeled vehicle		0.0256*	-0.257
		(0.0129)	(0.478)
Dummy=1 if village has a motorstation		-0.00436	-0.173
		(0.00622)	(0.183)
Dummy=1 if village has a railway		-0.00219	0.244
		(0.0149)	(0.398)
Dummy=1 if village has an airport		-0.00374	-0.794
		(0.0368)	(1.217)
N_{1990}		0.00113	-0.0792**
		(0.000878)	(0.0325)
Constant	0.161***	0.110***	-3.699***
	(0.00438)	(0.0151)	(0.395)
First Stage F	23.61	21.04	4.63
Observations	206,802	206,802	206,802
Year x Province FE	✓	✓	✓
Legal and Political Characteristics		✓	✓

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

Notes: See Section 3.3.2 for discussion. First stage regressions of access instrumented with distance to hypothetical least cost interconnection. In columns (1) and (2), a desa has $\Delta Access_{vpt} = 1$ if the desa receives access in year t and was not connected in 1991. N_{1990} is the baseline number of firms in 1990. The outcome variable in column (3) is the change in the distance to the nearest substation between year t and 1991 in km. Legal and political characteristics include whether a desa is governed by an elected or appointed official, development classification, and legal classification (such as transmigration settlements). Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.

Table 5: Estimates of the Effect of Electrification on Local Manufacturing Outcomes.

	Sample: Desa-Level		
ΔY	No. of Firms	No. of Workers in Manufacturing	Log(1+Output)
	(1)	(2)	(3)
	Panel A: IV		
$\Delta Access_{vpt}$	1.026***	303.9***	0.559**
	(0.316)	(78.84)	(0.218)
First Stage F	21.04	21.04	21.04
	Panel B: OLS		
$\Delta Access_{vpt}$	-0.0360**	-11.94***	0.00352
	(0.0150)	(3.372)	(0.0108)
	Panel C: Reduced Form		
Z_v	-0.00139***	-0.412***	-0.000759**
	(0.000428)	(0.109)	(0.000310)
Year x Province FE	✓	✓	✓
Controls	✓	✓	✓
Observations	206,802	206,802	206,802
$\bar{Y}$	1	104.17	.41
$\bar{Y}\|N > 0$	4.46	462.56	1.82

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

Notes: See Section 4.1 for discussion. Results from IV, OLS, and reduced form regressions of Equation (2). Controls include distance to nearest origin power plant, distance to nearest regional road, baseline number of firms, infrastructure controls and legal and political characteristics. Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.

Table 6: Estimates of the Effect of Electrification on Entry, Exit, and Turnover.

	Sample: Desa-Level			
ΔY	No. of Entrants	No. of Exiting Firms	Entry Rate	Exit Rate
	(1)	(2)	(3)	(4)
	Panel A: IV			
$\Delta Access_{vpt}$	1.997***	0.972***	0.356**	0.393**
	(0.508)	(0.298)	(0.150)	(0.156)
First Stage F	21.04	21.04	20.04	20.04
	Panel B: OLS			
$\Delta Access_{vpt}$	-0.0297	0.00638	-0.00658	0.00647
	(0.0265)	(0.0238)	(0.00952)	(0.0137)
	Panel C: Reduced Form			
Z_v	-0.00271***	-0.00132***	-0.000765***	-0.000845***
	(0.000570)	(0.000302)	(0.000244)	(0.000282)
Observations	206,802	206,802	39,015	39,015
Year x Province FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
$\bar{Y}$	.43	.21	.34	.21
$\bar{Y}\|N > 0$	1.89	.79	.34	.21

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

Notes: See Section 4.2 for discussion. Results from IV, OLS, and reduced form regressions of Equation (2). Controls include distance to nearest origin power plant, distance to nearest regional road, baseline number of firms, infrastructure controls and legal and political characteristics. Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.

Table 7: Impact of connection on input use at the firm level.

	Sample: Firm-Level							
$log(Y)$	Wage Bill	No. Workers	Capital	Materials	Energy Bill	Generator Fuels Spending	Electricity (kWh)	Generation Share (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: IV							
$Access_{vpt}$	0.334	0.358*	1.509***	0.826**	2.572***	-0.177	2.852***	-0.115***
	(0.342)	(0.198)	(0.379)	(0.377)	(0.633)	(0.266)	(0.987)	(0.0352)
First Stage F	39.40	39.44	39.26	40.94	41	43.23	39.44	41.55
	Panel B: OLS							
$Access_{vpt}$	0.272***	0.169***	0.316***	0.448***	0.502***	-0.0907*	1.045***	-0.0343***
	(0.0529)	(0.0352)	(0.0671)	(0.0872)	(0.109)	(0.0498)	(0.158)	(0.00635)
	Panel C: Reduced Form							
Z_v	-0.00202	-0.00217*	-0.00922***	-0.00507**	-0.0159***	0.00115	-0.0173***	0.000737***
	(0.00200)	(0.00117)	(0.00182)	(0.00215)	(0.00321)	(0.00171)	(0.00586)	(0.000218)
Observations	38,444	38,447	37,294	36,984	37,312	35,375	38,447	29,308
Industry x Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Province x Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 5 for discussion. Results from IV, OLS, and reduced form regressions of Equation (6). Dependent variables are in log, except for the generation share in Column (8). Desa controls include distance to nearest origin power plant, distance to nearest regional road, baseline number of firms, infrastructure controls and legal and political characteristics. Firm Controls include export, generator, and ownership dummies (central government, local government, domestic, foreign). Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.

Table 8: Impact of connection: Intensive v.s. Extensive Margin.

Sample: Firm-Level								
	Output				TFPR (ϕ_{it})			
$log(Y)$	All	Entry	Exit	Incumbents	All	Entry	Exit	Incumbents
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: IV								
$Access_{vpt}$	0.973***	1.747***	0.648**	0.757	0.297***	0.524***	0.103	0.316**
	(0.360)	(0.541)	(0.292)	(0.472)	(0.106)	(0.162)	(0.0755)	(0.135)
First Stage F	39.40	28.88	40.17	31.52	41.48	27.23	42.89	34.13
Panel B: OLS								
$Access_{vpt}$	0.411***	0.384***	0.169**	0.454***	0.0150	-0.0251	-0.0218	0.0326*
	(0.0742)	(0.103)	(0.0636)	(0.0874)	(0.0157)	(0.0270)	(0.0198)	(0.0178)
Panel C: Reduced Form								
Z_v	-0.00589***	-0.0130***	-0.00483**	-0.00407*	-0.00184***	-0.00380***	-0.000787	-0.00176**
	(0.00194)	(0.00281)	(0.00199)	(0.00238)	(0.000604)	(0.000819)	(0.000589)	(0.000679)
Observations	38,440	5,537	10,093	24,363	35,321	4,835	9,006	22,696
Industry x Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Province x Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 5 for discussion. Results from IV, OLS, and reduced form regressions of Equation (6). Entry is defined as the set of firms aged less than three years old. Exit is defined as the set of firms in the last three years they appear in the census. Desa controls include distance to nearest origin power plant, distance to nearest regional road, baseline number of firms, infrastructure controls and legal and political characteristics. Firm Controls include export, generator, and ownership dummies (central government, local government, domestic, foreign). Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.

For Online Publication

C Theoretical Appendix

C.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, 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.

C.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:

$$U = H + \int_{i\in I} \alpha q_i di - \frac{1}{2}\eta\Big(\int_{i\in I} q_i di\Big)^2 - \frac{1}{2}\gamma \int_{i\in I} {q_i}^2 di \tag{7}$$

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:

$$q_i = \frac{\alpha}{\eta N + \gamma} + \frac{\eta N}{\gamma(\eta N + \gamma)}\bar{p} - \frac{1}{\gamma}p_i \tag{8}$$

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 (8) to zero:

$$p^{max} = \frac{\gamma\alpha}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} \tag{9}$$

The residual demand for product i from (8) can therefore be written as:

$$q_i = \frac{1}{\gamma}(p^{max} - p_i) \tag{10}$$

C.3 Production

On the production side, consider a single input technology[30] where firm i produces according the the following production function:

$$q_i = \phi_i x_i \tag{11}$$

where ϕ_i is the firm's physical productivity and x_i is the input of production which is supplied inelastically at a constant[31] 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:

$$p(c_i) = \frac{1}{2}(p^{max} + c_i) \tag{12}$$

The equilibrium profit is:

$$\pi(c_i) = \frac{1}{4\gamma}(p^{max} - c_i)^2 \tag{13}$$

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:

$$c^* = p^{max} = \frac{\gamma\alpha}{\eta N + \gamma} + \frac{\eta N}{\eta N + \gamma}\bar{p} \tag{14}$$

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

$$p(c_i) = \frac{1}{2}(c^* + c_i) \tag{15}$$

[30]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.

[31]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.

$$\mu(c_i) = \frac{1}{2}(c^* - c_i) \tag{16}$$

$$q(c_i) = \frac{1}{2\gamma}(c^* - c_i) \tag{17}$$

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 a cutoff level for firm productivity:

$$\phi^* = \frac{w}{c^*} \tag{18}$$

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

C.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:

$$V^e = \frac{w^2}{4\gamma} \int_{\phi^*}^{\infty} \left(\frac{1}{\phi^*} - \frac{1}{\phi} \right)^2 dG(\phi) - s = 0 \tag{19}$$

Equation (19) pins down ϕ^* which summarizes the equilibrium. The equilibrium mass of firms N is determined using equations (12) and (14).

C.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:

$$\frac{d\phi^*}{dw} = -\frac{\partial V^e/\partial w}{\partial V^e/\partial \phi^*} > 0 \tag{20}$$

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:

$$\frac{d\phi^*}{ds} = -\frac{\partial V^e/\partial s}{\partial V^e/\partial \phi^*} < 0 \tag{21}$$

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.

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 (15)-(17) conditional on survival are:

$$\bar{p} = \frac{1}{2}(c^* + \bar{c}) \tag{22}$$

$$\bar{\mu} = \frac{1}{2}(c^* - \bar{c}) \tag{23}$$

$$\bar{q} = \frac{1}{2\gamma}(c^* - \bar{c}) \tag{24}$$

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 (22) predicts that the average observed prices conditional on firm survival is lower when c^* is lower. Equations (23) and (24) 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 (16) and (17), 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 $c^* = p^{max}$:

$$c^* = \bar{p} + \frac{\gamma(\alpha - \bar{p})}{\eta N + \gamma} \tag{25}$$

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

$$N = \frac{2\gamma(\alpha - c^*)}{\eta(c^* - \bar{c})} \tag{26}$$

These equations state that tougher competition (lower c^*) is associated with a higher mass of active firms N and a lower average price[32] $\bar{p}$. To see this[33], suppose N increases, and that surviving firms don't change their prices following entry, keeping $\bar{p}$ constant. From equation (25), c^* will decrease. From equation (22), $\bar{p}$ 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 (22)-(24) 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 the 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 (20) and (21) 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 C.1 summarizes the predictions of the model, split by the different channels :

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

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

Table C.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 locations 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 the 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 depends 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.

D 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):

$$y_{it} = p_{it} + q_{it} = f(\mathbf{x}_{it}, \beta) + a_{it} + p_{it} + \epsilon_{it} \tag{27}$$

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

$$\phi_{it} = a_{it} + p_{it} \tag{28}$$

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:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \beta_e e_{it} + \phi_{it} + \epsilon_{it} \tag{29}$$

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

$$\phi_{it} = h^{-1}(e_{it}, k_{it}, l_{it}, m_{it}) \tag{30}$$

Substituting back in Equation (29):

$$y_{it} = \kappa(k_{it}, l_{it}, m_{it}, e_{it}) + \epsilon_{it} \tag{31}$$

Estimating Equation (31) 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:

$$\phi_{it} = g(\phi_{it-1}) + \eta_{it} \tag{32}$$

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 (32):

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

and

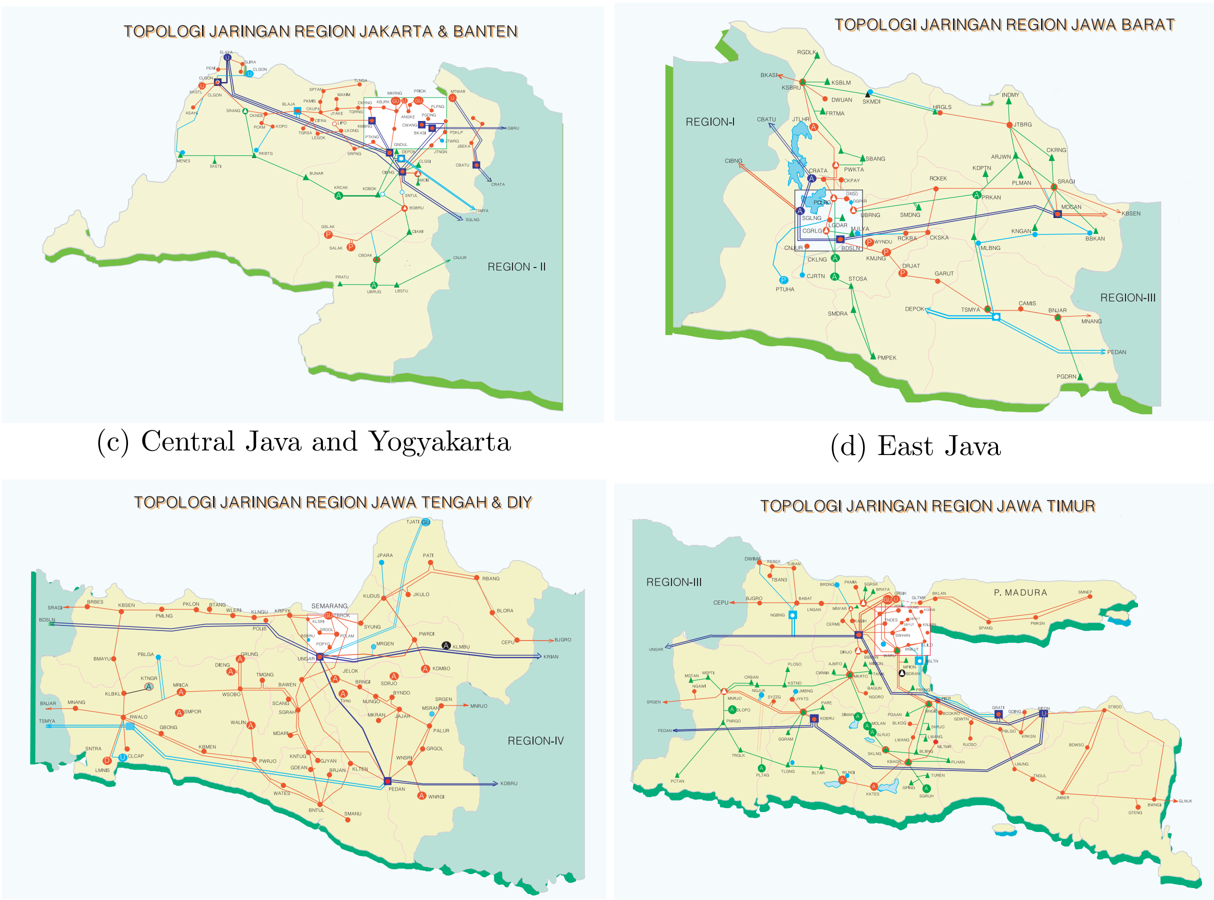
$$\phi_{it}(\beta_l, \beta_m, \beta_e, \beta_k) = \hat{\kappa}_{it} - \beta_l l_{it} - \beta_m m_{it} - \beta_e e_{it} - \beta_k k_{it}$$

I implement the one step GMM estimator as suggested by Wooldridge (2009) in *Stata* using the Rovigatti and Mollisi (2016) package.

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

E Additional Figures and Tables

Figure E.1: Example of Current Maps of the Transmission Network.



(a) Java and Banten

(b) West Java

(c) Central Java and Yogyakarta

(d) East Java

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

E.1 Compliers Characteristics

Identifying compliers is important for the interpretation of the LATE estimate. Table E1 explores the characteristics of complier desas. Each column corresponds to a separate first stage regression of access on the Z_v, interacted with two mutually exclusive indicator variables such as being close to a regional road (below median distance) and not being close to a regional road (above median distance). Each coefficient represents the effect of Z_v on access for that sub-population of desas. At the bottom of each column, I present the P-value of the F-test testing the hypothesis that the two coefficients are equal. The results indicate that complier desas are more likely to be further away from a regional road (Column (1)), rural (Column (6)), have a larger population (Column (7)), and being located in less elevated areas (Column (9)), but not more likely to be at the coast (Column (10)). Complier desas do not necessarily have access to a better local road (Column

(2)), have a lower share of households working in agriculture (Column (8)), are more industrialized (Columns (3) and (4)) nor have access to banking in 1980 (Column (5)). This suggests that complier desas have a lower level of development (hence the downward bias in the OLS), suggesting that the potential for growth in response to electrification for compliers is higher.

E.2 Robustness

Table E2 presents results from various robustness checks. Panel 1 reports the main estimates from Tables 5 and 6 for reference. Panels 2 to 6 are based on specification (2). Panel 2 presents the results with an alternative measure of access defined as an indicator for being within 20 km of the nearest substation instead of 15 km (as the main access treatment is defined). Panel 3 excludes desas within 20 km of the nearest origin power plant. Panel 4 presents the estimates with zero-preserving log transformations of the outcome variables in levels in the main analysis (number of firms, manufacturing workers, number of entrants, and number of exiting firms), and the estimates for output in levels (which is in logs in the main text). Panel 5 presents the results using Conley (1999) standard errors, accounting for spatial correlation (within 500 km) and for serial correlation across time. Panel 6 controls for an additional local geographic characteristic, land gradient. Finally, Panel 7 presents results from a cross-sectional specification analogous to equation (6) at the desa-level, using Z_v as an instrument for access:

$$Y_{vpt} = \alpha + \beta Access_{vpt} + \boldsymbol{V'}_{vpt}\boldsymbol{\eta} + \gamma_{pt} + \epsilon_{vpt} \tag{33}$$

Table E3 presents the firm level estimates from Equation (6) using the full sample of firms, including those already connected at baseline.

Table E1: Compliers Characteristics

Sample: Desa-Level					
	(1)	(2)	(3)	(4)	(5)
Dependent Variable			$\Delta Access_{vt}$		
$Z_v * \mathbb{1}(ClosetoRoad = 0)$	-0.0015***				
	(0.0003)				
$Z_v * \mathbb{1}(ClosetoRoad = 1)$	-0.0011***				
	(0.0003)				
$Z_v * \mathbb{1}(FourwheelsRoad = 0)$		-0.0007			
		(0.0005)			
$Z_v * \mathbb{1}(FourwheelsRoad = 1)$		-0.0014***			
		(0.0003)			
$Z_v * \mathbb{1}(LowBaselineN = 0)$			-0.0014***		
			(0.0004)		
$Z_v * \mathbb{1}(LowBaselineN = 1)$			-0.0013***		
			(0.0003)		
$Z_v * \mathbb{1}(AnyFactory1980 = 0)$				-0.0015***	
				(0.0003)	
$Z_v * \mathbb{1}(AnyFactory1980 = 1)$				-0.0010***	
				(0.0003)	
$Z_v * \mathbb{1}(Atleast1Bank1980 = 0)$					-0.0009***
					(0.0003)
$Z_v * \mathbb{1}(Atleast1Bank1980 = 1)$					-0.0016***
					(0.0005)
P-value of F-test	0.0400	0.158	0.679	0.0880	0.187
Observations	206,802	206,802	206,802	165,258	165,258
	(6)	(7)	(8)	(9)	(10)
$Z_v * \mathbb{1}(Urban = 0)$	-0.0014***				
	(0.0003)				
$Z_v * \mathbb{1}(Urban = 1)$	-0.0004				
	(0.0005)				
$Z_v * \mathbb{1}(LowPopuation = 0)$		-0.0015***			
		(0.0003)			
$Z_v * \mathbb{1}(LowPopuation = 1)$		-0.0011***			
		(0.0003)			
$Z_v * \mathbb{1}(LowAgr.Share1980 = 0)$			-0.0007**		
			(0.0003)		
$Z_v * \mathbb{1}(LowAgr.Share1980 = 1)$			-0.0012***		
			(0.0004)		
$Z_v * \mathbb{1}(LowElevation = 0)$				-0.0027***	
				(0.0005)	
$Z_v * \mathbb{1}(LowElevation = 1)$				-0.0003	
				(0.0004)	
$Z_v * \mathbb{1}(Coast = 0)$					-0.0013***
					(0.0003)
$Z_v * \mathbb{1}(Coast = 1)$					-0.0013***
					(0.0005)
P-value of F-test	0.0319	0.0426	0.273	0.000125	0.980
Observations	228,310	228,310	186,060	228,310	228,310
Province x Year FE	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓

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

Notes: See Section E.1 for discussion. Results from first stage regressions of desa-level $\Delta Access_{vpt}$ on the instrument interacted with various desa-level characteristics. For continuous variables (distance to road, population in year t, no. of households in agriculture in 1980 divided by population in 1980, elevation, N_{1990}), the indicator variables are defined based on the median value. Desa controls include proximity to coast, proximity to origin power-plant, elevation, distance to road, baseline number of firms in 1990, desa political and legal status, and infrastructure controls. Robust standard errors in parentheses clustered at the province-by-year and desa level. The report p-values correspond to the hypothesis test that the coefficients on the interactions are equal.

Table E2: Estimates of the Effect of Electrification on Local Manufacturing Outcomes.

ΔY	No. of Firms	No. of Workers in Manufacturing	Log(1+Output)	No. of Entrants	No. of Exiting Firms	Entry Rate	Exit Rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1- Main Estimates							
$\Delta Access_{vpt}$	1.026***	303.9***	0.559**	1.997***	0.972***	0.356**	0.393**
	(0.316)	(78.84)	(0.218)	(0.508)	(0.298)	(0.150)	(0.156)
2- Access within 20 km of nearest station							
$Access^{20}_{vpt}$	1.311***	388.4***	0.715**	2.553***	1.242**	0.292**	0.322***
	(0.452)	(118.7)	(0.274)	(0.838)	(0.492)	(0.114)	(0.119)
3- Excluding 20 km from origin plants							
$\Delta Access_{vpt}$	0.725**	264.5***	0.396*	1.367***	0.642**	0.296*	0.343**
	(0.291)	(77.59)	(0.201)	(0.430)	(0.242)	(0.149)	(0.162)
4- $\Delta log(1+Y)$, except output in levels							
$\Delta Access_{vpt}$	0.258***	0.578	105.6***	0.631***	0.414***	-	-
	(0.0752)	(0.351)	(34.71)	(0.147)	(0.128)	-	-
5- Conley Standard Errors							
$\Delta Access_{vpt}$	1.026***	303.9***	0.559**	1.997***	0.972***	0.356***	0.393***
	(0.39)	(81.53)	(0.24)	(0.64)	(0.28)	(0.12)	(0.12)
6- Controlling for land gradient							
$\Delta Access_{vpt}$	0.749*	360.6***	0.295	1.572**	0.823**	0.226	0.321**
	(0.408)	(121.0)	(0.286)	(0.640)	(0.398)	(0.153)	(0.158)
7- Cross-Sectional IV							
$Access_{vpt}$	0.343***	126.5***	0.605***	0.554***	0.211***	0.287**	0.277***
	(0.0844)	(35.05)	(0.0923)	(0.105)	(0.0436)	(0.108)	(0.0737)

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

Notes: See main text and Section E.2 for discussion. Results from differeced IV regressions of Equation (2) in Panels 1-6, and a cross-sectional IV regression of Equation (33) in Panel 7. All specifications include province-by-year effects and desa controls (distance to nearest origin power plant, distance to nearest regional road, baseline number of firms, infrastructure controls and legal and political characteristics). Robust standard errors in parentheses clustered two-way at the province-by-year and desa level, except in Panel 5 (Conley (1999)).

Table E3: Impact of connection on the sales and inputs at the firm level.

Sample: Firm-Level										
Dependent Variable	Wage Bill (1)	Workers No. (2)	Capital (3)	Materials (4)	Energy Bill (5)	Generator Fuels Bill (6)	Electricity (kWh) (7)	Generation Share (%) (8)	Output (9)	TFPR (10)
Panel A: IV										
$Access_{vpt}$	0.961	0.770**	2.441***	1.451**	3.436***	0.0798	3.973**	-0.258***	1.608**	0.207
	(0.576)	(0.354)	(0.734)	(0.596)	(0.933)	(0.402)	(1.589)	(0.0670)	(0.674)	(0.158)
First Stage F	17.10	17.12	16.06	21.09	16.76	18.76	17.12	21.83	17.10	19.20
Panel B: OLS										
$Access_{vpt}$	0.311***	0.207***	0.333***	0.448***	0.467***	-0.181***	1.244***	-0.0671***	0.398***	0.0109
	(0.0423)	(0.0284)	(0.0538)	(0.0665)	(0.103)	(0.0478)	(0.181)	(0.0103)	(0.0596)	(0.0141)
Panel C: Reduced Form										
Z_v	-0.00288**	-0.0023**	-0.0072***	-0.0046***	-0.0103***	-0.00026	-0.012***	0.001***	-0.0048***	-0.0007
	(0.0014)	(0.001)	(0.00140)	(0.0016)	(0.0021)	(0.0013)	(0.0042)	(0.0002)	(0.0015)	(0.0004)
Observations	137,599	137,603	130,723	131,388	135,490	126,798	137,603	118,769	137,542	123,972
IndustryxYear FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ProvincexYear FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 5 for discussion. Results from IV, OLS, and reduced form regressions of Equation (6). Desa controls include distance to nearest origin power plant, distance to nearest regional road, baseline number of firms, infrastructure controls and legal and political characteristics. Firm Controls include export, generator, and ownership dummies (central government, local government, domestic, foreign). Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.

Table E.4: Impact of connection: Intensive v.s. Extensive Margin.

Sample: Firm-Level								
	Output				TFPR (ϕ_{it})			
$log(Y)$	All	Entry	Exit	Incumbents	All	Entry	Exit	Incumbents
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: IV								
$Access_{vpt}$	0.973***	1.187	0.413	0.972**	0.297***	0.354	-0.0339	0.308***
	(0.360)	(0.719)	(0.333)	(0.382)	(0.106)	(0.228)	(0.119)	(0.113)
First Stage F	39.40	15.10	23.64	39.48	41.48	9.840	27.04	41.353
Panel B: OLS								
$Access_{vpt}$	0.411***	0.178	0.0925	0.438***	0.0150	-0.0251	-0.0218	0.0326*
	(0.0742)	(0.168)	(0.0925)	(0.0758)	(0.0157)	(0.0270)	(0.0198)	(0.0178)
Panel C: Reduced Form								
Z_v	-0.00589***	-0.00883*	-0.00274	-0.00578***	-0.00184***	-0.00280**	0.000244	-0.00189***
	(0.00194)	(0.00281)	(0.00199)	(0.00238)	(0.000604)	(0.00128)	(0.000850)	(0.000625)
Observations	38,440	818	2,996	34,626	35,321	642	2,551	32,056
Industry x Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Province x Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Desa Controls	✓	✓	✓	✓	✓	✓	✓	✓
Firm Controls	✓	✓	✓	✓	✓	✓	✓	✓

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

Notes: See Section 5 for discussion. Results from IV, OLS, and reduced form regressions of Equation (6). Entry is defined as the set of firms aged 0. Exit is defined as the set of firms in the last year they appear in the census. Desa controls include distance to nearest origin power plant, distance to nearest regional road, baseline number of firms, infrastructure controls and legal and political characteristics. Firm Controls include export, generator, and ownership dummies (central government, local government, domestic, foreign). Robust standard errors in parentheses clustered two-way at the province-by-year and desa level.