



FINANCIANDO EL DESARROLLO • AMÉRICA LATINA

CAF
DOCUMENTOS DE TRABAJO

CAF
WORKING PAPERS

DEVELOPMENT EFFECTS OF ELECTRIFICATION: EVIDENCE
FROM THE GEOLOGIC PLACEMENT OF HYDROPOWER
PLANTS IN BRAZIL

N° 2011/03

March, 2011

Lipscomb, Molly
Mobarak Mushfiq, Ahmed
Bahram, Tania

CAF - Ave. Luis Roche, Torre CAF, Altamira. Caracas, Venezuela 01060

© CAF, 2011 por Lipscomb, Molly, Mobarak Mushfiq, Ahmed y Bahram, Tania. Todos los derechos reservados. Pequeñas secciones del texto, menores a dos párrafos, pueden ser citadas sin autorización explícita siempre que se cite el presente documento.

Los resultados, interpretaciones y conclusiones expresados en esta publicación son de exclusiva responsabilidad de su(s) autor(es), y de ninguna manera pueden ser atribuidos a CAF, a los miembros de su Directorio Ejecutivo o a los países que ellos representan.

CAF no garantiza la exactitud de los datos incluidos en esta publicación y no se hace responsable en ningún aspecto de las consecuencias que resulten de su utilización.

EFFECTOS DE LA ELECTRIFICACIÓN SOBRE EL DESARROLLO: EVIDENCIA DEL POSICIONAMIENTO GEOPOLÍTICO DE PLANTAS DE ENERGÍA HIDROELÉCTRICA EN BRASIL

Lipscomb, Molly, Mobarak Mushfiq, Ahmed y Bahram, Tania
CAF Documento de trabajo N° 2011/03
Marzo, 2011

RESUMEN

Estimamos el efecto del desarrollo de la electrificación en Brasil durante el período 1960-2000. Brasil depende casi exclusivamente en energía hidroeléctrica, la cual requiere interceptar agua a altas velocidades. Construimos un modelo de ingeniería que solo toma como insumos variables geográficas (gradiente del río, flujo de agua y el Amazonas) y simula una serie de tiempo de las redes de electricidad hipotéticas para Brasil que muestran cómo seta hubiera evolucionado si las inversiones en infraestructura se hubieran realizado sólo tomando en cuenta los costos geológicos e ignorando los temas asociados a la demanda. Usando este modelo como un instrumento, documentamos que los grandes efectos positivos que ha tenido la electrificación sobre el desarrollo han sido subestimados cuando no se toma en cuenta la distribución política y hacia áreas menos desarrolladas de los proyectos de infraestructura. La mejora generalizada de la productividad laboral entre sectores y áreas en vez del reordenamiento por efectos de equilibrio general aparece ser el mecanismo más probable a través del cual estas mejoras en desarrollo son realizadas.

Palabras clave: electricidad, energía hidroeléctrica, Brasil

RETURNS TO ELECTRICITY: EVIDENCE FROM THE QUASI-RANDOM PLACEMENT OF HYDROPOWER PLANTS IN BRAZIL

Lipscomb, Molly, Mobarak Mushfiq, Ahmed and Bahram, Tania
CAF Working paper N° 2011/03
March, 2011

ABSTRACT

We estimate the development effects of electrification across Brazil over the period 1960-2000. Brazil relies almost exclusively on hydropower, which requires intercepting water at high velocity. We build an engineering model which takes as inputs only geography (river gradient, water flow and Amazon) and simulates a time series of hypothetical electricity grids for Brazil that show how the grid would have evolved had infrastructure investments been made based solely on geologic cost considerations, ignoring all demand-side concerns. Using the model as an instrument, we document large positive effects of electrification on development that are underestimated when one fails to account for the political allocation of infrastructure projects or its targeting to under-developed areas. Broad-based improvement in labor productivity across sectors and areas rather than general equilibrium re-sorting (in-migration to electrified counties) appears to be the likely mechanism by which these development gains are realized.

Molly Lipscomb
University of Notre Dame

Ahmed Mobarak Mushfiq
Yale University, School of Management
ahmed.mobarak@yale.edu

Tania Bahram
University of Colorado

Keywords: electricity, hydro-power, Brazil

Development Effects of Electrification: Evidence from the Geologic Placement of Hydropower Plants in Brazil*

Molly Lipscomb
University of Notre Dame

Ahmed Mushfiq Mobarak
Yale School of Management

Tania Barham
University of Colorado, Boulder

March 27, 2011

Abstract

We estimate the development effects of electrification across Brazil over the period 1960-2000. Brazil relies almost exclusively on hydropower, which requires intercepting water at high velocity. We build an engineering model which takes as inputs only geography (river gradient, water flow and Amazon) and simulates a time series of hypothetical electricity grids for Brazil that show how the grid would have evolved had infrastructure investments been made based solely on geologic cost considerations, ignoring all demand-side concerns. Using the model as an instrument, we document large positive effects of electrification on development that are underestimated when one fails to account for the political allocation of infrastructure projects or its targeting to under-developed areas. Broad-based improvement in labor productivity across sectors and areas rather than general equilibrium re-sorting (in-migration to electrified counties) appears to be the likely mechanism by which these development gains are realized.

*We thank the University of Colorado NICHD Population Center, Corporación Andina de Fomento, Center for Advancement in Research and Teaching in the Social Sciences at the University of Colorado, the National Science Foundation, and the Macmillan Center at Yale University for the financial support that made this data collection possible. We also thank Daniel Ortega, Taryn Dinkelman, Steven Puller, Rohini Pande, Arik Levinson, Erin Mansur, Sheila Olmstead, Judy Chevalier, Bill Evans, and seminar participants at NBER Environmental and Energy Economics Summer Institute, Corporación Andina de Fomento, UC-Berkeley Energy Institute, IDEI-World Bank International Conference on Infrastructure Economics, University of Virginia, Harvard University, Cornell University, Inter-American Development Bank, Center for Global Development, Fudan University, Yale University (Economics) and Yale School of Management for comments, and Vanessa Empinotti and Steven Li for excellent research assistance.

1 Introduction

Construction of large-scale infrastructure projects was a popular use of development funds until the 1970s but this was replaced by a trend toward smaller programs in health and education in the 1980s and 1990s. There is now renewed support for large infrastructure projects as a means to poverty reduction (World Bank, 2003; Ali and Pernia, 2003). Despite the renewed investment, there is relatively little causal evidence of the effects of large infrastructure investment in general,¹ ² and electrification in particular.³ This is because electricity networks and other infrastructure are expanded in a planned manner, leading to reverse causality and program placement bias. Unlike health and education programs, large infrastructure projects do not lend themselves easily to researcher manipulation and randomization. However, understanding the effects of investment in energy is important. Energy expenditures account for a significant fraction of total household expenditure, especially in developing countries (e.g. 24% in Cambodia) and unreliable energy access can have large effects on firm productivity (Straub 2008).

This paper examines the effects of electricity grid expansions in Brazil between the years 1960 and 2000 on local economic development. To address endogeneity issues, we develop a spatial engineering model of hydropower dam placement for Brazil which produces hypothetical maps that show how the electrical grid would have evolved over these 40 years had infrastructure investments been based solely on geologic cost considerations, ignoring demand-side concerns. This allows us to isolate the portion of the variation in grid expansion in Brazil that is attributable to exogenous cost considerations and use it as an instrument to estimate the development effects of the impressive growth in electrification in Brazil over this period. This empirical strategy takes advantage of the fact that Brazil relies almost exclusively on hydropower to meet its electricity needs, and the cost of hydropower dam construction depends on geologic factors such as water flow and river gradient, since hydropower generation requires intercepting large amounts of water at high velocity.

Geographic characteristics such as water and slope can affect development outcomes through

¹Some recent studies investigate the effects of irrigation dams (Duflo and Pande 2007), highways (Chandra and Thompson 2000, and Michaels 2008), and railroads (Atack 2009, Donaldson 2009, Banerjee, Duflo, Qian 2010). See Estache (2010) for a review of the literature on infrastructure impact evaluations.

²Aschauer (1989), Canning and Bennathan (2002), Esfahani and Ramirez (2003), Estache, Speciale, and Verdas (2005), Canning and Pedroni (1999), Hulten, Bennathan, and Srinivasan (2006), and Yeaple and Golub (2007) estimate macro growth effects of infrastructure expansion.

³Dinkelman (2010), Rud (2008), Assaduzzaman et al (2010), Ketlogetswe et al (2007), Grogan and Sadanand (2009), Khandker et al (2009), and Kammen and Mills (2009), Fan et al (2002) have examined the effects of electrification, and a subset of these studies have used instrumental variables strategies.

channels other than electrification, and thus using these variables as instruments in a cross-sectional regression would result in biased estimates. Instead, our approach is to produce a simulated time series of infrastructure development. To implement this, we provide an “engineer” (a Matlab model) with a budget of a certain number of hydropower dams and transmission lines in each decade based on actual data for the country as a whole. We then ask the model to allocate these dams and lines spatially so that the cost of construction is minimized. Water flow, Amazon location and river gradient are the only data the model has at its disposal to minimize cost, and it operates under the dynamic constraint that locations already electrified in a previous decade do not require new infrastructure. The modeled electricity grid evolves over time to minimize the engineering costs of dam and line construction, ignoring where people and businesses are located. We use the models output as an instrument to estimate instrumental variables (IV) regressions with location fixed-effects on county-level data for each decade.

We derive the formal econometric conditions under which this procedure can generate unbiased estimates of the effects of electrification on development outcomes in a two-stage IV estimation. That exercise reveals that unbiasedness depends on whether cost-side concerns in hydropower dam placement can truly be separated from demand-side concerns at the level of variation in the data.⁴ The procedure clearly fails under cross-sectional variation, since people and firms are more likely to be located in water-rich areas. The estimator is also biased in panel-data settings if people and firms move over time along the same spatial lines as the engineering model: from the lowest cost locations (robust water flow with a steep river gradient) in the early decades to slightly more expensive (flatter and less water-rich) locations in later years. Fortunately, expansion of economic and demographic settlements does not follow the pattern of modeled grid expansion, since the demand-side is attracted to water, but unlike hydropower plants, population settlements are repelled from areas with steep gradient. Taken together, this implies that the patterns of evolution of the demand side and of the engineer’s hypothetical hydropower dam placements over time are distinct, and demand and cost factors can be separated empirically, minimizing this bias. The evolution of other public services such as roads, sanitation and water infrastructure are not closely correlated with hydropower dam placement for similar reasons, and our results are robust

⁴If there is covariance between the geologic cost (C) and the demand (D) factors, that introduces bias in the estimated cost parameters (the coefficients on geology) in the “first stage” engineering model, since the demand-side factors are purposefully ignored in that model. This bias gets transferred to the second stage 2SLS estimates of the effects of electrification, and is a function of $cov(C, D)$. So $cov(C, D)$ conditional on observables (e.g. the location fixed effects) is the key identification challenge that can introduce bias in estimating the effects of electrification.

to controlling for these variables.

We find large effects of (lagged) electrification on two summary measures of development: a U.N. Human Development Index (HDI) computed for each county, and housing values imputed from census data on rents, under the assumption that improvements in living and working conditions in the county will be capitalized into rents. County HDI increases by 9 points (or 17 percent at the mean) when a county receives full access to electricity, with the gains concentrated in the income and education components. Housing values increase by 6 percent at the mean with a 10 percent increase in electrification. OLS regressions substantially under-estimate the gains from electrification, which is consistent with either the targeting of infrastructure to poorer areas, or the fact that compliers in the IV approach (i.e., the hydropower dams identified by the cost-minimizing engineering model) are the most cost-effective projects not built on the basis of political or other motivations.

These large development effects could be realized due to gains in firm and worker productivity, or they could simply reflect selective in-migration of the most productive workers and firms into electrified areas, creating a larger disparity between counties that get electricity and those that do not. To determine which of the two mechanisms is at play, we examine effects on a broader set of variables including in-migration, urbanization, salaries, employment, and population density. The small effects on across-county movement and county population density we estimate suggest that migration is unlikely to account for the large magnitude of development gains observed. We estimate large, positive effects of electrification on employment, salaries, formalization, returns to education and investments in education. The pattern of results suggests that electricity led to some broad-based improvements in labor productivity as workers gained both post-secondary education and work experience in the decade following electrification. The effects are of similar magnitude across sectors and across urban and rural economies. The development gains are concentrated in education, employment and income, but not in health.

Our estimation strategy is related to Duflo and Pande (2007) who use slope interacted with a time-varying state budget variable to predict irrigation dam placement, although our engineering model has a more complicated structure with multiple inputs (water flow and gradient) which help us distinguish the evolution of the electricity network from other infrastructure investments and from the evolution of demand. Our focus is on hydropower and not on irrigation dams, and we

document much larger development effects from electrification than Duflo and Pande (2007) find for irrigation. Our results are related to Dinkelman’s (2010) study of the employment effects of household access to electricity in a rural province of South Africa. While that study is able to delve into specific household mechanisms in one area, our data allow us to study the macro effects of electrification on a broad range of development outcomes over long time periods across a large developing country. In addition to addressing the identification challenges with the engineering model, another distinctive contribution of this paper is to report the long-run development effects of electrification over a forty year period.

The next section provides contextual information about the electricity sector in Brazil. Section three describes our data, section four describes our estimation strategy. Section five presents our estimates of the development effects of electrification and possible mechanisms, and section six concludes.

2 Background on the Electricity Sector in Brazil

Brazil provides an excellent setting to implement our estimation strategy because eighty-five percent of its electricity is generated from hydropower plants (US EIA 2010). This dependence on hydropower gives our engineering model strong predictive power in the first stage. Brazil is not alone in its dependence on hydropower: hydropower is the fastest growing source of electricity worldwide (US EIA 2010) and dominates the energy sector in Latin America.

Brazil experienced tremendous growth in electrification between the years 1950 and 2000, which provides variation with which to examine development effects and broadens the scope of our investigation and its external relevance for poorer countries that currently have low rates of electrification. The transmission network in Brazil grew at an average rate of 8.9 percent per year, increasing in size from 2,359 kilometers in 1950 to 167,443 kilometers in 2000 (SINDAT, 2000). Generation capacity has also increased: seven hundred and seventy-five major electricity plants have been constructed in Brazil since 1910 (SIGEL, 2008).

Electricity expansion in Brazil was not organized at the national level until a 1961 law that sought to coordinate the planning of the expansion of infrastructure in Brazil in order to increase economic development in the country. In addition to expanding the network in accordance with the government’s development goals, expansion plans in the Brazilian South and Southeast indicate

that load factors linked to local GDP and projected GDP growth played an important role in determining the expansion of the grid (Canambra Engineering, 1969). Development goals and anticipation of growth are important but opposing sources of selection bias in estimating the effects of electrification.

The 1961 legislation created a new national electricity company, Eletrobrás, which coordinated the financing of electricity projects and ensured that projects were in keeping with the government's overall development goals for the country. While Eletrobrás took control of the four existing regional (North, Northeast, South, and South Central) electricity companies, much of the planning of the electricity networks was devolved to the regional level. The fragmented system and the high cost of transmission explain why local infrastructure matters for local electricity access, and consequently, local long-term economic development.

The government initially expanded access in the 1960s and 1970s by increasing the number of isolated power generators, which provided power to local areas but was not transmitted further than the region (Canambra, 1968). Vast expansion in generating capacity during this period was made possible by high electricity rates and the easy availability of financing. Investment in the electricity network slowed in the mid 1970s due to reduced financial means and remained low through the 1990s, which caused a deterioration in network reliability (Gall, 2002). Consequently, the largest variation in our data is for the earlier part of our sample period.

Brazil reformed the electricity market in 1995, privatizing some of the state owned electricity companies and distribution companies, but political conflict and economic crisis weakened the reforms. The government continues to own eighty percent of the generation capacity in Brazil (Gall, 2002). Brazil invested in integrating transmission across regional quadrants in the 1990s but much local electricity continues to be sourced from local or relatively nearby plants (Gall, 2002).

The benefits of improved access to electricity accrue to many sectors of the economy. Industry is the largest energy consumer in Brazil—since 1970, the industrial sector has accounted for roughly half of Brazil's power usage. Agriculture represents a relatively small, but steadily growing, share of power usage. Its share of total power usage grew from less than 1% in 1970 to almost 4% by 2009. The public and commercial sectors' shares of power usage have not fluctuated or changed significantly since 1970, holding steady at about 10% and 15%, respectively. The residential sectors share of power consumption oscillates between approximately 20% and 25% (IPEA, 2010).

Twenty-seven percent of rural Brazilians still lack access to electricity (World Bank, 2005c), and infrastructure continues to be a development priority of the government of Brazil.

3 Data

3.1 Constructing Maps of Actual Electrification in Brazil 1960 - 2000

We assembled a GIS database of the locations of all major power plants and electricity transmission substations across Brazil from the 1960s until 2000 using a number of historical sources, such as the feasibility studies and inventories that the electricity companies in Brazil undertake prior to planning expansion of their networks. The power plant and transmission data come in two forms: (1) tables with inventories of all transmission lines that typically specify the county of origin, the destination county, length and voltage, and similar tables of power plants specifying location, type, and wattage; (2) large paper maps of generation plants and transmission lines by region of Brazil. Figures 3 and 4 provide examples of the maps and tables which were used to construct the data set. We digitize and combine this information into GIS maps of the Brazilian electricity network for the 1960s, 1970s, 1980s, 1990s and 2000.⁵ Power plants were placed on the digital map according to their reported latitude and longitude, while transmission substations were assumed to be located at the centroid of their county of record.

We collected data on generation plants and transmission lines, but not on the third component of the electricity grid: distribution networks. Transmission lines transfer electricity from the generation plants to the regions which are being supplied, while distribution networks transport electricity from the major local transmission substation to household, industrial and agricultural consumers of electricity. It was not possible to map these distribution networks over the period, because electricity distribution in Brazil is decentralized across sixty-four privatized electricity companies, and there is no central clearinghouse for data on their operations. We do, however, need to account for the distribution network, since assuming that only areas with substations have electricity is a very rough approximation. Based on conversations with electricity sector professionals in Brazil, we assume that on average distribution networks stretch one hundred kilometers across,

⁵The 1960s network is based on the comprehensive inventory taken by Canambra (1967) and Canambra (1969) for 1965 and 1967. The 1970s network is pieced together from various maps and tables from the different regions of Brazil from Eletrobràs. The 1980s network is based on another comprehensive inventory by SIESE (1987). The 1990s is again pieced together from various sources (e.g. Furnas 1993), and the 2000 network is based on SIESE (2000).

so all locations within a fifty kilometer radius of the centroid of a county containing a transmission substation are assumed to have access to electricity.⁶

Our spatial unit of observation is a grid of 33,342 evenly spaced points for all of Brazil set sixteen kilometers apart from each other. The use of grid points for this stage of the analysis allows us to circumvent the problem of endogenously drawn administrative boundaries and their changes over the period of the sample. Figure 5 shows the grid points and the construction of the data for Southern Brazil.

Figures 6 to 10 map the evolution of the electricity network in Brazil from the 1960s through 2000. The early development of the electricity network was focused in the relatively affluent and industrial south and from the 1970s onward the grid was expanded to the populous (but poorer) Southeast and Northeast. The network has expanded westward every decade since the 1970s, and by 2000, the coastal areas of the Southeast and Northeast had almost universal coverage. The Amazon and Pantanal areas have remained largely unconnected and continue to have substantially less access to electricity than the rest of Brazil.

3.2 Constructing the Instrument: An Engineering Model of Cost Minimization

Our instrument is a prediction of electricity access for each location across Brazil based on an engineering model which accounts for geologic cost factors over time. Our instrument–simulated electricity availability at each location in each decade–accounts for both the generation plants and the evolution of transmission lines and substations over time. We construct a simple engineering model of electrification in which decisions are made solely based on cost considerations determined by geography. The model generates predictions for whether each of the 33,342 evenly spaced grid points has electricity access in each of the five time periods of data between 1960 and 2000.

Our objective is to generate predictions for electricity availability for the five time periods of data on the electricity grid available to us: the 1960s based on inventories conducted in 1965 and 1967, 1970s based on maps from 1973 of the network, 1980s based on a comprehensive inventory conducted in 1987, 1990s and 2000 from the Brazilian National Electricity Agency (ANEEL) data on recent construction of major transmission lines. From the perspective of the engineering model, the specific dates for which we have data are essentially arbitrary, which implies that the scale of

⁶Figure 5 illustrates this assumption on distribution coverage: the dark blue polygons are counties which have transmission substations and the light blue circles surrounding them are assumed to be the distribution networks associated with those substations.

expansion between two periods—i.e. the number of new power plants and transmission lines built since the last inventory—is indeterminate. We match the scale of expansion between two periods to match the scale of investment in hydropower plants observed in the data. In other words, we allocate a budget of 240 power plants to the model in the 1960s because that is the number of hydropower plants in existence in Brazil by 1967—our inventory date for that decade. By similar reasoning, the budget for 1970s was 53 additional power plants, 36 additional plants for 1980s, 25 additional plants for 1990s, and 24 additional plants for 2000. The model takes these country-wide budgets as given and chooses the optimal location of hydropower plants and transmission lines within Brazil based purely on geological factors.

Even though electricity networks are planned in large part to meet local demand, access to electricity in a country that relies heavily on hydro-power has an exogenous geographic component to it because the cost of access depends on the suitability of local generation. In evaluating a new location for a hydro-power plant, engineers consider available head, flow duration, and daily peaking operation to determine generation cost (Gulliver and Arndt, 1991). Available head is determined by the amount of water flow and the change in elevation between the top and bottom of the dam. The head determines the amount of power that will be produced, and generating a given amount of power is cheaper in locations where the available head is larger. The flow duration is determined by the amount of time in a given month (day, or year) in which a given flow rate required by the turbines in use is equaled or exceeded. The daily peaking operation is the amount of the flow duration which occurs during peak demand hours (Gulliver and Arndt, 1991). In addition, in making site decisions, engineers typically consider distance to the existing transmission network, as developing new transmission lines is expensive and can comprise a large component of the overall budget for the network (Canambra, 1968).

The model begins by creating a measure of the cost of building a hydropower plant in each location by combining the various geologic factors and assigning a cost parameter to each factor. The geologic cost factors include whether the location (a circle of radius ten kilometers around each grid point) has a river, the average and maximum gradient of the river, maximum water flow accumulation anywhere within that circle, and an indicator for whether that location falls in the Amazon. The geologic inputs into the model are calculated based on GIS maps from U.S. Geological Surveys Hydro1k program. Using a GIS map of water bodies, we create two kilometer buffers on

either side of each river, and compute the gradient along the river using elevation maps.⁷ We use the coefficients from a probit regression of hydropower plant location in Brazil on the geologic measures to assign the cost parameters to each grid point on our map. Table 1 reports the results from the probit regression of hydropower placement on geologic factors. The model uses these cost data to rank all grid points by the cost of hydropower dam construction, and subsequently allocates plants to the lowest cost locations until the budget for the decade is exhausted.

As a robustness check we estimate the same probit regression of power plant placement on geologic factors using data for the United States.⁸ If we were concerned about the use of Brazil data capturing non-engineering cost factors, we would expect the US model to allocate the geologic cost factors differently than the Brazil data. We find that the allocation of electricity is quite similar across the two models. In the US model, we estimate the geologic factors in the same way as the factors were estimated in Brazil; we calculate flow accumulation, maximum and average slope, and regress the placement of hydropower in the US on these factors. We then generate the model in Brazil based on the US probit coefficients. The results are qualitatively similar, but we gain the greatest precision using the Brazilian data.⁹

Figure 11 maps the locations of the 240 power plants that were predicted based on 1960s geologic cost considerations. The red dots represent the predicted plants, the yellow dots the actual plant locations, and the color of the background reflects the elevation (darker colors are closer to sea level), and the blue lines show river locations. As expected, hydropower plants are most likely to be located in areas with steep river gradient, with greater water flow, and away from the Amazon. Our model predicts a large number of power plants along the Southeast to Northeast corridor (São Francisco river basin) where elevation changes quickly from the low-lying coastal areas, implying a steep increase in slope.

The model next predicts the locations of substations (i.e. directions of transmission lines) that deliver the electricity generated at each plant predicted from the previous step. We make the

⁷To create a separate control variable for average land slope to use in a few cross-sectional specifications, we draw circles of radius ten kilometers around the evenly spaced grid points throughout Brazil and compute average slope in those circles using the elevation map. See Appendix C, figure 2 for an example of the creation of the river gradient variable.

⁸The use of Brazilian data to parameterize the cost function may introduce concerns about endogenous placement bias, but the formal econometric derivation in the next section shows that in our panel specifications, only if endogenous demand-side parameters move over time in a similar sequence from the lowest-cost engineering locations and up the cost ladder would those factors introduce endogeneity bias.

⁹This is not surprising since hydropower plants and water resource conditions in the U.S. are different from those in Brazil, and geography therefore matters in slightly different ways.

simplifying assumption that all power plants have the same generation capacity. We assume that each plant has exactly two transmission substations which are connected through a single line; this assumption was made to be consistent with the data on the average number of transmission substations per hydropower dam (SINDAT, 2008). The electricity network is assumed to be fully durable and new substations and power plants cannot be placed in locations which have already received electricity in prior decades. New substations are also not placed in locations that receive generation plants in the same decade from the first step of the model.

The model arrives at the optimal lowest cost electricity network in each decade by computing costs for all possible arrangements of transmission lines. There are a finite but arbitrarily large number of possible permutations of transmission lines, and the numeric method we use to arrive at the lowest-cost grid in equilibrium is detailed in Appendix C. The model assumes that cost increases with distance and is prohibitively high when building substations in the Amazon (due to high material transport costs). The empirical section of the paper contains robustness tests controlling for an Amazon indicator and an Amazon-specific time varying trend to ensure that this assumption does not drive our results. The model predicts short transmission lines, so that substations end up being located very close to generation points.

Once the equilibrium set of transmission lines is determined, we assume that all grid points within a fifty kilometer radius of any substation will receive access to electricity, which accounts for the distribution network surrounding that substation. In other words, we purposefully remain agnostic about the direction in which the distribution networks are expanded. The chosen fifty kilometer radius is based on average size of distribution networks and mirrors our treatment of distribution networks in the electricity data. In subsequent decades, new power plants are placed in the highest probability circles among those that have not yet received electricity and locations for transmission substations are proposed from among the points that remain without electricity in the previous decades.

Figures 12 to 16 plot the areas predicted to receive electricity by this model by decade. There is a reasonably good cross-sectional spatial correlation with the actual electricity network for Brazil (figures 6-10), and there are encouraging signs of correlation in terms of direction of expansion. The strength of this correlation in a model with location fixed effects determines the predictive power in our first stage.

Ignoring the demand-side concerns forces the model to under-allocate electricity to places like São Paulo and Rio de Janeiro, which implies that the engineer remains endowed with that extra generation capacity which she must allocate elsewhere. This leads our hypothetical maps for the early decades to display broader spatial electrification coverage than what is observed in Brazil. This has the effect of weakening the predictive power of the model to some extent, but we find that the model still has a strong first stage.

4 Estimation Strategy

4.1 Dependent Variables and Equation Estimated

This study examines the effect of electricity on development outcomes over the period 1960 - 2000. We start with two summary measures of local development as dependent variables: a U.N. human development index measured for each county and average housing values in the county. Housing values are estimated from census questions on rents and imputed rents. Local development and improvements in amenities should be capitalized into rents, so we view the average housing value as a summary development measure that proxies for welfare changes in the locality.¹⁰ The U.N. Human Development Index is an index varying from zero to one, combining data on longevity, income and education. We also examine the effects of electrification on additional dependent variables in an attempt to uncover mechanisms through which electrification affects development. Summary statistics and definitions of these dependent variables are available in tables two and three.

Brazilian counties have changed borders numerous times over the period of our analysis. Since we do not have digital county maps from 1960, we are unable to assign the grid points (our spatial unit for the construction of the instrument) to the exact county every decade. The Brazilian statistical agency has defined larger standardized county units (a smaller set of old large counties), which we use to construct a valid county time series. Although electricity availability is predicted at the grid point level, socio-economic data is at the county level, so we collapse our dataset to the large standardized county units and cluster errors at this level. While this is a conservative way of estimating our standard errors (we actually have more variation in our independent variable),

¹⁰Land values have been used in the environmental literature to measure welfare changes through hedonic methods. See, for example, Greenstone and Gallagher (2008), and Davis (2004).

this estimation strategy guards against the possibility of mis-specification due to changes in county borders. We have more precise information on the variable of interest (electrification) and on the instrument (simulated electrification) in counties with larger areas, owing to the fact that there are a larger number of grid points in larger counties. We therefore run weighted regressions where the weights are county area.

We regress development outcomes in county, c , and decade, t , on electricity lagged by a decade in a 2SLS model with county fixed effects:

$$Y_{ct} = \alpha_c^1 + \gamma_t^1 + \beta \hat{E}_{c(t-1)} + \epsilon_{ct} \quad (1)$$

where \hat{E} is electricity provision, predicted on the basis of the engineering model in the first stage:

$$E_{c(t-1)} = \alpha_c^2 + \gamma_t^2 + \theta Z_{c(t-1)} + \eta_{ct} \quad (2)$$

E_{ct} is the proportion of grid points in the standardized county covered by the distribution network surrounding generation plants and transmission substations in each decade (figures 6-10). Z_{ct} is modeled electricity: the proportion of grid points in a standardized county predicted to be electrified by the simulated engineering model.

We lag infrastructure development by a decade in all specifications. The development of the distribution network may take several years to complete following the construction of the hydropower dams, so a lag of ten years is the appropriate timing for the estimation of the development effects of electrification. While there may be immediate regional economic effects from the construction of dams and hydropower plants, these would be primarily short term and focused in the vicinity of the power plant. We model the placement of transmission lines as well as the generation plants, so we are able to model where people actually have access to electricity, rather than the effects at the building sites of the dams.

4.2 The Source of Identification, and Bias Concerns

In this section we derive the formal econometric conditions under which the IV strategy based on the simulated engineering model yields unbiased estimates. This allows us to be clear about the underlying source of identification. We adopt the notation and adapt the methodology set out in

Hahn and Hausman (2002) (henceforth HH). HH study the following two-stage model:

$$y = E\beta + \epsilon = z\theta + \nu \quad (3)$$

$$E = z\theta + \eta \quad (4)$$

Applied to our problem, y represents the development outcome, E the endogenous independent variable of interest (Electricity), and z the cost variable (Modeled Electricity).¹¹ We are interested in formulating an expression for bias in the 2SLS estimator, $\beta_{2SLS} - \beta$. A potential for bias is introduced because we are estimating the first stage as $E = z\theta + \eta$, even though the true data generating process may be: $E = z\theta_1 + d\theta_2 + \eta$ (where d represents the demand factors).

Ignoring the demand term in the first stage equation potentially biases the estimate of θ_1 , and the size of this bias is exactly $\theta_2 \frac{\Sigma z' d}{\Sigma z' z}$. This bias can also be written as $\theta_2 \frac{cov(z, d)}{var(z)}$. Since β_{2sls} is a function of $\hat{\theta}_1$, this mis-specification in the first stage can also bias the 2SLS estimator.

$$\beta_{2SLS} = \frac{\Sigma \hat{E}y}{\hat{E}\hat{E}} = \frac{z\hat{\theta}_1 y}{z\hat{\theta}_1 z\hat{\theta}_1} = \frac{\Sigma z\hat{\theta}_1(\beta z\theta_1 + \nu)}{\Sigma z\hat{\theta}_1 z\hat{\theta}_1} \quad (5)$$

Plugging in the definition of $\hat{\theta}_1$ ($\hat{\theta}_1 = (z'z)^{-1}z'E = (z'z)^{-1}z'(z\theta + \eta) = \theta_1 + (z'z)^{-1}z'\eta$), the bias in the 2SLS estimator equals:

$$\beta_{2SLS} - \beta = \frac{\Sigma z\hat{\theta}_1(\beta z\hat{\theta}_1 - \beta z(z'z)^{-1}z'\eta + \nu)}{\Sigma z\hat{\theta}_1 z\hat{\theta}_1} - \beta \quad (6)$$

The numerator of the equation above can be simplified using the definitions of η and $\hat{\theta}_1$:

$$\beta_{2SLS} - \beta = \frac{\Sigma z\hat{\theta}_1(-\beta z(z'z)^{-1}z'\eta + \nu)}{\Sigma z\hat{\theta}_1 z\hat{\theta}_1} \quad (7)$$

The numerator of equation (7) can be simplified using the definitions of η and $\hat{\theta}_1$:

$$\Sigma z\hat{\theta}_1(-\beta z(z'z)^{-1}z'\eta + \nu) = \Sigma z\hat{\theta}_1(-\beta z(z'z)^{-1}z'(E - z\theta_1) + \nu) = \Sigma z\hat{\theta}_1(-\beta z(\hat{\theta}_1 - \theta_1) + \nu) \quad (8)$$

¹¹In this specification, we assume that all other right hand side variables (including fixed effects) have been partialled out (as in Hahn and Hausman, 2002). Hahn and Hausman (2000) shows that the case is similar when other right hand side variables are included.

The appearance of $\hat{\theta}_1 - \theta_1$ in the numerator implies that the bias in β_{2SLS} is positively proportional to the omitted variable bias in the first stage from ignoring the demand factor. In other words, the 2SLS estimator is biased if there is covariance between cost (our first stage instrument) and the omitted demand factor:¹²

$$\beta_{2SLS} - \beta = \frac{\left(\nu - \beta z(\theta_2 \frac{cov(z,d)}{Var(z)})\right) z' \hat{\theta}_1}{\Sigma(z' \hat{\theta}_1)^2} \quad (9)$$

The exclusion restriction for the instrument will not be met in our application if the exogenous cost factors and demand co-move at the level of variation in our data, conditional on our controls. This is because the instrumental variables approach relies on our ability to separate cost and demand factors in electricity allocation so that we can isolate the portion of the variation that occurs due to the geology-based cost minimization.

In cross-sectional data, cost is determined by water flow and river gradient, and these variables also determine the location of population and economic activity (which determine electricity demand), so our approach would not be valid. In a panel setting, our engineering model predicts that over time hydropower dams will move from the lowest cost (robust water flow with a steep gradient) locations to less water-rich and flatter locations.

The question of interest to determine potential bias in the FE-IV specification is whether demand evolves over time along the same spatial pattern as the engineering model's movement from lowest cost to higher cost locations. Conceptually, this appears unlikely because people and firms are likely to locate in water-rich areas (a positive factor in the engineering model), but they also prefer flat areas (a negative factor). So the spatial evolution of the demand side and the cost side are likely to be distinct and $cov(z, d)$ likely to be low once location fixed effects are added.

We can also examine the empirical relevance of this concern in the data. Table four shows that while in the cross-section population density is correlated with the engineering cost of construction, this correlation becomes close to zero once location fixed effects are added. Similarly, the correlation of engineering cost with a county-level GDP and industrial GDP per capita measures become smaller (and statistically indistinguishable from zero) with location fixed effects. The bottom two rows of the table show that these population and economic concentration measures are more correlated with actual electricity infrastructure placement than with the models prediction of engineering costs. This suggests that the simulated instrument which combines river gradient, budget, and

¹²HH simplifies the denominator to $R^2 \Sigma y^2$ where R^2 measures the fit of the first stage regression

water flow does isolate the exogenous component of the variation in electrification.

5 Empirical Results

5.1 First stage Results: Predicting Electrification using the Engineering Model

Table 5 shows our first stage results that predict the locations of actual electricity infrastructure using the simulated instrument produced by the engineering model. Standard errors are clustered by the 2,184 standardized counties in all specifications. The fit is very strong when using all cross-sectional variation in the data as visual comparison of the actual and simulated maps in figures 6 to 10 and figures 12 to 16 suggests. The second specification indicates that much of this is driven by the large differences in electrification between the Amazon and non-Amazon regions of Brazil. The coefficient on the simulated instrument is reduced considerably once state fixed effects are added. Our preferred specification adds location (standardized county) fixed effects in the third column. This first stage regression is the basis for all the two-stage regressions reported in this paper. Using only the within-county variation in the data, we find that areas predicted to have electricity in a given decade by our engineering model are thirty-two percentage points more likely to have electricity that decade. The F-statistic on the first stage is 34.7.

5.2 Second Stage Results: Effects of Electrification on Development

Tables 6 and 7 analyze the effects of electrification on two summary measures of a county's development: the average value of the housing stock and a human development index measured for the county. Across all OLS and IV specifications, the effect of electrification on subsequent changes in average housing values is large and positive. In OLS regressions without fixed effects (i.e. using all cross-sectional variation in the data, including the variation between Amazon and non-Amazon regions), moving from no electricity to full electrification is associated with a 4,667 reais increase in average housing value, which represents a 35% increase at the mean. Adding region, state and location fixed effects in columns two through four reduces the magnitude of this effect to 1,326 reais. The IV estimates using the simulated instrument from the engineering model are larger than the corresponding OLS estimates. In the preferred IV specification with county fixed effects, a 10% increase in electrification of the county increases average housing values by 6% at the mean.

As shown in table 7, the estimated effect of electricity on the human development index is also positive. In OLS regressions with or without state or region dummies, electrification is associated with about a 3 point increase in the county's human development index score. With the addition of county fixed effects, the magnitude reduces to a statistically insignificant 1 point increase. Our preferred IV specification (with county fixed effects) suggests that moving from zero to full electrification increases the county HDI score by 9 points. Since the HDI is an index score based on sample values, it is instructive to interpret this result in the following way: the 9-point increase would take the median county in Brazil in 1980 to the human development level of the 69th percentile county. This represents a significant move within the distribution of HDI.

Comparing across the OLS specifications in tables 6 and 7, we see that cross-sectional estimates are large, and the magnitude of the effects of electrification decrease when it is estimated within-region, within-state or within county. We also see that IV estimates of the effects of electrification are substantially larger than the OLS estimates. There are three possible reasons for the downward bias in OLS relative to IV in our data. First, the electricity variable is measured with substantial error (we have the exact placement of hydro-power plants, but we can only place transmission lines in the centroid of the county in which they are located, rather than their exact GIS location), while the geography variables used to predict the placement of electricity are measured quite precisely (based on 1km by 1km satellite maps). The IV estimates may be correcting the measurement error in the independent variable and addressing the associated attenuation bias.

Second, the IV estimates measure the returns to electrification from a selected set of the most efficient, lowest-cost hydropower plants. The OLS estimates, in contrast, report the development effects of the current distribution of hydropower dams and sub-stations in Brazil, many of which may have been placed on the basis of political and other considerations. Thus the larger IV estimates may reflect the fact that the rates of return on geographic cost-minimizing placement of electricity infrastructure are larger than infrastructure allocated on the basis of politics or other non-cost considerations. Areas that received electricity primarily because of the low cost of provision rather than a socio-economic, political or other demand-side pull may have derived the greatest benefits. Consistent with this interpretation, Cadot, Roller and Stephan (2006) show that transport infrastructure is highly susceptible to politically motivated allocations. Engel, Fischer, and Galetovic (2009) show that even once developed, public works projects may not be adequately

maintained because of political considerations. And in Brazil, the allocation of publicly provided health services has been shown to be subject to political considerations (Mobarak, Rajkumar and Cropper 2011).

Third, as described in detail in section 2, the electricity network in Brazil was designed and expanded primarily by the government or government managed utility companies during the period covered by our data (1950-2000). The demand-side endogeneity bias that the IV estimation corrects may have been of the form of the government targeting poorer areas important for maintaining political support (such as the program Luz para Todos) rather than more intensive expansions in developed areas where demand is likely to be greatest. OLS estimates would be biased downward due to the government's promotion of its development objectives.

5.3 Exploring the Mechanism underlying the Development Effects of Electrification

Estimating the large positive county-level development effects of electrification does not by itself imply that there are true productivity gains that justify greater national-level investment in infrastructure. This is because electrification can induce movement of people and firms in general equilibrium, and the large gains in human development and housing values we observe may simply be a result of the re-sorting of productive workers and firms toward electrified areas. Alternatively, electrification may lead to real gains in employment and labor productivity if it allows firms to use capital more efficiently or if it changes workers' incentives or ability to invest in human capital. In this section, we will examine the effects of electrification on a broad range of outcomes on which county-level long-term time series data are available as a way to gauge which of the two mechanisms underlie the observed positive development effects of electrification.

In table 8, we estimate the impact of electrification on each of the three components of human development - longevity, education, and income. We find that the development gains are concentrated in the income and education sectors, and not in health. The effect of electricity on life expectancy is both statistically insignificant and very close to zero. This is consistent with the possibility that electrification has conflicting effects on health - it allows for improvements in health technology and service delivery, but it may also increase pollution and strain through expansions of heavy manufacturing industries.

The estimated effect of electrification on average household income is quite large and positive in the FE-IV specification, but negative in OLS (suggesting a downward bias from the government’s targeting of under-developed areas). Going from no electrification to full electrification takes the median Brazilian county in 1980 to the income-HDI level of the 75th percentile county.

The education component of the U.N. Human Development Index is comprised of literacy and school enrollment. A county gaining access to full electrification leads to a gain of 14 index points in its education score according to the FE-IV specification. This represents a move from the 50th percentile to the 93rd percentile county in 1980.

The last few columns of the table examine effects on direct household income and poverty measures, as opposed to index values. The development gains are large - a 10% increase in electrification reduces poverty by about 5% at the mean.

5.3.1 Employment Effects by Sector

Having established that the gains from electrification are concentrated in income and education (and not in health), we next try to get a sense of (a) whether the income effects are realized due to better employment conditions, and (b) the sectoral distribution of the gains - in formal and informal sectors and in urban versus rural areas. Table 9 reports the effects of electrification on a broad notion of being “economically active” (including self and informal employment), then on the narrower concept of formal sector employment, and finally, on formal employment in urban and rural areas separately.

We see that electrification holds similar large positive benefits to both formal and informal or self employment in table 9. Going from zero to full electrification is associated with a 13-15 percentage point increase in the probability of employment. Again, the effects are much larger in the preferred fixed effects IV specification than in the OLS specification. Furthermore, the gains are very similar in both urban and rural areas within counties.

The 13-15 percentage point increase in employment probability represents a 40% improvement in mean employment conditions across Brazilian counties. This helps rationalize the large gains in income we observed in table 8. The similarities in the size of these effects across sectors and locations suggest that the broad-based improvements in labor skills and education (observed earlier) are more likely to explain these effects than narrowly focused improvements in the quality and

efficiency of capital in manufacturing or any one other sector. While it is possible that industry gains when factories can use better electric-powered capital, and agriculture simultaneously gains with better irrigation technology, and there are downstream general equilibrium effects on the informally employed and self-employed in small-scale manufacturing and services, the remarkable consistency in the magnitude of the employment effects across sectors and locations seems quite unlikely. The more likely mechanism at play is a broad-based gain in labor productivity following electrification.

5.3.2 The Source of Labor Productivity

Improvements in labor productivity are likely related to the large improvements in educational attainment observed in table 8. In table 10 we try to identify at what level these education gains occurred: were they concentrated in reducing illiteracy, improving primary education, or increasing school enrollment at all levels? Going from none to full electrification leads to a seven percentage point drop in the illiteracy rate in the county, which represents a 22% drop at the mean. The proportion of the population with less than four years of education drops twelve percentage points, or a 19% decrease at the mean. School enrollment across all levels experiences the largest gain: years of schooling in the population increases by 1.2 years, which represents about a 45% increase at the mean. In summary, while there were significant gains in reducing illiteracy and improving primary enrollment, the largest magnitude of gains from electrification appear to be in post-primary education.

Beyond education, the greater work experience that workers accumulate due to improved employment conditions may have also contributed to the improvement in labor productivity. To examine this channel, we look at the effect of electrification on a stock of human capital variable created by the Instituto de Pesquisa Econômica Aplicada (IPEA). IPEA estimates the human capital variable by running Mincer regressions using census and survey data to estimate the returns to education and experience, and uses these estimates to monetize and combine the gains in average education and work experience in counties. The effect of electrification on the stock of human capital in a county is quite large—going from no electrification to full electrification generates a 53% increase at the mean. This again suggests that the gain in labor productivity is associated with the accumulation of skills at the middle and upper end of the distribution through both primary

and post-primary education and work experience.

5.4 Real Productivity Gains or the Effects of Migration and Sorting?

We have documented positive income and employment effects of electrification across sectors and locations, and these effects appear most consistent with broad-based labor productivity improvements in newly electrified areas. The productivity gains may have been a real effect of electrification (e.g. increases in efficiency and returns to education allow workers to invest more in education, quality of complementary inputs like capital increases, or there is better accumulation of work experience as employment conditions improve), or it may be explained by re-sorting of workers through migration into electrified areas. Mobility of workers and firms can increase real productivity gains (e.g. employment conditions improve as firms move in to electrified areas, and this subsequently leads to greater accumulation of human capital). Our goal in this sub-section is to examine whether the large development effects we have observed can be fully explained by migration and re-sorting.

We look for evidence on this directly by examining the effects of electrification on in-migration in each county. Migration data is only available for the 1990 and 2000 census, which considerably shortens our panel. The point estimate in the preferred fixed effects IV specification indicates that a move from zero to full electrification doubles the influx of migrants into counties, but the standard errors around this estimate are very large, possibly due to the short panel. However, an important point to note is that only 7% of the average county's population is comprised of recent migrants, and thus even a doubling of the in-migration rate does not change the composition of the population dramatically. A 10% increase in electrification would increase the migrant share of the population from 7.2% to 8.0%. It is unlikely that this increase in the migrant share could account for the 5% (or 3.4 percentage point) reduction in poverty and 4% improvement in employment conditions associated with that 10% increase in electrification. Therefore, even taking the large (but imprecisely estimated) coefficient in the migration regression at face value, it can only explain a small portion of the development gains.

We next examine effects on county population density and within-county urbanization to look for further evidence of changes in population composition using variables for which we have a longer panel. The preferred FE-IV estimate of the effect of electrification on population density is actually smaller than the OLS estimate, which suggests that the government may have targeted

electricity to more densely populated (but poorer, less developed) regions. The correctly estimated effect of electricity on population density is small and statistically insignificant, further bolstering the case that (a) a change in population composition is unlikely to explain the large magnitude of development and productivity gains from electricity that we have documented, and (b) the influx of migrants induced by electrification does not change overall population composition that much, since migrants are a small share of the overall population. The urbanization measure does provide evidence of substantial within-county sorting following electrification. Going from zero to full electrification leads to 12 percentage point more of the county population being classified as “urban,” which could either be a result of rural residents shifting towards the population centers within counties, or because the greater agglomeration leads to more of the county being classified as urban by the statistical agency. Either way, this is a within-county move, and cannot explain away the cross-county estimates of productivity gains associated with electrification.

To summarize, we first observe large improvements in overall development measures following electrification, with gains concentrated in income and education (but not health). The improvements are equally large across sectors and across urban and rural areas, which are suggestive of broad-based increases in labor productivity. Workers accumulating both more education and more experience appear to be contributing to the productivity increase. Migration and re-sorting of workers and firms cannot explain away the magnitude of the development gains observed, which suggests that electrification brought about some real improvements in labor productivity.

5.5 Robustness Checks

This section reports the results of additional regressions designed to examine various threats to identification. The first important concern we address is the possibility that electricity proxies for a broader package of infrastructure investments. Infrastructure is sometimes delivered as a package, and solving the logistics associated with constructing transmission lines may itself lead a parallel road being built or other infrastructure services being delivered more efficiently. Our IV strategy is designed to mitigate this concern, and it is particularly helpful that our engineering model based instrument’s reliance on both water (which likely lowers cost of delivery and attracts infrastructure) and gradient (which likely increases costs and repels other infrastructure) makes the spatial patterns of hydropower generation and of the delivery of other types of infrastructure distinct. Nonetheless,

table 12 examines the sensitivity of our results to inclusion of other infrastructure control variables.

We include controls for the percentage of households in the county with running water and with improved sanitation access computed from the decennial census data. Unfortunately, we do not have any direct measure of the evolving development of roads over time. We do have car ownership data from the last two censuses, but only one of those two years overlaps with our sample, so even this imperfect proxy for roads is not directly useful for our regressions. Instead, in table 12b we show that the growth in car ownership is positively correlated with water availability in the county interacted with a time trend, and negatively correlated with land slope interacted with a time trend. Motivated by these two correlations, we use the water trend and the land slope trend as proxies for the road network.

Table 12 shows that the effects of electrification on our summary measures of local development (average housing values and the county human development index) remain positive, statistically significant and of similar magnitude after controlling for other these other measures of infrastructure (access to running water, improved sanitation, and indirect proxies for roads). Two important shortcomings of this robustness check are (a) the provision of other infrastructure may also be endogenous, and we do not have exogenous instruments for their availability, and (b) our proxy for the growth of the road network is indirect.

The first two columns of tables 13 and 14 use a census-based measure of household electrification instead of the measure constructed from historical maps of generation plants and transmission lines. The fraction of households with electricity access may be less susceptible to concerns about some types of measurement error, and also provides a direct measure of household-level connectivity. The within-year correlation between the infrastructure variable and the census variable on household electrification rates is 0.545. Receiving electricity infrastructure in the county raises the household electrification rate by 9 percentage points in a regression controlling for year and county fixed effects.

The instrument has a poorer fit for the census electrification variable in the first stage, probably because our engineering model is designed to predict the optimal placement of electricity infrastructure, whereas variation in the extent of household electrification is more directly determined by other demand factors. Reassuringly, estimates of the development effects of electrification remain qualitatively and quantitatively very similar when the census variable is used. A 9 percentage point

increase in the household electrification rate (which is the average effect of getting infrastructure in the county) increases the Human Development Index by about 9 points, which is almost exactly the magnitude of this effect when estimated with our infrastructure data. The estimated effects on housing values are also very similar across the two specifications.

The engineering model we develop to construct the instrument uses three geographic variables as inputs: availability and flow of water, the gradient of the river, and the location of the Amazon. The Amazon is a unique region in many respects, and also has very low rates of electrification. While we include location fixed effects in all our preferred specifications, there may be a concern that the Amazon follows a very different development trend, and the differential evolution of Amazon and non-Amazon regions coupled with the differential rates of electrification explains our results. In the third column of tables 13 and 14, we flexibly control for a differential Amazon trend by including a full set of interactions between the Amazon dummy and the decade dummies as regressors. Furthermore, we also include a full set of interactions between a water availability dummy and the decade dummies as regressors to allow for water rich areas to have a differential time trend. Reassuringly, the estimates of the development effects of electrification remain quantitatively and qualitatively very similar after these controls are added. Going from no electrification to full electrification increases the human development index by about 10 points in this specification.

The first step of our engineering model uses data from Brazil to parameterize the cost function for building hydropower dams. A possible concern with this approach is that assigning the relative importance of water flow, river gradient and the Amazon using Brazilian data may introduce an element of demand-side preferences specific to Brazil in determining hydropower dam placement. We therefore re-estimate the cost function using data from the United States rather than Brazil, re-calibrate the engineering model and generate a new instrument. This instrument is used in the fourth columns of tables 13 and 14. The geology of hydropower generation in the U.S. is very different from Brazil because the two countries have very different levels of water resources and differential reliance on hydropower relative to other sources of energy. Accordingly, we lose some predictive power, and also find that the estimated effect of electrification on development is only half as large, although still positive under this specification.

6 Conclusion

Unreliable energy in the developing world is viewed by firms as a significant obstacle to doing business (Straub 2008), and donors and governments have been persuaded to invest in large-scale infrastructure projects. There remain good competing demands on development funds, which make it important for social scientists to inform policy-makers about the returns to infrastructure investment. The findings of this study are also relevant for Brazil specifically, as there is strong debate over the construction of new hydro-electric dams, particularly in the Amazon. Former President Luiz Inacio Lula da Silva committed to increasing hydro-electric energy in the country as a priority of his government, and current president Dilma Rousseff is known to be expanding infrastructure as a priority of her government (New York Times, 2005; The Economist, 2010).

Our study addresses this challenging question by isolating an exogenous portion of the variation in electrification across Brazil based on engineering cost factors. The methodology and engineering model we develop can be useful for studying the effects of hydropower investments in other countries, and the general concept can be applied to a broader range of infrastructure projects. The empirical section documents large development gains from investments in electricity, which are substantially under-estimated when one fails to account for the endogenous placement of projects. These gains are concentrated in education, employment and income. The development gains are of similar magnitude across sectors and areas, and they are suggestive of broad-based improvements in labor productivity in the decade following electrification. Importantly for national policy, the magnitude of the development effects is unlikely to be explained by selective in-migration of productive people and firms. The productivity and development gains are thus a value-added to society, and not just re-sorting of already productive resources in general equilibrium.

A References

- Asaduzzaman, M., Douglas Barnes, and Shahidur Khandker, (2010). "Restoring Balance: Bangladesh's Rural Energy Realities." World Bank Working Paper #181.
- Atack, Jeremy, Fred Bateman, Michael Haines and Robert A. Margo. 2009 "Did Railroads Induce or Follow Economic Growth? Urbanization And Population Growth in the American Midwest, 1850-60" NBER Working Paper 14640.
- Ali, I. and E. Pernia (2003). "Infrastructure and Poverty Reduction: What is the Connection?" ERD Policy Brief Series No. 13, Economics and Research Department, Asian Development Bank, Manila, Philippines.
- Arellano, Manuel and Stephen Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*. V.58 No.2. p.277-297.
- Aschauer, D.A. (1989). "Is Public Expenditure Productive?" *Journal of Monetary Economics*, 23, 177-200.
- Banerjee A., E. Duflo, N. Qian (2009). "On the Road: Access to Transportation Infrastructure and Economic Growth in China," mimeo.
- Baldacci, E. N. Guin-Siu, L. de Mello (2003). "More on the Effectiveness of Public Spending on Health and Education: A Covariance Structure Model," *Journal of International Development*, 15(6), 709-725.
- Bensch, Gunther, Jochen Kluge, and Jorg Peters. "Rural Electrification in Rwanda: an Impact Assessment using Matching Techniques." mimeo.
- Bloom, D., D. Canning, and J. Sevilla (2001). "Health, Human Capital, and Economic Growth Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons." Commission on Macroeconomics and Health Working Paper #WG1:4.
- Briceno-Garmendia, C. Stache, A. Shafik. (2004). "Infrastructure Services in Developing Countries: Access, Quality, Cost and Policy Reform." World Bank Policy Research Working Paper No 3468, Washington D.C.
- Cachapuz, Paulo Brandi de Barros (2002). "O Planejamento da expansão do setor de energia elétrica : a atuação da Eletrobràs e do Grupo Coordenador do Planejamento dos Sistemas Elétricos." Eletrobràs, Rio de Janeiro.
- Cadot, Olivier, L. Roller, and A. Stephan (2006). "Contribution to productivity or pork barrel? The two faces of infrastructure investment." *Journal of Public Economics*. V.90 pp. 1133-1153.
- Calderon, C. and L. Serven (2004). "The Effects of Infrastructure Development on Growth and Income Distribution." World Bank Policy Research Working Paper Series 3400, World Bank, Washington DC.
- Canambra engineering consultants, 1968 *Power Study of South Eastern Brazil*. Rio de Janeiro: UNDP.
- Canambra engineering consultants, 1969 *Power Study of South Brazil*. Rio de Janeiro: UNDP.
- Canning and Bennathan (2002). "The Social Rate of Return on Infrastructure Investments," Policy Research Working Paper Series 2390, World Bank, Washington DC.
- Chandra, Amitabh and E. Thompson. 2000. "Does Public Infrastructure Affect Economic Activity? Evidence from the rural interstate highway system" *Regional Science and Urban Economics* 30(2000): 457-90.
- Coelho dos Santos, Silvio, and Maria Jose Reis. *Memória do Setor Elétrico na Região Sul*. Sao Paulo: UFSC.

- Davis, Lucas (2004). "The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster." *American Economic Review*. v.94, n. 5 pp. 1693-1704.
- de Souza, Antonio Carlos Caetano (2007). "Assessment and statistics of Brazilian Hydro-Electric Power Plants: Dam Areas Versus Installed and Firm Power." *Renewable and Sustainable Energy Review*, 12(7): 1843-1863.
- Dinkelman, Taryn (2010). "The Effects of Rural Electrification on Employment: New Evidence from South Africa." Forthcoming, *American Economic Review*.
- Donaldson, David. 2008 "Railroads and the Raj: the economic impact of transportation infrastructure" mimeo LSE.
- Duflo, E. and R. Pande (2007). "Dams," *Quarterly Journal of Economics*, 122(2):601-646.
- The Economist. December, 2010. "Brazil's new President: Coming down to Earth."
- Eletrobrás, (1983). *Setor de energia elétrica : fontes e usos de recursos, série retrospectiva 1967-77*. Rio de Janeiro : Eletrobrás.
- Eletrobrás, (1973). "Eletrobrás região sudeste: Usinas e linhas de transmissão ACIMA de 66KV em Operação/construção ao estudo ou projeto" Map.
- Eletrobrás, (1973). "Eletrobrás região sul: Usinas e linhas de transmissão ACIMA de 66KV em Operação/construção ao estudo ou projeto" Map.
- Engel, E. R. Fischer, and A. Galetovic (2009). "On the Efficient Provision of Roads." mimeo.
- Escobar, G. and M. Torero (2005). "Measuring the Impact of Asset Complementarities: The Case of Rural Peru," *Cuadernos de Economía*, 42(May), 137-164.
- Esfahani, H.S. and M.T. Ramirez (2003), "Institutions, Infrastructure, and Economic Growth," *Journal of Development Economics*, 70, 443-477.
- Estache, Antonio, (2010). "A Survey of impact evaluations of infrastructure projects, programs, and policies." ECARES Working Paper 2010-005.
- Estache, A. B. Speciale, and D. Veredas (2005), "How much does Infrastructure Matter to Growth in Sub-Saharan Africa?" mimeo.
- Fan, S., L. Zhang, and X. Chang (2002). "Growth, Inequality and Poverty in Rural China: The Role of Public Investments," Research Report 125, International Food Policy Research Institute, Washington D.C.
- Fernald, J.G. (1998). "Roads to Prosperity? Assessing the Link between Public Capital and Productivity." *American Economic Review* 89:619-38.
- Gall, Norman. (2002). "Brazil's difficulties in making decisions; blackout in energy policy." Braudel Papers n. 31. Fernand Braudel Institute of World Economics, São Paulo.
- Glewwe, P., M. Kremer, and S. Moulin. 1998. "Textbooks and Test Scores: Evidence from a Randomized Evaluation in Kenya," Washington, DC: World Bank Development Research Group.
- Glewwe, P. and E. Miguel. (2008) "The Impact of Child Health and Nutrition on Education in Less Developed Countries," in T. P Schultz and J. Strauss eds., *The Handbook of Development Economics* Vol. 4. North-Holland Press, Amsterdam.
- Gramlich, E. (1994). "Infrastructure Investment: A Review Essay," *Journal of Economic Literature*. 32(30), 1176-1196.
- Gulliver, John S. and Roger E. Arndt. (1991). *Hydropower Engineering Handbook*. New York: McGraw-Hill.
- Greenstone, Michael and Justin Gallagher, (2008). "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program," *The Quarterly Journal of Economics*, MIT Press, vol. 123(3), pages 951-1003, August.
- Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu. (2005) "Engines of Liberation." *Review of Economic Studies*, 72, 109-133.
- Grogan, Louise and Asha Sadanand (2009) "Electrification and the Household." mimeo.

- Hahn, Jinyong, and Jerry Hausman. (2002) "A new specification test for the validity of instrumental variables." *Econometrica*. V.70, n.1, 163-189.
- Hahn, Jinyong, and Jerry Hausman. (2000). "A new specification test for the validity of instrumental variables." Mimeo, available at <http://econ-www.mit.edu/faculty/hausman/papers>.
- Hulten C.R., Bennathan, E. and S. Srinivasan (2005). "Infrastructure, Externalities, and Economic Development: A Study of the Indian Manufacturing Industry," mimeo, World Bank.
- Instituto de Pesquisa Econômica Aplicada (2010). Website database, available at: <http://www.ipeadata.gov.br/>
- International Monetary Fund (IMF). (2004). *International Financial Statistics*, Washington D.C.
- Jurgowski, Paul. (2005). "Omitted Variable Bias," in Kempf-Leonard, ed. *Encyclopaedia of Social Measurement*, Elsevier.
- Kammen, Daniel M. and Andrew Mills (2009). "Community-Based Electric Micro-Grids Can Contribute to Rural Development: Evidence from Kenya." *World Development*: 2008.
- Keller, Wolfgang and Carol Shue. April (2008). "Institutions, Technology, and Trade" NBER Working Paper No 13913.
- Ketlogetswe, C., T.H. Mothudi, and J. Mothibi (2006). "Effectiveness of Botswanas policy on rural electrification." *Energy Policy*, 35(2007), 1330-1337.
- Khandker, Shahidur, Hussain Samad, Nguyen Minh. (2009). "Welfare Impacts of Rural Electrification: Evidence from Vietnam." World Bank Policy Research Working Papers #5057.
- Kremer, M. (2003). "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons," *American Economic Review* 93(2) , May 2003, pp. 102-106.
- Kremer, Michael, Josh Angrist, Eric Bettinger, Erik Bloom, and Elizabeth King (2002). "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment," *American Economic Review*, 92(5), 1535-1558.
- Londono, J. (1996). "Poverty, Inequality, and Human Capital Development in Latin America 1950-2025," The World Bank, Washington D.C.
- Mehrotra, S. and E. Delamonica (2002). "Public Spending for Children: An Empirical Note," *Journal of International Development*, 14 (8), 1105-1116.
- Memoria de Minas e Eletricidade, Government of Brazil. (2000). *Banco de Imagens, 1883-1999*.
- Michaels, Guy. (2008). "The Effect of Trade on the Demand for Skill Evidence from the Interstate Highway System." *Review of Economics and Statistics* V.8, N.4, p. 683-701.
- Miguel, E. and M. Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72(1), 159-217.
- Mingat, A. and J. Tan (1996). "The Full Social Returns to Education: Estimates Based on Countries Economic Growth," Human Capital Working Paper 16131, World Bank, Washington D.C.
- Mobarak, Ahmed Mushfiq, Andrew Sunil Rajkumar, and Maureen Cropper (2011). "The Political Economy of Health Services Provision in Brazil," *Economic Development and Cultural Change*, Forthcoming.
- Parker, S., L. Rubalcava, and G Teruel. (2008). "Evaluating Conditional Schooling and Health Programs" In T.P. Schultz and J. Strauss (Eds), *Handbook of Development Economics*, vol 4. North-Holland Press, Amsterdam.
- New York Times (2005). "Brazil Weighs Costs and Benefits of Alliance With China," September 20.
- Rud, Juan Pablo, (2010). "Electricity Provision and Industrial Development: Evidence from India." Mimeo, LSE.
- Schlenker, Wolfram, M. Hanemann, A. Fisher, (2006). "The Impact of Global Warming on US Agriculture: an Econometric Analysis of Optimal Growing Conditions." *Review of Economics and Statistics*. 88(1) 113-125.

- Schultz, T. Paul (2004). "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program," *Journal of Development Economics*, 74(1), 199-250.
- Sistema de Informações Empresariais do Setor de Energia Elétrica (SIESE). (1987). *Relatório Estatístico de Linhas de Transmissão e Subestações 1970 1975 1980-86*. Eletrobrás: Rio de Janeiro.
- Sistema de Informações Geográficas Cadastrais do SIN (SINDAT) available at <http://www.ons.org.br> last accessed 6/1/2008.
- Sistema de Informações Georreferenciadas do Setor Elétrico, <http://sigel.aneel.gov.br/brasil/viewer.htm>, last accessed 6/1/2008.
- Straub, S. (2008). "Infrastructure and Development: A Critical Appraisal of the Macro Level Literature." World Bank Policy Research Working Paper 4590.
- Strauss, J. and D. Thomas. 2008. "Health over the Life Course. In T.P. Schultz and J. Strauss (Eds), *Handbook of Development Economics*, vol 4. North-Holland Press, Amsterdam.
- Summers, L. (1994). *Investing in all the People: Educating Women in Developing Countries*, The World Bank, Washington D.C.
- Taylor, M. (2005). "Electrifying Rural Guatemala: Central Policy and Rural Reality," *Environment and Planning C: Government and Policy*, 23, 173-189.
- Tennessee Valley Authority. 2010. "Hydroelectric Power." <http://www.tva.gov/power/hydro.htm>.
- Thomas, D. et al. (2003) "Iron Deficiency and the Well-Being of Older Adults: Early Results from a Randomized Nutrition Intervention," mimeo, UCLA.
- The U.S. Geological Survey. *The Hydro 1K Derivative Database*. Available at: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro.
- U.S. Energy Information Administration (2010). *International Energy Outlook 2010*, <http://www.eia.doe.gov/oiaf/ieo/electricity.html>
- The World Bank (2008). *The Welfare Impact of Rural Electrification: A Reassessment of the Costs and Benefits*. The World Bank, Washington, D.C.
- The World Bank (2005a). "Ten Things You Never Knew About the World Bank," <http://www.worldbank.org/tenthings/>, accessed November 14, 2005.
- The World Bank (2005b) *Infrastructure*, <http://www.worldbank.org/infrastructure/>, accessed November 11, 2005.
- The World Bank (2005c). *Brazil: Background Study for a National Rural Electrification Strategy: Aiming for Universal Access*, The World Bank, Washington, D.C.
- The World Bank (2005d). World Development Indicators, Washington, D.C.
- The World Bank (2005e). Energy Working Notes: May 2005, Energy and Mining Sector Board, The World Bank, Washington D.C.
- The World Bank (2004). *World Development Report 2004: Making Services Work for Poor People*, World Bank, Washington D.C.
- The World Bank (2003). *Infrastructure Action Plan*. The World Bank, Washington D.C., July.
- The World Bank (1994) *World Development Report 1994: Infrastructure*. The World Bank, D.C.
- The World Bank (1993) *World Development Report 1993: Investing in Health*. The World Bank, D.C.
- Yeaple, S. and S. Golub (2007), "International Productivity Differences, Infrastructure, and Comparative Advantage." *Review of International Economics*. 15(2) 223-242.

B Data Appendix

There is no comprehensive source for electricity data over our period of analysis across Brazil. Most of the network was privatized in the 1990s, and the overseeing agencies, Operador Nacional do Sistema Elétrico (ONS) and Agência Nacional de Energia Elétrica (ANEEL) were formed in the early 1990s and have little institutional memory for the period prior to their charters. In order to assemble our data set, we traveled to Brazil and met with professionals in the field at major electricity companies and government agencies in Brasília, São Paulo, Rio de Janeiro, Curitiba, Salvador, and Foz do Iguaçu (Itaipu). The meetings with local professionals were informative not only in terms of data with which they provided, but also for the broader understanding of the development of the electricity network in Brazil.

We collected data from the Ministério de Minas e Energia and ANEEL in Brasília, Eletrobrás, ONS, the Memoria de Eletricidade, and Furnas in Rio de Janeiro, Companhia Hidrelétrica de São Francisco (CHESF) in Salvador, Copel in Curitiba, and Itaipu Binacional in Foz do Iguaçu. Data on the location and year of creation of plants was assembled from a database of important power plants from Sistema de Informações Georreferenciadas do Setor Elétrico (SIGEL) and a historical study of hydroelectricity in Brazil by the Memoria de Eletricidade.

Data on the state of the network in each decade was assembled from a combination of sources. Data from the 1960s was procured primarily through feasibility studies conducted by the Canambra Engineering Consultants who did a comprehensive survey of electricity in Brazil in the 1960s, focusing on the South and South-Central Brazil. Inventories of the network as of the publication dates in 1965 and 1967 were included as part of the surveys, and maps were also included to show the placement of the network. CHESF also provided limited information about the state of the network in North Eastern Brazil from 1960 through the present.

The 1970s data was assembled from maps drawn by Furnas and Eletrobrás in 1973. Data from the 1980s is from a comprehensive inventory done by SIESE in 1987. The survey includes detailed data on both transmission lines and generation plants. Data from the 1990s is from a listing by SIGEL which is a survey of the current electricity network done by ANEEL. Data from 2000 is from both SIGEL and SINDAT, the database of the current electricity network done by ONS.

We had the data from the inventories in each period entered into Excel spreadsheets by firms in India and Bangladesh. Data on line voltage was used to ensure comparability of inventories conducted by different sources—only lines of at least 69 kilovolts were included as transmission lines—13 kilo volt lines were considered part of the distribution networks. Data from the tables were transferred into GIS maps which we then compared against maps which were drawn of the electricity network in each decade in order to insure the accuracy of our final decade-by-decade networks.

C Appendix: Modeling Electricity Networks in Brazil: An Engineering Cost Approach

In choosing locations for electrification, engineers focus on two factors: geographic factors which affect the cost of generating electricity and transporting it, and load factors related to the demand for electricity. In order to generate an instrument for electricity provision, we focus on the cost factors which determine engineers' least cost estimates. This means that we abstract from the demand side factors and focus on the geographic factors determining areas which are low cost for the generation of electricity. The cost side of the engineer's decision mainly consists of two components: the placement of the generation plants and the placement of the transmission substations. In our model, we predict the expansion of electricity networks according to the geologic factors which

engineers evaluate when determining the optimal placement of hydropower plants.

The force of falling water is the primary factor used in generating hydropower. Large generation plants in Brazil typically consist of a large dam which separates the upstream reservoir from the downstream portion of the river, the intake gates which allow limited amounts of water from the reservoir into the penstock through which water is allowed to run through the dam to the turbine. The pressure from the upstream water in the reservoir pushes the turbine which transmits current to the electric coils, and the current is then collected at high voltage levels for transmission through the high voltage transmission lines (Gulliver and Arndt, 1991). A diagram of electricity generation is shown in figure 11. Generation plants therefore require large amounts of water flow and large changes in elevation. The large levels of flow ensure that the river does not run dry over parts of the year, and the large changes in elevation allow for higher amounts of pressure on the turbines with relatively smaller amounts of water.

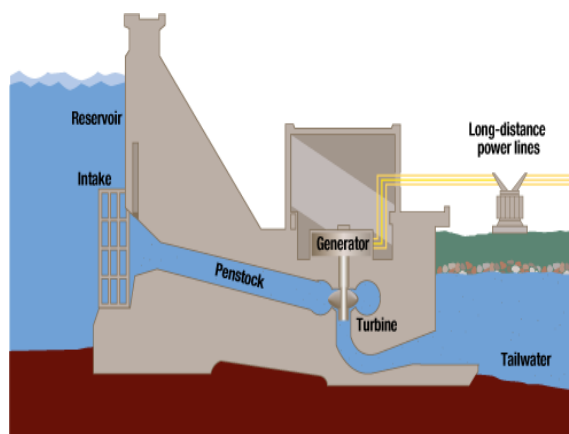


Figure 1: The generation of hydro-electric power. *Source: Tennessee Valley Authority, 2010.*

The length of the transmission lines increases the cost of the electricity. The high voltage lines are expensive, and there are electricity losses over space as the electricity is transported further from the area where it was generated.

We generate measures of the geographic factors used by engineers in determining the optimal placement of hydro power plants by using data in the United States Geological Survey’s Hydro 1K database. We overlay the map of Brazil with a grid of evenly spaced grid points each 15km apart. We also draw a buffer of 2km around each piece of each river, to proxy the actual placement of rivers. We then draw circles of radius 12km around each grid point, and extract data on the maximum slope and average slope on land, maximum and average slope in water, and the maximum and average flow accumulation of water in the rivers for each circle surrounding a grid point. This data is then attributed to the grid point, which has a unique identifier. An example of how we generate the geological observations is shown in figure 2.

We also overlay the map of actual placement of hydropower plants on our map of grid points. A grid point is assigned to have a hydropower plant if a hydropower plant is located within the grid point’s circle.

Our engineering cost model begins by selecting the locations of generation plants based on probabilities estimated in a probit equation based on the geographic factors. An indicator for whether or not a hydropower plant occurs within a grid circle is regressed on the average and maximum slope of the river within the grid circle, the log of the maximum flow accumulation (the number of raster grids which flow into each raster), a water indicator which takes the value of one if the circle has a river or stream in it, zero otherwise, and an Amazon indicator. The Amazon indicator allows for the fact that building hydropower plants in the Amazon is costly, as materials

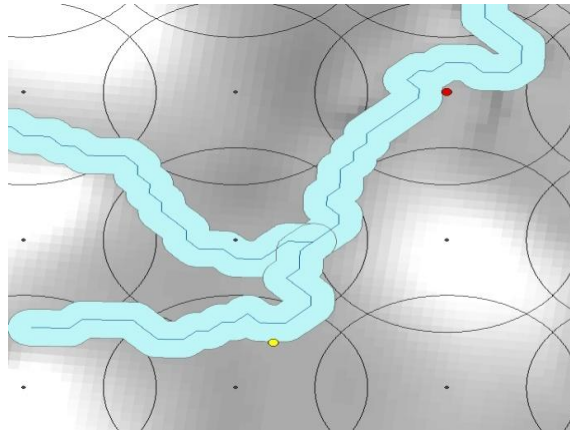


Figure 2: Construction of River Gradient Instrument *Source: Authors' calculations.*

need to be transported into the Amazon and the transmission lines to major cities would have to be quite long.

The initial placement of transmission substations is randomly selected from the remaining grid points across Brazil. Two substations are allocated to each power plant. The cost of the transmission lines is calculated based on distance and an Amazon indicator. (High materials costs are assumed for transmission stations within the Amazon as transport costs are high). Costs are calculated by the length of each line—longer lines are more expensive. The program randomly chooses alternative points and calculates the cost of the transmission lines through those points. If the line cost for the alternative points is lower, the alternative points are retained. This process is repeated until a lowest cost allocation is achieved for each decade (30,000 iterations).

Distribution networks surrounding each transmission substation and power plant are generated by selecting all points within a 50 km radius of a transmission substation as the local distribution area of electricity.

The process is repeated for subsequent decades. In each decade, power plants are allocated to the highest probability points for generation from the probit equation which have not yet received electricity. Initial locations for transmission substations are randomly chosen from the points across Brazil which have not yet received substations or power plants in earlier decades. Alternative points are proposed for the new substations. Substations and plants from previous decades are not altered. The program is again iterated until the lowest cost equilibrium is reached.

Grid points assigned to receive electricity in a given decade are given a value of 1, while those which are not projected to receive electricity in that decade are given a value of 0. The vector of projected indicator values for each grid point is then collapsed to the county level, and average predicted electrification within each county in each decade is used as our instrument for electricity provision in the instrumental variable regressions.

For robustness, we generate a similar model using geographic data from the US and the placement of hydropower plants in the US. We develop the geographic measurements in the same way as our geologic measurements in Brazil—with evenly spaced 15km apart grid points across the US overlaid with 12km circles in which we generate average and maximum values of slope and flow accumulation, and locations of hydropower plants. We then run a probit regression similar to our probit regression of hydropower plant locations in Brazil, and use these US coefficients to generate predictions for the grid points in Brazil which have the highest probability for the location of a hydropower plant based on their geographic characteristics. We then predict the placement of the hydropower plants and transmission networks in each decade using the predicted plant locations based on the coefficients from the US probit regression, and test whether using the geographic

weights from a similar probit regression run in the United States on locations of hydropower plants makes a difference in placement of electricity. We find that the results are similar across the two models.

D Tables

Table 1: Probit regression, estimating probabilities of locating a hydropower plant at a given grid point

Dependent variable: Indicator for location has a hydropower plant		
Mean of Dependent variable: 0.0113		
	Probit coefficient	Mean value in Sample
Log of Maximum Flow	0.029**	0.452
Accumulation	(0.014)	
Average Slope	0.044	0.301
in the river	(0.030)	
Maximum Slope	0.062***	0.822
in the river	(0.012)	
Amazon Indicator	-0.753***	0.452
	(0.066)	
Indicator for location	-0.030	0.462
has a river in it	(0.063)	
Observations	33342	33342

Observations are at the grid point level, and the dependent variable is an indicator for whether or not the grid point had a hydropower plant in any year between 1960 and 2000. Grid points are evenly spaced points with 15 kilometers between each grid point across Brazil. The geographic variables are calculated as explained in appendix C, and shown in figure 2. Measured values of the geographic variables are extracted from the US Geological Survey's Hydro 1K database. The coefficients shown are the probit coefficients, the mean value of each of the variables in the sample are shown to the right. Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2: Summary Statistics

Variable		Obs	Mean	Std. Dev.	Min	Max
Electricity	An indicator variable for the presence of electricity infrastructure within a county boundary or within 50km of the centroid of a county which has electricity infrastructure. (From primary data generated from inventories of generation plants and transmission substations by Brazilian Electricity companies).	8730	0.721	0.421	0	1
Percent Electrified	Percent of total households in the county which are connected to the electricity network. (From the Brazilian Census)	8730	0.602	0.328	0	1
Modeled Electricity Instrument	Percent of the evenly spaced grid points which are allocated electricity within a county within a given decade	8730	0.592	0.472	0	1
Measures of Demand						
Population density	Number of people per square kilometer	8730	62.709	292.186	0.041	10097.86
GDP per capita	Thousands of R\$ per person, in constant 2000 R\$	8728	3.765	6.057	0.084	252.13
Industrial GDP per capita	Estimated value added in the industrial sector, in thousands of R\$, in constant 2000 R\$	8730	1.196	3.708	0.002	112.18
Summary Measures of Development						
Average value of housing	Thousands of R\$ per household, calculated as the present value of monthly rent in perpetuity, or hedonically estimated values for owned housing.	8730	13.048	8.370	0.428	62.533
Human Development Index*	Index from 0 to 1, the mean of the longevity, education, and income component HDI indices.	8730	0.557	0.169	0.155	0.886
HDI Components						
HDI Longevity	Calculated according to the formula: $(LifeExpectancyatBirth - 25)/60$	8730	0.569	0.117	0.170	0.881
HDI Salaries	Calculated according to the UNDP formula: $(Log(PPPAdjustedGDPpercapita) - 2)/2.60206$ base changed in 1991, so a ratio of old to new 1991 observations was used to adjust the observations for the year 2000.	8730	0.472	0.287	0.013	1.135
HDI Education	Calculated according to the formula: $(2 * AdultLiteracyRate + GrossEnrollmentRate)/3$. Where Gross Enrollment Rate is the number of students enrolled in primary, secondary, or tertiary education, regardless of age, as a percentage of the number of people of official school age.	8730	0.515	0.154	0.075	0.877
Gross Monthly Income per capita	Gross monthly income from all sources including formal and informal employment, pensions, transfers, and rents. In thousands of R\$.	8730	0.114	0.085	0.004	0.787
Poverty	Head count ratio: Percentage of people with family income below 50% of the minimum wage	8730	60.469	24.920	3.81	99.88

All R\$ values are given in constant 2000 R\$. *The calculation of the HDI variables changed in 1991—the HDI variables have been made comparable across decades by multiplying 2000 values by the ratio of old to new 1991 values. The HDI component indices are of the general format: (observed value-minimum value)/(maximum value-minimum value).

Table 3: Summary Statistics

Variable		Obs	Mean	Std. Dev.	Min	Max
Employment						
Percent Economically Active	Number of people in the county at least 10 years of age who are employed for pay, are self-employed, or are employed without pay for at least 15 hours per week, divided by total population in the county.	8730	0.364	0.070	0.175	0.630
Percent Employed	Number of people in the county who are employed for pay, divided by total population of the county.	8730	0.347	0.063	0.132	0.613
Urban Employment	Population living in urban areas of the county reporting that they are employed for pay, divided by total population living in urban areas of the county.	8730	0.338	0.069	0.100	0.606
Rural Employment	Population living in rural areas of the county reporting that they are employed for pay, divided by total population living in rural areas of the county.	8685	0.349	0.068	0	0.741
Human Capital						
Illiteracy	The percentage of people aged 15 and above who can not, with understanding, both read and write a short, simple statement on their everyday life	8730	32.00	17.96	1.8	89.9
Less than 4 years of Education	Percent of adults over age 15 with less than four years of formal education.	8730	65.248	21.160	9.07	99.8
Average years of Education	Ratio between the sum of the number of years of schooling completed by people at least 25 years of age and the number of people in this age group.	8730	2.77	1.551	0	9.61
Human Capital per capita	The estimated difference between labor market incomes and expected income for worker with no education or experience.	6549	19.06	7.115	6.569	59.01
Population Changes						
New Migrants	Percent of total population in the county who has migrated to the county in the past five years.	4366	0.072	0.041	0	0.329
Life Expectancy	Calculated at birth assuming constant mortality rates, in years.	8730	60.098	7.809	38.4	76.921
Population Density	People per square kilometer.	8730	78.502	372.90	0.042	11732.17
Percent of population in Urban Areas	Percent of total population living in areas of the county categorized by the census as urban.	8730	0.517	0.241	0.023	1
Other Infrastructure						
Percent of Households with Running Water		8730	0.393	0.282	0	0.993
Percent of Households with Sanitation		8730	0.194	0.267	0	0.978
Percent of Households with Cars		4366	0.006	0.009	0	0.132
Geography						
landslope	Gradient over 1 square kilometer	8730	1.642	2.013	0	21.763

All R\$ values are given in constant 2000 R\$.

Table 4: Conditional Correlations between engineering model and demand side variables

		Population Density	GDP per 1000 residents	Industrial GDP per 1000 residents	Percent Industrial GDP
Engineering model prediction of lowest cost locations	Cross sectional with state dummies	0.062*** (0.02)	-2.767** (1.21)	-1.700 (1.09)	-0.010 (0.02)
	Fixed Effects	0.005 (0.02)	-1.320 (0.93)	-1.363 (1.31)	-0.029 (0.03)
Actual Electricity Infrastructure Placement	Cross sectional with state dummies	0.131*** (0.05)	0.806 (1.43)	1.632 (2.04)	0.074 (0.05)
	Fixed Effects	0.112 (0.09)	-2.831 (2.64)	-5.140 (3.93)	-0.127* (0.07)

The coefficients are the conditional correlation between the demand side variables (Population density, GDP per 1000 residents, Industrial GDP per 1000 residents and Percent of Industrial GDP), and the electricity variables (modeled predictions and actual placement from the data). All R\$ values are in constant 2000 R\$ per resident. Decade dummies are included in all specifications. The fixed effects specifications include county fixed effects.

Table 5: First Stage regressions

Dependent variable: Actual electricity availability from infrastructure inventories.

Modeled Electricity Availability	0.507*** (0.03)	0.102*** (0.03)	0.323*** (0.05)
Land slope	0.045*** (0.01)	0.006 (0.01)	
Year FE?	Y	Y	Y
State FE?	N	Y	N
County FE?	N	N	Y
r2	0.379	0.669	0.840
N	8730	8730	8730
F-Stat	232.64	16.05	34.71
p-value	0	0	0

Dependent variable is prevalence of electricity infrastructure in the county. Errors are clustered at the county level. County size weights are included in all specifications. Data is from 1960-1990 (because dependent variables in second stages are 1 period forward with the last available data in 2000).

Table 6: Housing Values

Dependent variable: Average Value of Housing

	OLS Regressions				IV Regressions using Modeled Electricity Instrument	
Lagged Electricity	4.667***	4.125***	3.262***	1.326***	9.067***	7.792***
Infrastructure	(0.81)	(0.51)	(0.37)	(0.35)	(1.60)	(1.72)
Land slope	0.357	0.022	-0.232**		-0.266	
	(0.27)	(0.26)	(0.12)		(0.32)	
Year FE?	Y	Y	Y	Y	Y	Y
Region FE?	N	Y	N	N	N	N
State FE?	N	N	Y	N	N	N
County FE?	N	N	N	Y	N	Y
N	8730	8730	8730	8730	8730	8730

Dependent Variable is average housing value in thousands. Year Dummies are included in all regressions. All regressions have county size weights and year dummies, errors are clustered by county. The average housing value in the sample is 13.048.

Table 7: Human Development Index

Dependent variable: Human Development Index

	OLS Regressions				IV Regressions using Modeled Electricity Instrument	
Lagged Electricity	0.030**	0.033***	0.027***	0.009	0.099***	0.091***
Infrastructure	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)
Land slope	0.006	0.003	-0.003**		-0.004	
	(0.00)	(0.00)	(0.00)		(0.01)	
Year FE?	Y	Y	Y	Y	Y	Y
Region FE?	N	Y	N	N	N	N
State FE?	N	N	Y	N	N	N
County FE?	N	N	N	Y	N	Y
N	8730	8730	8730	8730	8730	8730

The dependent variable is the human development index. Year dummies are included in all regressions. All regressions have county size weights, and errors are clustered by county. The average HDI value in the sample is 0.557.

Table 8: Dependent Variables: Human Development Index Components and other Poverty Measures

Dependent variable: Human Development Index Components and other Poverty Measures

	HDI: Longevity		HDI: Income		HDI: Education		Gross Income pc		Poverty	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Lagged Elect. Infrastructure	-0.004 (0.01)	-0.010 (0.04)	-0.045* (0.02)	0.274** (0.11)	0.029*** (0.01)	0.146*** (0.04)	-0.009 (0.01)	0.073** (0.03)	-1.293 (2.01)	-34.829*** (8.78)
N	8730	8730	8730	8730	8730	8730	8730	8730	8730	8730
Mean of dep var	0.569	0.569	0.472	0.472	0.515	0.515	0.114	0.114	60.469	60.469

Dependent Variables are the component indices of the HDI index and other poverty measures. Year Dummies included in all regressions. All regressions have county size weights and year dummies, errors are clustered by county.

Table 9: Dependent Variables: Measures of Employment Effects

Dependent variable: Employment

	Economically Active		Formal Employment		Formal Employment (Urban)		Formal Employment (Rural)	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Lagged Electricity Infrastructure	0.018** (0.01)	0.138*** (0.03)	0.019*** (0.01)	0.149*** (0.03)	0.016** (0.01)	0.132*** (0.03)	0.023*** (0.01)	0.155*** (0.04)
N	8730	8730	8730	8730	8730	8730	8685	8678
Mean of dep var	0.364	0.364	0.347	0.347	0.338	0.338	0.349	0.349

Dependent Variables are Employment Variables. Year Dummies included in all regressions. All regressions have county size weights and year dummies, errors are clustered by county.

Table 10: Dependent Variables: Measures of Education Effects

Dependent variable: Education

	Illiteracy		Less than 4 years Education		Years in School		Human Capital	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Lagged Electricity Infrastructure	-3.131*** (1.01)	-7.493** (3.30)	0.755 (1.28)	-12.270** (5.22)	-0.005 (0.13)	1.243** (0.52)	2.102*** (0.49)	10.117* (5.93)
N	8730	8730	8730	8730	8730	8730	6549	6549
Mean of dep var	32.00	32	65.248	65.248	2.77	2.77	19.06	19.06

Dependent Variables are Education Variables. Year Dummies included in all regressions. All regressions have county size weights and year dummies, errors are clustered by county.

Table 11: Population Changes

Dependent variables: Measures of Population Effects								
	In-Migration		Life Expectancy		Population		Urban	Per-
	Rate				Density		cent of Pop.	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Lagged Electricity	0.011	0.079	-0.742*	-1.680	6.062**	0.908	-0.009	0.124*
Infrastructure	(0.03)	(0.16)	(0.39)	(1.79)	(2.48)	(10.06)	(0.02)	(0.07)
N	4366	4366	8730	8730	8730	8730	8730	8730
Mean of dep var	0.072	0.072	60.098	60.098	78.502	78.502	0.517	0.517

Dependent Variables are population change variables. Migration data is available only for 1990 and 2000, making the panel substantially shorter. Year Dummies included in all regressions. All regressions have county size weights and year dummies, errors are clustered by county.

Table 12: a. Robustness Tests: Including other Infrastructure as Controls

	Average Housing Value		Human Development Index	
	OLS	IV	OLS	IV
Lagged Electricity	0.945*	13.708*	0.016	0.286**
Infrastructure	(0.53)	(7.14)	(0.01)	(0.14)
Lagged	0.011	-0.352	-0.102***	-0.110***
Running Water	(2.11)	(1.60)	(0.04)	(0.02)
Lagged Sanitation	-2.319	-3.378*	-0.102***	-0.124***
Access	(2.20)	(1.93)	(0.02)	(0.02)
Water	-0.061	-0.455	-0.004	-0.012
Trend	(0.34)	(0.37)	(0.01)	(0.01)
Land slope Trend	0.338***	0.056	0.004**	-0.002
	(0.12)	(0.16)	(0.00)	(0.00)
Year Dummies	Y	Y	Y	Y
N	6549	6549	6549	6549
Mean of dep var:	13.048	13.048	0.557	0.557

Dependent Variables are average housing value and HDI. Decade Dummies are included in all regressions. All regressions have county size weights and year dummies, errors are clustered by county. Water trend and land slope trend are included as proxies for the evolving availability of road infrastructure, for which we do not have available data spanning the time period of interest. Table 12b. shows the correlation between car ownership and water trend and land slope trend during the census years 1990 and 2000 in which they were collected, suggesting that these may be appropriate proxies for road availability.

Table 12: b. Land slope trend as a control for roads

Percent of Households with cars		
Land slope Trend	-0.002**	-0.004
	(0.00)	(0.00)
Water Trend	0.016***	0.052***
	(0.00)	(0.02)
Year dummies?	Y	Y
County FE?	N	Y
N	4366	4366

Table 13: Additional Robustness Tests

Dependent Variable: Average Housing Value				
	OLS	IV	IV	IV*
Percent of Houses Electrified	11.406*** (1.43)	44.172* (25.77)		
Lagged Electricity Infrastructure			8.760*** (3.03)	3.075 (1.91)
Year Dummies?	Y	Y	Y	Y
Water year dummies	N	N	Y	N
Amazon year dummies	N	N	Y	N
County FE?	Y	Y	Y	Y
N	8732	8732	8730	8730
F-Stat in first stage regression for North America instrument		2.08		40.32

*The Electricity Instrument in this regression has been constructed using data on the importance of slope and water flow in the US. The Dependent Variable is average housing value. Decade dummies included in all regressions. All regressions have county size weights and decade dummies, errors are clustered by county.

Table 14: Additional Robustness Tests

Dependent Variable: Human Development Index				
	OLS	IV	IV	IV*
Percent of Houses Electrified	0.212*** (0.02)	0.933* (0.52)		
Lagged Electricity Infrastructure			0.109** (0.04)	0.045 (0.03)
Year Dummies?	Y	Y	Y	Y
Water year dummies	N	N	Y	N
Amazon year dummies	N	N	Y	N
County FE?	Y	Y	Y	Y
N	8732	8732	8730	8730
F-Stat in first stage regression for North America instrument		2.08		40.32

*The Electricity Instrument in this regression has been constructed using data on the importance of slope and water flow in the US. The Dependent Variable is human development index. Year Dummies included in all regressions. All regressions have county size weights and decade dummies, errors are clustered by county.

E Figures

CANAMBRA ENGINEERING CONSULTANTS LIMITED												
TABLE III B-1												
INVENTORY OF TRANSMISSION LINES IN SERVICE - JUNE 1968												
PARANÁ												
FROM BUS	TO BUS	BASE VOLTS kV	LENGTH km	NO. OF CTS	CONDUCTOR		IMPEDANCE PER UNIT 100 MW BASE		CHARGING kVAR AT BASE VOLTS	REMARKS		
					SIZE	TYPE	R + jX	JK				
Chaminc	Capatema	69	53.8	2	1/0	AWG	6/1	ACSR	0.7871	0.6574	0.80	double circuit structures
Ponta Grossa	Itaiti	69	73.0	1	286.8	MCM	26/7	ACSR	0.3863	0.7524	1.15	
Itaiti	Salto Grande do Iguaçu	69	96.0	1	286.8	MCM	26/7	ACSR	0.4819	0.8902	1.52	
Itaiti	Guarapuava	69	93.7	1	286.8	MCM	26/7	ACSR	0.4714	0.9686	1.48	
Salto Grande do Iguaçu	União da Vitória	69	26.5	1	2/0	AWG	6/1	ACSR	0.3102	0.3281	0.41	operating at 33 kV
Harmonia	Mauá	69	39.8	2	4/0	AWG	6/1	ACSR copper	0.2627	0.1961	0.53	
Cianorte	Mourão	69	55.6	1	286.8	MCM	26/7	ACSR	0.2765	0.5682	0.87	
Mourão	Mamboré	69	32.3	1	2/0	AWG	6/1	ACSR	0.3718	0.3987	0.50	
Morretes	Marembó	69	15.3	1	397.5	MCM	26/7	ACSR	0.0517	0.1455	0.23	
Morretes	Paranaíba	69	38.0	1	397.5	MCM	26/7	ACSR	0.1218	0.3430	0.62	
Caacavel	Toledo	69	37.8	1	286.8	MCM	26/7	ACSR	0.1899	0.3891	0.60	operating at 33 kV
Mourão	Barbosa Ferraz	69	36.3	1	2/0	AWG	6/1	ACSR	0.4230	0.4494	0.56	

Figure 3: Transmission lines inventory from Canambra report 1969
Canambra, 1969

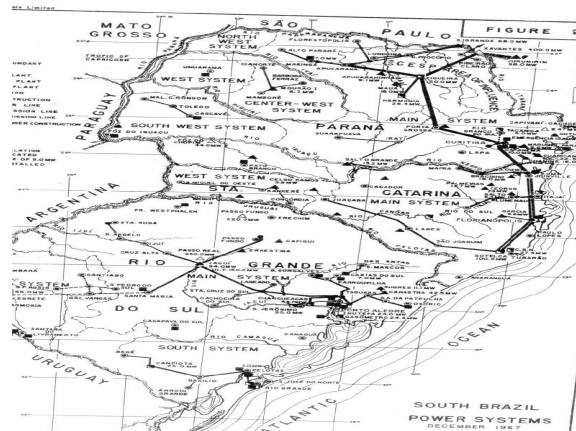


Figure 4: South Brazil Transmission as of 1967
Canambra, 1967

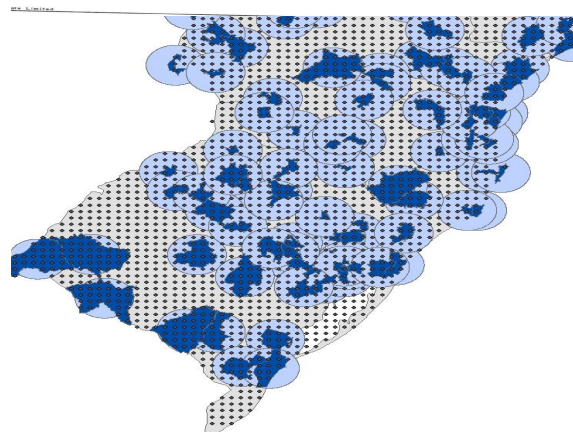


Figure 5: South Brazil Electricity Access as of 1967
Canambra, 1967

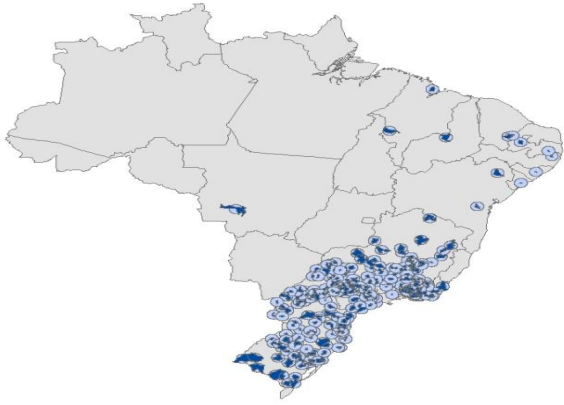


Figure 6: 1960s Transmission and Distribution

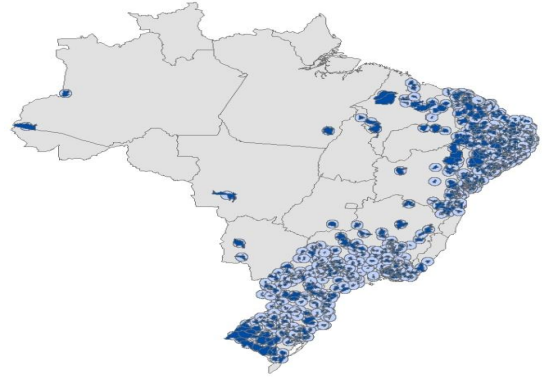


Figure 7: 1970s Transmission and Distribution

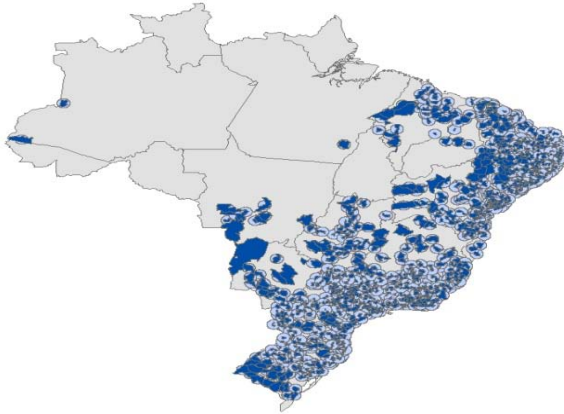


Figure 8: 1980s Transmission and Distribution

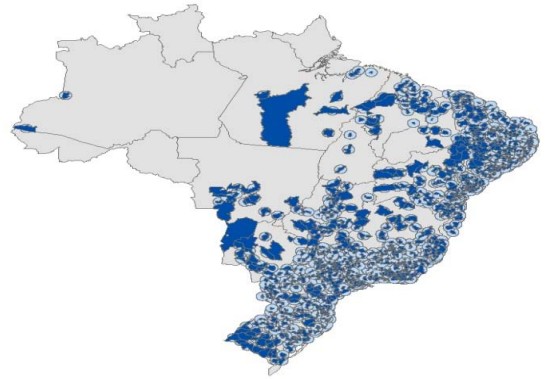


Figure 9: 1990s Transmission and Distribution

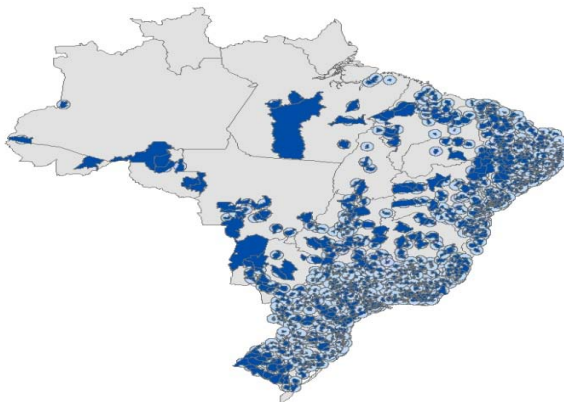


Figure 10: 2000 Transmission and Distribution

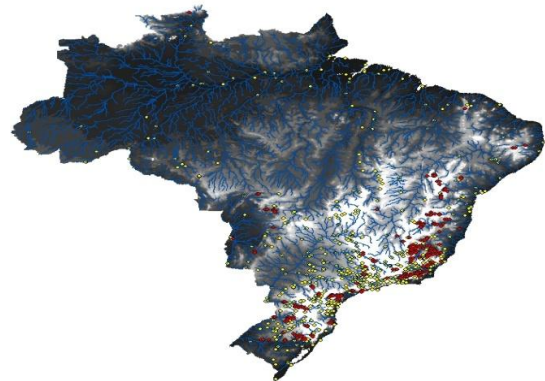


Figure 11: Predicted Locations of Hydropower plants, actual plants, rivers, and elevation.

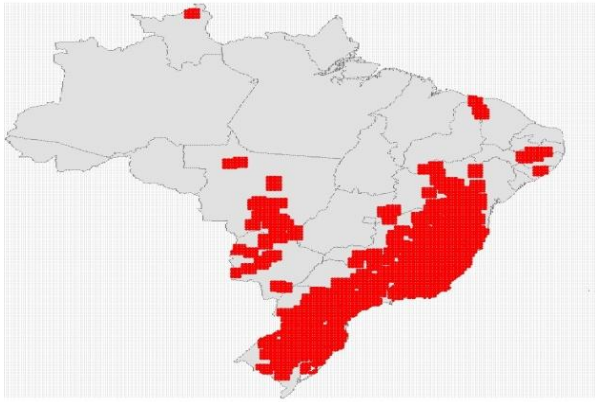


Figure 12: 1960s modeled power allocation

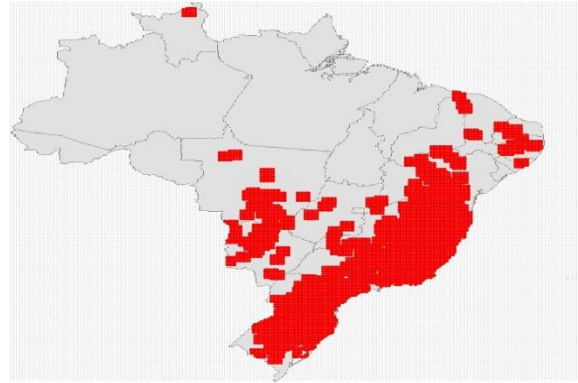


Figure 13: 1970s modeled power allocation

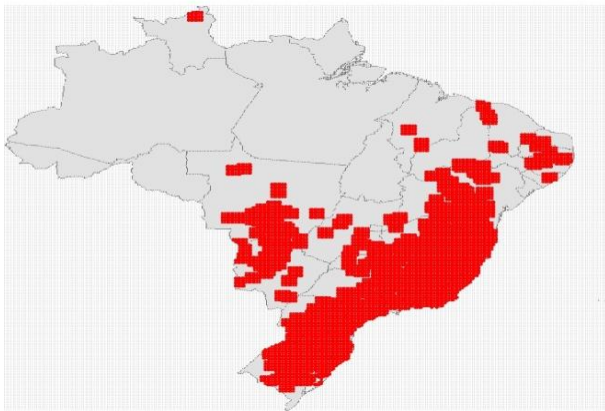


Figure 14: 1980s modeled power allocation

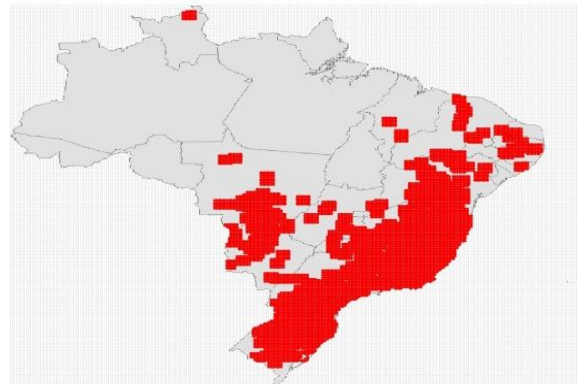


Figure 15: 1990s modeled power allocation

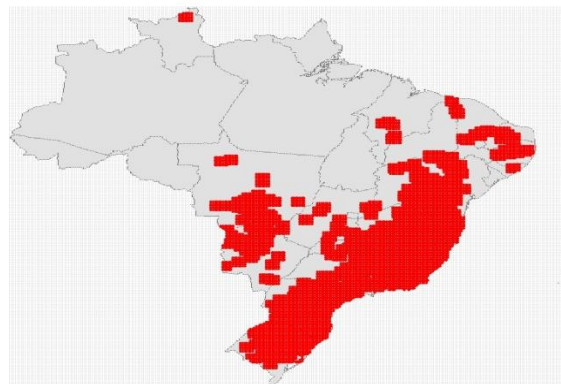


Figure 16: 2000 modeled power allocation