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

By Molly Lipscomb, A. Mushfiq Mobarak and Tania Barham

Online Appendix

Appendix 1: Data Sources

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, Eletrobras, ONS, the Memória de Eletricidade, and Furnas in Rio de Janeiro, Companhia Hidroelétrica 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 Informaes Georreferenciadas do Setor Elétrico (SIGEL) and a historical study of hydroelectricity in Brazil by the Memória 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 put together primarily through maps which were drawn by Furnas and Eletrobras 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. Examples of the tables and maps that we used in putting together the dataset are shown in Appendix Figures A1 and A2.

Appendix 2: Modeling the expansion of electricity based on cost parameters

In choosing locations for electrification, engineers focus on two factors: geographic factors that 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 that determine engineers' least cost estimates. This means that we isolate the geographic factors determining areas that are relatively low cost for the generation of electricity from the demand side factors that also affect the decisions of where to expand electricity networks.

The cost side of the engineer's decision consists of two components: the placement of the generation plants and the placement of the transmission substations. Generation plants must be located in areas suitable for the generation of large amounts of electricity: this means that they require large amounts of water flow and large changes in elevation. Transmission expenses increase quickly with the length of the transmission lines--this is because high voltage lines are extremely expensive, and there are large electricity losses in transmission across long distances.

The force of falling water is the primary factor used in generating hydropower. Large generation plants in Brazil typically consist of a large dam that 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 (intake structure) 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 that 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 A5. 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.

We begin by evenly distributing 33,342 grid points across Brazil, each located 15 kilometers on each side from the nearest grid point. These grid points are the spatial units that we use in constructing the model, and have the advantage of abstracting from political borders, and being uniform across the country. We collect data on local geographic factors by drawing buffer circles of radius 12km around each grid point, and extracting USGS data on slope and water accumulation¹ within each of the buffer circles. We also draw buffers of 2km around each piece of the river in order to proxy for average river location and width. We then attribute the average geographic statistics found in each circle to the corresponding grid point observation. The data for each buffer is then attributed to the corresponding grid point. An example of how this information is extracted from GIS is shown in figure A4. Hydropower plants are located at specific GPS coordinates in our maps, and grid points are assigned to have a hydropower plant if a hydropower station falls within 12 km of the point.

We had data on the GIS coordinates of all major hydropower plants in Brazil, which we overlaid on the gridded map. We then extracted data for each circle surrounding a grid point on whether there was a hydropower plant located within the

¹ USGS Hydro 1k Online Database.

buffer circle. Grid points which had hydropower plants built in the vicinity as of 2000 were given a value of 1 while grid points which did not have hydropower plants in the vicinity as of 2000 were given a value of 0. We had data on the counties in which transmission stations were located, so transmission stations in the data were assigned to the centroid of the county in which they were located, and any grid point in a county within 50 km of the centroid with a transmission station was classified as electrified.

The model begins by estimating geographic costs for each location of the hydropower plants. These location costs are based only on geographic factors, and the cost parameters are assumed not to vary over time. This cost vector is estimated using a probit regression for the probability of each grid point receiving electricity based 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.² The Amazon indicator allows for the fact that building hydropower plants in the Amazon is costly, as materials need to be transported into the Amazon and the transmission lines to major cities would have to be quite long. Results are shown in table 1.

In each decade, the Matlab model takes the number of hydropower plants which were built, and chooses the locations for the plants based on the vector of costs for each location. This leads to 240 being built in the 1960s, 53 built in the 1970s, 36 built in the 1980s, 25 built in the 1990s, and 24 built in 2000s. The estimated location of the hydropower plants are in the probit-estimated lowest geographic cost locations where the model has not already predicted electricity access to have reached.

Following each decade's selection of the lowest cost placement of power plants, transmission lines are chosen using a simulated annealing optimization process in Matlab. The initial placement of transmission substations is randomly selected from the grid points across Brazil that have not been assigned power plants. The initial placement is made without restriction, except that they cannot be located on points already categorized as having access to electricity. Two substations are allocated to each power plant. The cost of the transmission lines is then calculated based on distance from the power plant and an Amazon indicator. (High materials costs are assumed for transmission stations within the Amazon as transport costs are high). Costs are calculated based the length of each line--longer lines are more expensive.

In each iteration of the optimization process, 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 for that power line are retained. When lower cost transmission lines are identified for a power plant, the existing transmission line is dropped for the plant and replaced with the lower cost line. The new lower cost line is then tested against potential alternative lines that have been randomly selected from points without a power plant or substation and again replaced when a lower cost line is identified. This process is repeated until a lowest cost allocation is achieved 30,000 iterations, at which point we found in testing that the placement of the transmission lines no longer changed. Following the annealing process

² For robustness, we 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.

for a decade, the model continues to the next decade to place new power plants and transmission lines (two substations each) for those plants.

More explicitly, the modeling process takes place as follows. First 1960s power plants are placed, then 1960s transmission lines are determined using the simulated annealing method with 30,000 iterations. After the optimization process for the 1960s transmission lines has been completed with 30,000 iterations of randomly chosen points, existing transmission lines and power plants can not be moved. The power plants for the 1970s are placed in the lowest cost remaining grid points that has not yet been assigned either a transmission substation or a power plant. The transmission lines for the 1970s are optimized with the existing 1960s power plants and transmission lines remaining constant in the grid points assigned to them during the 1960s optimization. When the 1970s transmission lines have been assigned using the simulated annealing process with the points that have not yet received a power plant or transmission substation according to the model, the 1980s optimization process begins.

The process is repeated for each decade. In each decade, power plants are allocated to the highest probability points for generation from the probit equation that have not yet received electricity. New power plants and transmission stations can not be placed on existing power plants or transmission stations, but they can be placed over points in the distribution radius of older plants. Initial locations for transmission substations are randomly chosen from the points across Brazil that have not yet received substations or power plants. 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.

After the optimization process has been completed for each decade, 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.

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. Grid points are then collapsed to the county level and the average value of forecasted electrification variable is used as the instrument for electricity provision in the instrumental variable regressions.

Figure A6 shows a map of the construction of the instrumented areas of electrification in the US in the NE across the 5 decades of our study. Generation plants and transmission substations are larger circles, and other points assumed to be electrified in a given decade (because of the distribution network) are smaller circles. There is substantial variation across locations in terms of decades in which electricity access is added.

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 that 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.

Appendix 3: Variable Descriptions, Sources, and 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 Infrastructure	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: $(Life expectancy at birth - 25)/60$	8730	0.569	0.117	0.170	0.881
HDI Salaries	Calculated according to the UNDP formula: $(Log(PPP Adjusted GDP per capita) - 2)/2.60206$ 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: $(2 * Adult Literacy Rate + Gross Enrollment Rate)/3$. 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).

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\$.						

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

Appendix Figures for Online Publication

[illegible]

Figure A1: Transmission lines inventory from Canambra report 1969

Source: Canambra, 1969

Figure A2: South Brazil Transmission as of 1967

Source: Canambra, 1967

Figure A3: South Brazil Electricity Grid as of 1960s

Source: Authors' calculations

Figure A4: Construction of River Gradient Instrument

Source: Authors' calculations

Figure A5: Predicted Locations of Hydropower plants, actual plants, rivers, and elevation
Source: Authors' calculations

Figure A6

FIGURE A6.1: INFRASTRUCTURE DEVELOPMENT IN CENTRAL BRAZIL

FIGURE A6.2: INFRASTRUCTURE DEVELOPMENT IN NE BRAZIL

Source: Authors' calculations

Figure A7: Evolution of population settlements across Brazil

Source: Authors' calculations

Figure A8: Hydropower generation

Source: <http://www.tva.gov/power/hydroart.htm>

Figure A9: Modeled electricity placement in the NE 1960-2000

Source: Authors' calculations

Appendix 5: Tables

Table A1: Agricultural Effects of Local Dams

	Production (in constant 2000 R\$)			Area Harvested (hectares)		
Hydropower Dam built within 5 years	-4,712.720 (10,675.617)			-11,979.844* (7,244.856)		
5 years after Hydropower Dam built		8,207.510 (10,305.636)			8,571.454 (5,928.720)	
10 years after Hydropower Dam built			16,474.108** (7,689.646)			14,465.667 (9,404.744)
Year fixed effects	Y	Y	Y	Y	Y	Y
Observations	25,425	25,425	21,776	25,447	25,447	21,798
R-squared	0.013	0.013	0.016	0.002	0.002	0.003
Number of counties*	3,653	3,653	3,650	3,652	3,652	3,651

Robust standard errors in parentheses

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

*Data aggregated to the 1970s standardized counties rather than the 1960s data found in our main regressions, as available from the Brazilian agricultural census.