The effects of rural electrification on employment: New evidence from South Africa*

Taryn Dinkelman
Department of Economics
University of Michigan

Preliminary - please do not quote or cite without permission.

October 6, 2007

Abstract

Technology shocks biased towards home labor have ambiguous effects on market labor supply in a simple model of home production. This paper investigates the employment effects of a significant change in home production technology: the mass roll-out of domestic electrification in rural South Africa. Using two waves of aggregate Census data matched with administrative and geographic data I collected on the electricity grid, I exploit variation in project timing to estimate district fixed effects models of changes in employment rates, controlling for baseline variables. I instrument for project placement using land gradient that affects the cost of grid expansion but is unlikely to directly affect changes in employment outcomes. IV results indicate asymmetric responses by gender: female employment rates increase by 13.5 percentage points in treated areas, but there are no significant effects for men. These results do not appear to be driven by spatial spill-overs or an expansion of major female employers, suggesting that electrification is not directly generating new demand for labor. Women in their thirties experience the largest increases in employment and there is some evidence that the lack of male response is related to the types of men who remain behind in these largely migrant labor sending communities. These results contribute to a growing literature on the effects of public infrastructure in developing countries and provide new evidence on the factors influencing the extensive margin of market work for women.

*PhD candidate. Email tdinkelm@umich.edu. Electrification data and institutional background kindly provided by Jenny Barnard, Sheila Brown, Ed Bunge, Jim Stephenson, Steven Tait, Innocent Nxele and Amos Zuma. Important conversations with Matias Busso, HwaJung Choi, John DiNardo, Ben Keyes, David Lam, Claudia Martinez, Justin McCrary, Dori Posel, Samara Potter, Jeff Smith and Gary Solon substantially improved the paper. Thanks to seminar participants at the University of Michigan, Michigan State University, University of Western Ontario and the University of KwaZulu-Natal. This work was partially funded through a National Science Foundation Doctoral Dissertation Grant and research funds from the Eva Mueller New Directions in Demography fund. All errors are mine.

1 Introduction

Electricity is pervasive in all industrialized countries and largely absent in developing ones. An estimated 1.6 billion people currently do not have access to electricity [Saghir, 2005]. Eighty percent of these people live in rural sub-Saharan Africa and south Asia. African time use data indicates that significant amounts of time are consumed in collecting fuel wood and preparing food using less efficient traditional fuels. For example, assuming a 16 hour work day, women in Ghana spend approximately 3.8% of total annual work hours collecting fuel for home use, while men spend 3% of total annual hours [Charmes, 2005]. In South Africa, over three quarters of fuel wood collectors are female and women spend between one and two hours daily gathering wood for home production [Budlender et al., 2001]. For many of the poorest, lack of access to infrastructure for basic household services constrains their ability to use the one resource that they have in relative abundance: labor.

Over the next several decades, many more poor countries will expand access to electricity: the World Bank has increased power project investments in sub-Saharan Africa from $447 million in 2001 to $790 million in 2007[1], South Africa will spend $20 billion to expand generation capacity and the world's largest hydro electric power plant is planned for the Democratic Republic of Congo.[2] While these initiatives are geared towards industry, they may have the potential to affect outcomes for women and children if they include domestic electrification. Since women and children are primary fuel wood collectors and food preparers, it is routinely argued that they benefit disproportionately from electrification (see for example [Saghir, 2005] and [United Nations, 2005]. However, microeconomic evidence of these effects is sparse.

In this paper, I focus on the effects that electrification may have on rural labor markets.[3] I ask what happens within households and communities when people get access to electricity, and whether women exhibit larger responses? If so, for which women is the short run impact of this infrastructure likely to be largest? Although electrification in general has the potential to shift rural areas to a new labor market equilibrium through changes in labor demand, I argue and present evidence that household electrification more plausibly operates as a labor saving shock to home production technology which can in turn release female time into the market.

Blanket roll-out of grid infrastructure in South Africa provides an unusual opportunity to evaluate the effects of domestic electrification. In 1993, over 65% of African households were without electricity. The end of apartheid in 1994 preceded a new commitment to universal electrification by Eskom, the national utility. By 2001, over 2 million households had been newly connected to the grid. A key feature of this roll-out was its focus on low capacity

[1] World Bank, $http://ppi.worldbank.org/explore/ppi_exploreSector.aspx?sectorID = 2$

[2] http://www.irn.org/programs/congo/

[3] School enrollment is not analyzed in this paper since South African enrollment rates are very high up to the legal school-leaving age of 15.

household connections rather than industrial activities [Gaunt, 2003]. Within the context of this continuing roll-out, I measure the impact on employment rates of men and women in KwaZulu-Natal province: a rural, former homeland part of eastern South Africa.[4]

While the correlation between public infrastructure and economic activity is significantly strong and positive in most countries, the question of whether infrastructure causes this activity or follows it has always been difficult to answer.[5] Endogenous placement of infrastructure in time and space confounds causal inference. In some cases though, technological constraints on infrastructure roll-out can provide legitimate exogenous variation in allocation. In this paper, I use data on the key constraints driving connections costs to construct causal estimates of the effect of electrification on employment using an instrumental variables strategy similar to that in [Duflo and Pande, 2005].[6]

I construct a two wave panel of community aggregate Census data from 1996 and 2001 to estimate district fixed effects models of employment growth. I collected spatial data on the location of physical infrastructure in 1996, project data describing when and where electricity projects were implemented and GIS data on average land gradient within a community and matched these to the Census data. In addition to comparing employment growth across areas with and without projects (in the flavor of a difference-in-differences analysis), I use district fixed effects to account for differences in local labor market conditions over time. Controlling for a range of baseline variables including measures of proximity to local labor markets also adjusts for some differences in growth paths across communities.

Defining treatment status with project data that identifies when a community gets access to infrastructure is preferable to defining treatment based on use, since use is strongly correlated with wealth. However, even comparing treatment and control areas defined in this way does not solve the identification problem. I demonstrate that treatment status may be assigned to communities with some error (thus biasing OLS downwards) and argue that projects may be targeted at growing areas (imparting an upwards bias to OLS) or to politically important areas that are lagging behind (pulling OLS estimates down again). To address this indeterminate bias I instrument for allocation of an electricity project to a community using average community land gradient.

Gradient directly affects the average cost per connection. During the period, cost was a primary factor in prioritizing areas for electrification as Eskom was financing all roll-out and had to meet annual household connections targets. My identification assumption is that conditional on district fixed effects, baseline controls and detailed measures of proximity to

[4]Homelands were pockets of land designated for African settlement which functioned largely as labor reserves for the white economy under apartheid. In 1994, all homelands were legally reintegrated into South Africa.[Christopher, 2001]

[5]The World Bank Development Report 1994 on Infrastructure contains a brief review of the literature. [Jimenez, 1995] also discusses the difficulties in establishing causality.

[6]In that paper, the authors consider the impact of large scale dam construction on farming productivity and poverty. Their IV strategy also relies on geography.

local labor markets, land gradient is unlikely to directly affect *changes* in employment rates.

My results indicate that in areas treated with electrification projects, the proportion of households using electric lighting rises significantly and the proportion of households cooking with wood falls significantly in both OLS and IV specifications. Female employment rates are sensitive to the presence of electrification infrastructure in rural areas, but district-fixed effects results provide a downwards biased estimate. Employment rates are between 0.9 percentage points *lower* for men and 0.1 percentage points higher for women. These estimates are contaminated by measurement error in the treatment variable and unobservable differences in growth paths across treated and non-treated areas. In particular, treatment is more likely in poorer areas, controlling for other factors. Instrumental variable results suggest that female employment rises by a significant 13.5 percentage points (lower bound of 5 percentage points, upper bound of 45 percentage points) in treated areas, while the change in the male employment rate is not statistically significantly different from zero. These positive, significant changes for women are notable, since over the same period national unemployment rates are rising.

I argue that the IV strategy deals with the concern that electrification follows local growth, or development priorities. My research design is unable to definitively establish whether electricity generates an employment response by releasing labor time into the economy or by directly stimulating new demand for labor. The labor supply channel is, however, more plausible for several reasons. First, roll-out was driven by household and not firm level targets, and the capacity supplied was too small to stimulate even mid-size manufacturing or service enterprizes. Second, if firms were opening up in response to new electricity, we might expect to see spill-overs between communities so that effects would be different when comparing treated and adjacent versus non-adjacent control areas. This is not apparent in my data. Finally, the types of firms stimulated by electricity would need to be biased towards female labor in order to generate my results. There is no reason to expect this to be the case, and I show that changes in two major sources of female employment are uncorrelated with the instrument.

An issue not often addressed in studies of infrastructure relates to which group benefits initially from infrastructure. Since several papers use IV methods for identification, the question of which marginal group experiences the effect of this investment would seem key.[7] To investigate channels through which electrification affects labor market outcomes, I isolate which women are affected most by the expansion. I find that employment for women in an age group with relatively fewer child-care responsibilities is most likely to respond to treatment. There is some evidence that the lack of male employment response is partly driven by differences in the types of men remaining in these historical mine-worker migrant sending areas.

This paper contributes to a growing microeconomic literature on the effects of physical

[7] A nice example is provided by [Jalan and Ravallion, 1999], who use propensity score matching techniques to show that only households with better educated mothers exhibit short run improvements in child health in response to piped water access.

infrastructure in developing countries in two ways.[8] First, I estimate the effects of electrification using actual project data that captures changes in access directly rather than changes in use. Second, my paper places a new emphasis on employment outcomes. Poverty, health and educational outcomes are typical outcome variables in studies of infrastructure impact; far fewer studies consider labor market impacts.[9] Existing evidence on whether infrastructure affects work and wages is very limite, and varies by type of infrastructure: [Banerjee et al.,] find that Chinese wages are higher in areas transected by railroads while [Akee, 2006] estimates large positive effects of road construction on wage employment and large negative effects on agricultural employment. More broadly, my paper also contributes to what we know about factors facilitate women's entry into the labor market: a topic that has recently attracted attention in the macroeconomic-labor literature.[10]

I begin with a conceptual discussion of the ambiguous labor supply effects of a positive shock to home production technology. This model suggests for which groups effects may be larger. Sections 3 and 4 describe the data and context of South Africa's electrification. Section 5 outlines the empirical strategy and section 6 presents main results and robustness checks. Section 7 probes the channels through which electrification affects employment and section 8 concludes with a discussion of the paper's main findings.

2 Conceptual framework: effects of technology shock biased towards home production

In Becker's (1965) model of time allocation, households combine time and market goods to produce time and goods intensive commodities. The shadow price of time in the home (also the marginal cost of time) depends on household preferences for time and goods intensive commodities and on existing home production technologies. In equilibrium, an individual supplies her labor to the market up to the point where the market wage equals the shadow price of time in the home. She will not work at all if the market wage is too low. For the same market wage, women with a higher value of marginal product of time in the home are less likely to work at all and will work fewer hours. Any factor affecting home production technology may therefore alter labor supply on the extensive and intensive margins by altering the shadow price of time.

The arrival of infrastructure for domestic electricity may be characterized as a positive shock

[8] There is an established macroeconomic literature which estimates the effects of public infrastructure on total factor productivity using time series data. See for example [Aschauer, 1989], [Canning, 1998] and [Fedderke and Bogetic, 2005] for evidence from South Africa.

[9] See [Cutler and Miller, 2005] for the effects of clean water technology in the USA; [Loshkin and Yemtsov, 2005] for effects of a package of infrastructure upgrades in Georgia, Russia; [Duflo and Pande, 2005] on the effects of Indian dam construction; [Cattaneo et al., 2007] for the effects of clean cement floors in Mexico.

[10] See for example [Greenwood et al., 2005] and [Bailey and Collins, 2006].

to time productivity.[11] Labor-saving electrification increases the effective amount of labor available for producing commodities: it reduces the need to fetch wood, speeds up cooking time and allows households to shift activities from daytime into night-time. As the effective amount of labor available for production increases, the household's production possibilities frontier shifts out and shadow prices of time- and goods-intensive commodities fall asymmetrically. For a given time intensity of production, the shadow price of the time intensive commodity falls relatively more.[12]

This change induces a substitution and an income (or more accurately, an endowment) effect. The increase in effective time raises real income and so the demand for leisure and all other normal commodities. The fall in relative shadow prices causes the household to substitute towards the more time intensive commodity. However, as the household re-optimizes over consumption, the time intensity of production of both commodities may change, leading to an indeterminate change in the relative shadow prices of commodities. If the shadow price of time falls below the market wage, exporting labor to the market allows the household to produce more of both commodities with a lower time intensity and higher goods intensity. However, as [Gronau, 1986] points out, the net effect on time supplied to home and market production is ambiguous. The effect that electrification will have on labor supply to the market is therefore an empirical question.

These models do suggest differential effects of this technology shock on labor supply for different types of households. Such differences are linked to heterogenous preferences for time- and market-intensive commodities (which we can't measure), differences in the relative value of time in the home versus market and differences in initial home production technology (some of which we can measure). Increases in market labor are also more likely when the shadow price of time is initially close to the market wage. For example, women with good outside opportunities are more likely to meet the participation constraint when an improved technology reduces the shadow price of a marginal hour in home production, while women with a high value of marginal product in the home due to child care responsibilities will be less likely to enter the labor market for a given shock to home production technology. Labor supply responses are also more likely the larger the technology shock: for example, households that obtain appliances complementary to electricity will have a larger positive shock to time endowments.

It is also possible that electrification affects the technology of market production directly

[11] This is in similar vein to [Michael, 1973] who models the impact that human capital has on non-market productivity.

[12] In a series of papers in the South African journal *Reality*(1990), various authors note that Africans who do not have electricity expect it to lengthen the day for productive activities and allow them to do household activities with more ease. The various authors note that it is likely to be the wealthier households in villages that are able to initially utilize the service. In other household survey data (KwaZulu-Natal Income Dynamics Study), households that get connected to the electricity grid between 1993 and 1998 report significant increases in ownership of kettles, fridges and use of electric lighting (Author's own calculations). More recently, results from a survey conducted in the rural Eastern Cape indicate that the vast majority of households rank electricity as the second most important basic service (below water) (University of Fort Hare, 2006).

and hence the demand for labor. The type of new access rolling out to households during this period was limited to low capacity that would not sustain even medium size businesses. So, new market work that could be stimulated by electrification would be limited to craft work and other very small-scale informal jobs. Ideally, wage data over time and space could provide evidence consistent with either a net labor supply (if wages fall) or a demand (if wages rise) effect. Obtaining accurate wage data on informal small scale work is always difficult and in this context, impossible: the Census contains no wage data nor do any surveys that capture wages contain enough spatial information to be useful. Instead, I argue that if electrification was predominantly affecting labor demand, we should expect (i) no gender or age asymmetry in the employment response (ii) different employment effects when comparing treated areas to non-adjacent control areas that are less likely at risk of experiencing labor demand spill-overs and (iii) growth in the number of major employers in response to treatment. In what follows, I present some evidence that these three effects are not observable in the data.

3 Details of the electrification roll-out

Eskom, South Africa's national electricity utility, is entirely responsible for electricity generation and transmission and is the sole distributor of power for most rural areas.[13] By 1990, "most economic units were electrified [Gaunt, 2003] and since white farms had been electrified in the 1980s for political reasons, there was a good distribution of infrastructure across rural and urban areas. Access to the electricity grid had been denied to many African households. This was particularly true in the homelands. Some homeland areas were transected by high voltage lines carrying power from the coal fields to white farms and towns, but were without power themselves. Addressing the backlog in domestic connections became a development priority under the National Electrification Programme (NEP).[14]

As part of the NEP, Eskom committed to electrify 300,000 households annually from 1995 onwards. These targets were regarded as "firm and non-negotiable" [Eskom, 1996] and all connections were fully subsidized by the utility [Gaunt, 2003].[15] Between 1993 and 2003, over 10 billion Rands (about USD1.4 billion at a 2006 Rand/Dollar exchange rate) were spent on domestic electrification and over 470,000 households were grid electrified in KZN province. Once areas had been targeted for electrification, kitchens were fitted with the *Ready-board*

[13]Details in this section were collected from a combination of written sources [Gaunt, 2003] and UCT (2002)) and personal interviews with Eskom engineers and planners (Ed Bunge, Eskom Electrification Engineer, Amos Zuma, prior head of Electrification in Pietermaritzburg, Innocent Nxele, prior head of Electrification in Margate) and energy experts (Gisela Prasad, Energy Research Development Council at the University of Cape Town, Trevor Gaunt in the Department of Engineering at the University of Cape Town) conducted in Durban, Cape Town and Johannesburg between May 2006 and May 2007.

[14]The National Electrification Programme (NEP) was piloted in small scale from 1989-1991. Approximately 50,000 households were connected in the Eastern region during this time.

[15]In early years, connection fees were charged to consumers but never collected.

which contained the electric circuit board, a pre-payment meter, three plug points and one light bulb. Households received a default supply of 2.5Amps or voluntarily upgraded to a 20Amp supply for a small fee (ZAR40 or about USD6.00). The default supply was sufficient for television, radio, one or two lights and one of a toaster, bar heater or single hot plate. This upgraded supply could additionally support a fridge and one of the following combinations: an iron and double hotplate; a kettle and single bar heater; an iron and two bar heater; or a small geyser.[16] The majority of Eskom's 3 million rural customers opted to be on the 20A capacity supply [Gaunt, 2003].

The networked nature of most physical infrastructure (phones, roads, rail, electricity, piped water and waterborne sanitation) is such that not all identical consumers can be connected simultaneously and households need to be connected in some order. All annual reports and interviews with planning engineers point to the central role of costs in determining allocation of projects to places. Barnard (2006) describes how mountainous terrain complicates the extension of the grid network to rural communities in KZN. She writes: "In the case of an electrical network, ideally the best route would run along the least slope, avoid forests, wetlands and other ecologically sensitive areas, be routed near to roads and avoid households, while running near densely populated areas in order to easily supply them with electricity."

The dual pressures of internal financing and a connections target provided strong incentives for the utility to prioritize areas with lowest average cost per connection. Three main factors influence the average cost per connection:

1. Distance: The bulk of electrification cost is in laying distribution lines. How far communities are from the existing substation infrastructure and high voltage lines that are necessary for access is a key cost variable. Transportation of personnel and equipment also becomes more expensive as distance from the grid grows, and there is a reduction in reliability as lines become longer [Eskom, 1996].

2. Household density: more densely settled areas have a lower average cost per connection, as shorter cables are required to connect a given number of households.

3. Land gradient/terrain: the less of an incline the land has, the fewer hills and valleys to cross and the softer the soil, the cheaper it is to lay power lines and erect transmission poles [Eskom, 1996].

I have assembled a data set that contains measures of all three of these cost factors. Distance from the grid and household density are both likely to be correlated with economic opportunities that may directly affect changes in employment. In contrast, land gradient is much less likely to directly affect employment growth, conditional on other spatial variables and district fixed effects. Land gradient therefore forms the basis of my IV strategy which is discussed further in section 5.

[16] Department of Minerals and Energy (2004).

4 Data

My sample covers rural former homeland areas in KwaZulu-Natal. There are three reasons for this restriction. First, rural households are more likely to be using time-consuming traditional fuels than urban households. Census micro data from 1996 indicates that 2.7% of urban African households used wood for cooking whereas 63.4% of rural households did so. Second, there are potentially fewer confounders in rural areas than in urban areas. In my empirical work, I control for changes in access to other development services which is a likely source of confounding during the period. Finally, although urban electrification expanded most quicklys in the early 1990s, by the mid -1990s appropriate technology for supplying small power loads to rural areas had been developed. The five year period from 1996-2001 is therefore a relevant window for examining rural electrification effects, even though the NEP had begun prior to 1996.[17]

I combine five sources of data in what follows: aggregated data from two publicly available Census surveys, two data sets which I collected using Eskom infrastructure and administrative data and one geographic data set which I constructed using spatial mapping software (ArcGIS). This software was used to link the Census data as well as match the other data sets into Census locations. More details on this exercise are provided in the appendix.

4.1 Census data

Community level Census data from 1996 and 2001: To use the data in these two Census waves as a short panel of communities, I match geographic enumeration areas across waves in ArcGIS. I use 2001 spatial boundaries of communities as the main unit of analysis, and aggregate 1996 areas up to the 2001 boundaries, assuming a uniform distribution of people over the 1996 areas that span 2001 boundaries. A community is small and roughly equivalent to a US Census tract. The median number of households is 197 in 1996 and 265 in 2001 and 95% of communities have 750 households or fewer. My sample of communities consists of 1,992 areas in rural KZN; their boundaries are illustrated in Figures 2 and 3. The map illustrates the fragmentation that characterized the former KwaZulu; the apartheid government forcefully resettled Africans to areas deemed inhospitable for white settlement[Christopher, 2001].

Community aggregate data provides full population totals for each year for different combinations of variables. Key variables that can be constructed include the fraction of households with electricity in each year, the fraction of African adults in different age groups in different labor market states and the fraction of households living below a poverty line.[18] Since all are

[17]There is also a practical reason for focussing on this period: neither Census data nor Eskoms administrative or technical data stretch back beyond 1996. South Africa did not enumerate African homelands in the Census waves 1970-1991 and Eskom records on grid infrastructure have not been retained.

[18]The full micro data are not publicly available. The 10% micro data does not contain enough geography to match to Eskom project data, grid position, gradient or distance to roads and towns.

derived from the full population Census, I do not weight any results.

In 1996, all adults are asked: 'Does the person work?', where work includes working for pay, profit or family gain. The following activities were listed as work: formal work for a salary or wage, informal work such as making things for sale or selling things or rendering a service, work on a farm or the land, whether for a wage or as part of the household's farming activities. In 2001, adults were asked: 'Did the person do any work for pay, profit or family gain for one hour or more?', where possible responses were: yes (formal, registered, non-farming), yes (informal, unregistered, non-farming), yes (farming) and no (did not have work). Since the 2001 question includes any work for even one hour in the past week, it may be more expansive than the 1996 variable. However, with the change in employment rates as the main outcome variable, these differences are unlikely problematic as long as part time or informal work does not differentially represent a majority of new employment in steep versus flat areas.

In addition to relevant demographic and economic variables provided by the Census, I construct variables measuring the distance from each community to the nearest tarred road and the nearest small town in 1996. These distance measures capture community access to local economies.

4.2 Project data

To assign treatment status to each community, I collected administrative data from Eskom on the number of new household connections made annually by location for the period 1990 to 2007. In most locations, there is a spike in household connections in one year, indicating a concentration of project activity. I define the year of the spike as the treatment year. The main treatment variable is defined as $T = 1$ if the community had its first Eskom project between 1996 and 2001 (inclusive) and $T = 0$ if it never received an Eskom project or only had a project post-2001. Areas with projects occurring pre-1996 are excluded from the main analysis; there are 406 of these out of the total 2398 tribal areas in the sample (17%). I use these communities to conduct a false experiment in support of instrument validity. Two other treatment measures are constructed for sensitivity tests: a measure of time since treatment (T_{time}) which is $= 0$ if not treated during the period and $= 1, 2, 3, 4$ or 5 if treated between 1996 and 2001; and a treatment exposure measure that calculates the cumulative proportion of households that were connected between 1996 and 2001 ($T_{connect}$).

Measurement error in this variable arises because Eskom region boundaries do not line up with Census boundaries. I overlay Census boundaries with Eskom's spatial infrastructure data linked to project information. Treatment status is assigned to communities in the following way: for any community that lies even partially inside an Eskom project area, all of the information from that project is assigned to that community. Some communities will be assigned full treatment status when only a small percentage of households in the area were treated. In addition, not all households treated with a connection may be able to use this

electricity. To use electricity, customers had to purchase pre-paid electricity cards from local stores. Figure 1 verifies that getting new access to infrastructure did translate into increasing electricity use at the community level. The figure plots the cumulative distribution functions for the proportion of households reporting that electricity is their main source of lighting.[19] Treated areas do begin with higher rates of electrification, but coverage is low, with 80% of areas having less than 10% coverage. By 2001, the distribution for treated areas has shifted strongly to the right: only 30% of treated areas have 10% electric coverage or less, while about 70% of control areas have 10% coverage or less.[20]

4.3 Infrastructure data

I collected technical data on the location of the distribution network in 1996 from Eskom planning engineers. I combine the infrastructure database with Census spatial data and construct a measure of straight line distance (in kilometers) from each community to the closest 1996 substation. This "distance to grid" variable capture a key cost of electrification projects, since substations are necessary for stepping down electrical power to lower voltages appropriate for domestic use. Figure 2 illustrates the position of these sub-stations as triangles. Not all substations are concentrated in towns. While each town has at least one substation, there is a good distribution of substations across the province. This is the result of the politically motivated extension of power to white farmers in rural areas at the end of the 1980s.

Substations are also connected to *each other* by power lines: that electricity must travel between substations is a feature of the networked technology. This feature enables me to construct an additional instrument for allocation to treatment. I connect each substation to all other substations in the region using straight lines. Then, I create a variable indicating whether any part of a community lies on any of these connector lines or not. The hypothesis is that it is easier to expand the network by setting up transmission lines between substations than it is to build new substations.[21]

4.4 Geographic data

Using digital data on land elevation, I create a variety of land gradient measures for each community.[22] Gradient is calculated at a point for each 90 meter interval in the following

[19] Interview with Gisela Prasad, University of Cape Town Energy Research Centre: "Electric lighting was synonymous with the roll-out".

[20] Kolmogorov-Smirnov tests of differences between these empirical distribution functions reject equality of each comparison i.e. treatment vs control in the before period, treatment vs control in the after period, and before-after for each of treatment and control groups.

[21] New substations take between three and five years to build. In the mid 1990s, there was still enough capacity in the system to build lines out from existing substations; by 2006, this situation had reversed and Eskom is now in a big-build phase of expansion of substation infrastructure.

[22] These digital elevation model data are provided by the 90-meter Shuttle Radar Topography Mission (SRTM) Global Digital Elevation Model. Radar satellites capture elevation data at regularly spaced (90m) intervals.

way: it is the maximum rate of change between a point and it's eight nearest neighbors. The topography of the community is described by summarizing statistics about all gradient points within a community: mean and modal gradient, the range and variance of the gradient points. Gradient is measured in degrees from 0 (flat) to 90 degrees (vertical). This is a typical way to measure gradient.[23]

5 Empirical strategy

Let y_{jdt} be outcome y for community j and district d in time period $t = 0, 1$ with T_{jdt} indicating an Eskom electrification project in community j, district d by time period t. If treatment T_{jdt} is randomly assigned across communities, we could estimate the average treatment effect α_2 by OLS:

$$y_{jdt} = \alpha_0 + \alpha_1 t + \alpha_2 T_{jdt} + \epsilon_{jdt} \tag{1}$$

As with any infrastructure, electricity projects are unlikely to be randomly assigned. Particularly in any levels comparison, positive or negative selection on community and district level unobservables is likely. Specific terms affecting selection can be written in an error components framework:

$$\epsilon_{jdt} = \mu_j + \delta_j * t + \lambda_d * t + \nu_{jdt} \tag{2}$$

where μ_j is a community fixed effect, δ_j is a community trend term, λ_d is a district (local labor market) trend term and ν_{jdt} is remaining idiosyncratic error. To eliminate the community fixed effect, re-write equation (9) in first differences:

$$(y_{jdt+1} - y_{jdt}) = \alpha_1 + \alpha_2 \Delta T_{jdt} + \delta_j + \lambda_d + \Delta\nu_{jdt} \tag{3}$$

where $\Delta T_{jdt} = 1$ if the community had an Eskom project between t and $t + 1$.[24]

There are three reasons to suspect that even in first differenced form, OLS will not provide the correct answer to the question: what is the causal effect of electrification on changes in employment? First, positive selection on δ_j or λ_d may occur if electrification projects are allocated to communities or districts that are growing faster for unobservable reasons and $\alpha_{2,OLS}$ would be biased upwards. This is a typical concern when estimating effects of infrastructure development.

Second, negative selection on δ_j or λ_d may occur if projects are targeted to more disadvantaged areas conditional on being low cost. In the South African context, where electrification

[23] This is the base unit of the gradient variable used by Qian (2006) and[Duflo and Pande, 2005]

[24] An alternative specification would allow for X-variables to impact employment in a non-linear way by generating the logistic transformation of the outcome variable.

was driven by a socio-political compact between Eskom and the newly elected government, political concerns with disadvantaged communities could very well have directed some of the placement. Commenting on the roll-out, Gaunt (2003:91) writes that although objective criteria were identified for ranking communities, political pressures were part of the not-easily-identifiable but good reasons for selecting particular target groups. Political influence is not something I can measure in my data and must be treated as an omitted variable. If political target areas are also poorly performing regions, then $\alpha_{2,OLS}$ would be biased downwards since the control group would overcompensate for trend. [25]

Measurement error in ΔT_{jdt} presents a third practical challenge for estimating (3). Not all households in a Census community were electrified since Eskom project boundaries are typically smaller than Census community boundaries. In addition, households could still apply for access to electricity outside of the program if they could pay for the connection. Finally, since Census community boundaries cut across Eskom project boundaries, any community that even partially overlapped with an Eskom location was assigned treatment status. In the presence of this type of measurement error in a binary variable, $\alpha_{2,OLS}$ would be downwards biased.[26]

The net effect of these three sources of bias is ambiguous. I take two approaches to dealing with selection on unobservables and measurement error in the treatment variable.[27] First, I control for baseline co-variates (X_{jd0}) that should affect a community's growth path (δ_j). These variables include 1996 household density, community poverty rates, adult sex ratio (female/male), fraction of female headed households, distance to the 1996 grid, distance to the nearest road and town in 1996, fraction of adults that are White or Indian , measures of adult educational attainment in the area. Female headed households and adult sex ratios are included as additional indicators of community poverty.[28] I also include district fixed effects in this first differenced regression to take out common differences across local labor markets over time (λ_d). Since Census co-variates are limited, measures of poverty are noisy and there are no variables capturing the degree of political influence each community may have, selection on δ_j is still a concern. To overcome these issues, I instrument for program placement using mean community land gradient (Z_j).

[25] For example, [Rosenzweig and Wolpin, 1986] find evidence for negative selection of family planning project placement. This negative selection biased estimates of the impact of family planning on child health: OLS results that did not take this selection into account suggested that child health deteriorated in areas with new family planning projects. [Banerjee and Somanathan, 2007] report that gains in access to public goods in India appear to be allocated to the more politically mobilized disadvantaged groups.

[26] See [**?**] for a discussion of the effects of measurement error in a binary treatment variable.

[27] One alternative would be to use a third difference to eliminate unobservable economic growth trends. This is not possible with only two waves of data. It is also not entirely sensible in this context where the transition to democracy occurred in 1994, bringing with it new national governance and policies.

[28] [**?**] argue that both measures are good indicators of an area's poverty status in South Africa. The September 2001 Labor Force Survey indicates that self-employed and employers in rural KZN are disproportionately White or Indian, between ages 20 and 70 and with at least a grade 8 level of education. White and Indian households are also most likely to hire domestic workers.

The first stage regression for ΔT_{jdt} is therefore:

$$\Delta T_{jdt} = \pi_0 + \pi_1 Z_j + X_{jd0}\pi_2 + \gamma_d + \tau_{jdt} \tag{4}$$

The identification assumption is that, conditional on baseline community characteristics, proximity to local economic centers and grid infrastructure, land gradient of the community should not affect changes in employment outcomes independently of being assigned to an Eskom electrification project. In addition, gradient should not be correlated with measurement error in the treatment variable. This latter assumption is reasonable, given that the measurement error arises from mismatch in administrative boundaries. [29]

Geographic variables have been used in other empirical development work. Estimation of agricultural production functions often includes measures of land gradient since gradient affects soil fertility through differential run-off and erosion.[30] More recently, time-invariant topographical variables have been used to generate random variation in ability to develop infrastructure ([Duflo and Pande, 2005]and intensity of agricultural crop type ([**?**]).[31]

Gradient is a plausible candidate for an instrument in this context for several reasons. First, it is theoretically one of the three main cost drivers of electrification. Empirically, gradient is a good predictor of treatment assignment. Second, gradient predicts community level changes in use of electric lighting and wood for cooking but does not predict changes in other services related to major development projects like water and sanitation. It is therefore unlikely that gradient simply picks up 'ease of access to development projects' more generally. Third, although we might expect non-random allocation of individuals across flat and steep areas, there are reasons to believe this is less likely in the homeland context. In these areas, apartheid spatial planning forced individuals in designated areas, and tribal authorities were largely responsible for allocating land within these areas[Christopher, 2001].

It is worthwhile considering the interpretation of α_2 in the above system. The estimate of α_2 captures the local average treatment effect (LATE) of Eskom electrification projects on employment growth at a community level. Although it is typical to think about LATE's in terms of marginal effects for individuals who are affected by the instrument, these individuals aggregate to communities, which are the primary units being targeted for treatment. Understanding the characteristics of communities which have their probability of treatment manipulated by

[29] As a result of this mismatch, treatment is measured with less error in small communities and more error in larger places. For gradient not to be correlated with this measurement error, it should not differentially predict treatment in large and small places. To test this, I interact the number of households in 1996 with gradient and include this in the regression of treatment on gradient and all other controls. The interaction term is small and not significantly different from zero.

[30] For example, see [**?**].

[31] [Duflo and Pande, 2005] use the interaction of average district-level river gradient and state-level dam incidence as the exclusion restriction to model dam placement within district. They make the point that the probability of dam construction is non-linear in gradient unless the dam is for hydro-power. [**?**] uses average county gradient to instrument for the intensity of tea planting within county.

the instrument helps to put the estimate in context. For example, responses to electrification may be larger in flatter areas because these areas also have lower (fixed) commuting costs of getting to places of work.

Both OLS and IV results for encapsulate any effects of electrification on migration and migrant employment.[32] Selective migration is always a concern in studies of community program effects, as migrants may have unobservables that differ from incumbents and are correlated with treatment.[33] In this case, the migration response is of interest in itself. Suppose individuals are responding to electrification projects by moving in to treated areas faster than before. Then, employment rates may grow faster in treatment than in control areas because these migrants bring jobs with them or because they are more likely to find jobs than incumbents. In this situation, in-migration may be the major labor supply response to provision of services. To assess how much migration contributes to any measured employment effects, I re-estimate all results in a simple bounding exercise where I assume all recent in-migrants are employed and exclude them from the numerator of each employment to population ratios.

6 Results

6.1 Descriptive statistics

The spatial distribution of treated and control areas is shown in Figures 2 and 3. All communities in this sample are rural, former tribal areas of KZN province. The map highlights several important features of placement. Not all treated areas are positioned close to 1996 grid infrastructure and many areas adjacent to the grid are control areas. Being close to the original grid is neither necessary nor sufficient for subsequent electrification although we will see that it does raise the probability of treatment. Not all treated areas are clustered near towns– proximity to a town is not necessary for treatment. There is also a good distribution of treated areas across the entire province. This is helpful as it allows me to include district fixed effects which account for constant differences in growth rates across local labor markets.

Table 1 presents descriptive statistics for the sample in 1996. Means and standard deviations are provided for entire sample of 1,992 communities and for each of the treatment and control groups. The last column presents coefficients from a regression of each control variable on gradient, controlling for the 10 district fixed effects. This table provides some initial evidence on how treatment and control areas are different, and on the characteristics of communities in steeper and flatter areas.

[32] This issue confronts [**?**]) who estimate employment effects in local labor markets that are affected by coal booms and busts. In that paper, they find that a larger percentage of men lived in treated areas 5 years before the Census than in control areas, suggesting that out-migration fell as a result of changes in the coal industry.

[33] [**?**] provide a good discussion of selective migration concerns in program placement studies in developing economies.

On average, over 50% of households are female headed and the female/male adult sex ratio is well over 1. Sex ratios in these areas are skewed towards females since homelands were historically mine migrant labor reservoirs. The area is very poor: 61% of households live on less than 6,000ZAR per year.

Treatment areas are somewhat less poor than control areas and the adult sex ratio and fraction of potential employers is not significantly higher. Treated areas do have higher proportions of female headed households, high school educated men and women and are about 2.8 kilometers (1.7 miles) closer to the nearest road and town compared to control areas. There is not a great deal of evidence for targeting to wealthy areas.

Given that low average cost areas were desirable targets, it is not surprising that treated areas have on average a higher household density in 1996, are about 4.5 kilometers (2.8 miles) closer to the nearest Eskom substation in 1996, and have a 2-degree flatter average gradient than control areas. This gradient difference represents an increase in percent slope from 36% to 40%. To put "average gradient" in context: a mean gradient of 22 degrees implies a 40% slope. According to the FAO, a slope of between 20 and 25 degrees is "strongly sloping".[34] The descriptive statistics reflect the fact that these areas are extremely hilly (the middle part of the province is called the "Valley of a Thousand Hills"). Figure 4 illustrates the overlap in mean gradient distributions for treatment and control areas: it is not the case that the treatment areas are all flat while non-treated areas are all steep.

Although gradient is usually captured as an average community measure in empirical work, the average can mask variation in terrain within a community. The lower part of the table provides additional information on gradient. Treated areas have a lower modal gradient value, a somewhat lower variance and a larger range of gradient points, while median gradient is about 2 degrees lower than in control areas. These are all statistically significant differences.

Table 1 illustrates some of the difficulties with inferring causality from a comparison of outcomes across treatment and control areas: treated areas are slightly richer, but more importantly, are significantly closer to towns and roads and have a higher household density. We would expect these areas to be different in terms of growth in employment simply because of where they are situated in space. The final column illustrates the correlation between each baseline characteristic and gradient. In a completely randomized experiment, we would expect all X-variables balanced across values of the instrument; however, in a natural experiment setting this is unlikely to be the case. Looking down the column, gradient does an excellent job of balancing the community poverty rate, distance to town and road variables as well as the density and distance to the grid variables. Steeper areas are more likely to have more female headed households and a higher female to male sex ratio. This affects interpretation of the IV coefficient, because areas that are more likely to be treated because of gradient are also more likely to have less skewed sex ratios and fewer female headed households.

[34] http://www.fao.org/docrep/006/T0165E/apend.htm

Table 2 provides initial evidence that treated areas experience large positive changes in the use of electricity for lighting, large reductions in the use of wood as the main source of fuel for cooking, and little other changes in basic services that might also affect home production. Each cell presents coefficients (standard errors) from a different regression where the outcome variable is the change in proportion of households using each different service or fuel source. OLS results compare treatment and control areas, and IV results instrument for treatment using average community gradient. Results are presented without adjusting for any controls and then adjusting for all covariates and ten district fixed effects.

Average electrification rates rise by 23 percentage points more in treated areas. Reliance on wood for cooking falls by 4.2 percentage points. Treated areas do not have a significant difference in access to nearby water sources and experience a small significant increase in access to flush toilets. Instrumented coefficients are larger and significant for the variables related to electrification and largely not significant for the other basic services. In areas that are induced to be treated by virtue of having a flatter gradient, use of electric lighting increases by a large 71 percentage points, while wood use for cooking falls by 28 percentage points. This alerts us to the fact that employment responses may be larger for instrumented results, as the effects of the treatment on household fuel use are much larger than in the treatment and control area comparisons.[35]

Table 3 shows unadjusted comparisons of changes in employment rates for men and women and population growth rates in treatment and control areas. The main outcome variable is the employment to population rate of Africans aged 15 to 59 inclusive. Over the period, employment rates fall by 3.7 percentage points for men in these areas. Female employment rates remain steady on average across communities but very low, at about 7%. Comparing changes in employment rates in treated areas to the same change in control areas, the unadjusted estimate for women is not different from zero while for men it is a statistically significant -1.8%. These results are unusual; electrification is not expected to reduce employment. These negative employment changes indicate that treatment is probably occurring in places that are doing poorly over time. The second part of table 3 also suggests that differential population growth may contribute. Treated areas begin with higher populations in 1996 but also grow faster over time. Population growth in treated areas is 20 percentage points higher than in control areas over the entire period. This is roughly equivalent to a 5% growth rate per year. If people coming in to the area are predominantly unemployed men, this would contribute to the fall in male employment.

[35]The IV results for water services are also in the opposite direction to what we would expect if gradient was simply a noisy measure of wealth: areas that are slightly flatter have larger reductions in access to water sources close by, although these estimates are not significant once all controls are added. Hemson (2004) discusses why service provision worsens in some municipalities between 1996 and 2001.

6.2 OLS and IV main results

6.2.1 First stage

First stage estimates for assignment to treatment are presented in Table 4.[36] A one standard deviation increase in gradient (about 10 degrees) reduces the probability of being treated between 1996 and 2001 by 4%. As more controls are added (notably the district fixed effects), precision improves. For the main analysis, I implement inference tests that are robust to potentially weak instruments since the F-statistic is just below 10 once all controls are accounted for.[37] Using alternative definitions of treatment provides the same answer: an increase in gradient of 10 degrees reduces the probability of being treated early in the period by 12 percentage points (i.e. leads to projects happening later or not at all), and decreases the fraction of households treated by 2 percentage points. Although the strength of the first stage varies across dependent variables, gradient certainly reduces the probability of being treated between 1996 and 2001 regardless of how treatment is defined. I focus on outcomes for the treatment dummy using mean gradient as the instrument since this provides the strongest first stage.

The other two cost coefficients also have the expected signs: a one standard deviation increase in distance from the grid (about 13 kilometers) reduces the probability of treatment by 2%. Distance from the grid appears to matter more across districts, as the coefficient falls substantially with the inclusion of district fixed effects. A one standard deviation increase in household density (30 households) increases the probability of treatment by 3%.

Proximity to the nearest town and road are not significant predictors of assignment to treatment. The only other significant predictors of the treatment dummy outcome are a lower fraction of female headed households, a higher fraction of women with a high school education and a lower fraction of White and Indian adults. These coefficients are large, but should be scaled by the appropriate mean of the variable in Table 1.

The first stage provide mixed evidence on whether treated areas are selected on wealth. While areas with more female headed households (i.e. poorer) are less likely to be treated, areas with more White and Indian adults (i.e. richer) are also less likely to be treated. The community poverty rate and sex ratio variables also have positive signs in most specifications, suggesting that treatment is being assigned to poorer areas.

[36] Results from a logit model of the treatment are very similar to these linear probability model results. Using modal gradient as an instrument, modal and mean gradient, mean gradient and the standard deviation of gradient produce very similar results but smaller F-statistics in the first stage. Results available from the author upon request.

[37] Specifically, I compute heteroscedasticity-robust Anderson-Rubin confidence intervals that have the correct coverage properties in the presence of weak instruments. See Moreira and Cruz (2005), Mikusheva and Poi (2006) and Chernosokuv and Hansen (2007) for a description of Anderson-Rubin tests and a motivation for why they are robust to weak identification even under heteroscedasticity.

6.2.2 Employment: second stage

Tables 5 and 6 present OLS and IV results for African women and men aged 15 to 59 inclusive. The dependent variable is the change in sex-specific employment to population rates between 1996 and 2001. Column (1) reflects mean differences presented in Table 1: treatment areas experience a negative change in employment rates compared to control areas and this estimate is larger for men than for women. Adding controls and district fixed effects increases the coefficient on treatment somewhat. Female employment rates are growing faster in poorer places, indicated by the positive and significant coefficients on community poverty rate, sex ratio and female headed households. Distance from the 1996 grid, household density and distance to the road largely do not matter for employment growth of men or women. However, being closer to a town in 1996 predicts a smaller change in employment growth for men and women.

Differences in the coefficient on female/male sex ratio in the male and female results deserves some attention. In areas where the adult sex ratio is skewed toward women, female employment is more likely to increase: the mean sex ratio (1.4) leads to an increase in employment of 2.8 percentage points. However, male employment is not significantly different by sex ratio in the IV results. Part of the female effect may be that in areas with smaller fraction of men, women make up more of the group that responds to a given shock to home production. I explore this channel further in section 8.

IV estimates of the treatment effect are larger than OLS estimates, and significantly positive for women. Female employment increases by 13.5 percentage points in areas induced to get the treatment by gradient. Male employment increases by a much smaller 4.2 percentage points and is not significantly different from zero. Since gradient is correlated with some of the control variables as evident in Table 1, it is preferable to focus on results in columns (6) to (8) where we add successively more co-variates. To address concerns about over-optimistic inference with a possibly weak instrument, I implement heteroscedasticity-robust Anderson-Rubin tests on the second stage parameter estimate for both men and women. The standard confidence interval rejects zero and extends from 1% point to 26% points. The AR test for women can strongly reject zero and the confidence interval is wider, between 5 and 40 percentage points. The male test cannot reject zero. In a community with the median number of adult women in 1996 (N=264), a 13.5 percentage point increase in female employment translates into an increase from 21 women working to about 57 women working (35 more employed women).

In both male and female regressions in columns (6)-(8), the coefficients on variables other than treatment are remarkably consistent in sign and magnitude. For instance, the poverty variables all have the same sign and significance as in the OLS results. This is reassuring, as it implies that the instrument is not strongly correlated with observable aspects of communities that are related to wealth. Density and proximity to road, town and grid are all unimportant for employment growth compared to some of the other coefficients. The coefficient on the

proportion of potential employers in the area is negative and large, but scaled back to the mean proportion of Indian and white adults, this effect is small (below 0.001) and not significant in the IV results. The IV results are robust to the inclusion of changes in other basic households services, reassuring us that employment effects are not driven by access to other services which also possibly affect the technology of home production.

6.3 Robustness and specification checks

6.3.1 Measurement error in the treatment variable

To evaluate how much of the bias in the OLS coefficients is coming from measurement error in the treatment variable, I investigate whether OLS results are larger when restricting to (i) smaller areas that are more likely to be fully treated as part of an Eskom project; (ii) areas with at least 80% of households connected in a project between 1996 and 2001; and (iii) areas in which the change in electric lighting was at least 10%. In each exercise, the sample is restricted to those areas that are more likely to be strictly treated or untreated.

Table 7 reproduces the OLS results for females for the full sample and for successive definitions of treatment and sample limitations.[38] OLS results become large and positive in each restriction. However, even with a cleaner measure of treatment, the treatment effect is under 2 percentage points for female employment. Measurement error alone is therefore not likely to explain the entire gap between OLS and IV results. In addition, measurement error in a binary explanatory variable would bias coefficients towards zero– it could not account for the negative male OLS coefficient in the main result. OLS results are much more likely confounded by a community level effect not adequately controlled for by X's and district fixed effects.

One concern we might have is that since part of the control areas already have electricity, they may not form a good comparison group for the treated areas. I restrict to the sample of communities in which no household uses electric lighting in 1996. Table 9 part (a) presents first stage results for these communities. Gradient is still a significant predictor of treatment in these areas with a very similar coefficient. In fact, the F-statistic is larger in this restricted sample. The OLS and IV results are extremely similar to the main analysis results for women: treated areas have an increase of 12 percentage points in women working. Male OLS results are strongly negative, suggesting that there is even more selective targeting in these areas. IV results are again insignificant for this outcome.

6.3.2 Threats to validity

Any direct effect of gradient on changes in employment most likely operates through (i) changes in the value of agricultural productivity, (ii) economic shocks that are correlated with gradient

[38] Using a restriction on treatment areas alone and including the entire control group did not substantially change results.

or (iii) trend growth correlated with gradient. Addressing each of these in turn:

1. Gradient may affect crop mix through its effect on agricultural productivity (e.g. it alters water-retention properties of land and the extent of soil erosion (Lal, 1998)). Since gradient is fixed over time, it's effect on crop intensity should also be fixed; however, if profitability of these different crops changes over time (e.g. market prices change) and if farming is a large part of employment, then areas of different gradient may experience differences in farming employment growth.

   Although my study area is rural, it is not predominantly agricultural. Baseline occupational distributions for men and women in the Census are shown in Figures 4a and 4b: agriculture contributes very little to total employment. [39][40] Hence, gradient here is highly unlikely to be affecting real opportunities for work through changes in commercial farming opportunities.

2. For reasons unrelated to electrification, businesses may experience lower costs of set up in homeland areas immediately after the end of apartheid. These shocks to labor demand that are correlated with gradient would obscure the effects of electrification. There is no firm level data that could assist in testing this, but there is helpful information about the majority employers of women in the area: schools and White and Indian households. Data from the 10% micro data Census sample indicate that 75% of African women in rural KZN working as professionals or associate professionals are actually teachers. Since schools generate a demand for teachers, one 'business' we can examine the expansion of is new schools. Using two waves of the South Africa Schools Register of Needs that fall just before each of Census wave (1995 and 2000), I construct a variable measuring the change in the number of schools in each community over time.[41] Table 8 shows results from a regression of the change in the number of schools on community gradient and all other controls. There is no significant relationship between gradient and the growth in schools over time. While school placement (and hence teacher hiring) is probably related to the distribution of children in space, this distribution does not appear to be correlated with the instrument.

   As a second indirect check, I proxy for "employment opportunity" using the change in the proportion of adult population that is Indian or White adult with at least grade 8

[39] Data from other household surveys (OHS's), indicate that only 30% of households do any farming. The majority of this farming provides the sole source of food for the household (60%): very little is for market (4%) or as an extra source of food (4%) while about 30% is for leisure.

[40] This is primarily because Africans were moved [**?**] for a discussion of the collapse of farm yields in the homeland areas of South Africa under the pressure of rapid population expansion due to forced resettlements and high birth rates. Also see [**?**] chapter 6 and Aliber (2002) for an outline of how homeland areas do not generate the majority of income from agricultural produce.

[41] The Schools Register of Needs provides GPS coordinates of each school. This allows me to allocate schools to communities using the Census spatial data.

education. Column (2) of Table 8 indicates no significant relationship between gradient and the change in potential employer households over time.

3. If steep and flat areas evolve differently, the gradient IV would be invalid. In Table 9, I present the results from regressions run using two different IV sets in panel (b). In similar fashion to [Duflo and Pande, 2005], I control for gradient in the main regression and use the interaction of district level gradient and community level gradient as the instrument. Controlling for gradient takes account of differences in how flat and steep areas evolve over time. Because there are only 10 districts in the sample, the power of the first stage is markedly reduced. I find a (statistically imprecise) and slightly smaller coefficient on treatment using just this interaction as the instrument.

   As a second alternative, I generate a set of straight lines connecting substations to each other, under the assumption that it is cheaper to erect a distribution network along straight lines (ignoring topography). Communities that lie along these lines are "transected" communities. Using this transect variable alone as well as in combination with gradient, I find very similar coefficients for men and women.

   Finally, as a check on whether steep and flat areas evolve differently, I perform a false experiment for the sample of areas that have already been treated before 1996. I check to see whether gradient is correlated with changes in employment rates *after* treatment in the 1996-2001 period. A reduced form relationship between gradient and employment growth in areas where treatment status does not vary would raise concerns that gradient has a direct influence on employment growth. Table 10 provides no evidence of such a significant reduced form relationship.

6.3.3 Does electrification drive new labor demand?

Since infrastructure for electricity must expand out from existing infrastructure, control areas adjacent to treatment areas may also be affected by the treatment. For example, if workers can travel to work from control to treatment areas, then an electrification project that generates new demand for labor would have positive spill-over effects on neighboring untreated areas. This would dampen estimates of employment growth in response to electrification. Alternatively, if employable individuals move households across space towards treated areas to find jobs, electrification projects may increase the employment growth gap between treated and control spaces. Either type of spill-over could manifest in both OLS and IV estimates, and we should be able to see this if effect sizes differ depending on the set of control areas used for comparison.

I re-estimate OLS and IV regressions by restricting control areas that are at least one or five kilometers away from an area treated prior to 1996. Table 11 shows these results for each sample restriction. OLS results are never significantly different from zero while IV coefficients are large, positive and close to the main IV estimate: a coefficient of 0.114 could not be

rejected in the full sample. That areas more than five kilometers away from a prior treatment area exhibit roughly the equivalent employment response suggests that positive or negative spill-overs do not drive the result.

7 Channels

7.1 Migration

Any migration effects of electrification are captured in both OLS and IV coefficients. Recall from Table 3 that population growth in treated areas was significantly higher than in control areas over the period. The relevant question is then: how much of the employment effect is working through employed/employable migrants moving to these areas? To address this question, I bound the migrant effects on employment by redefining the dependent variable in the following way: I remove the total number of recent in-migrants from the numerator of each year's employment to population rate. This variable captures the lower bound changes in employment rates for incumbents only.

Table 12 provides results for men and women. For women, the OLS and IV results are remarkably similar across the full definition of employment and the migrant-excluded definition. A Hausman test on the treatment coefficient across each specification of female employment cannot reject that they are the same. The AR confidence intervals are tighter for this definition of employment outcomes. For men, the bounding procedure reduces the size of the OLS coefficients and increases the IV point estimate somewhat. Male employment is 8.4 percentage points higher in treated regions compared to non-treated regions but still not statistically significantly different from zero. Once again, the AR test rejects a zero effect for women (now with a lower bound of 5 percentage points), but not for men. While migration may occur in response to electrification or may have been a pre-existing trend, these bounding results suggest that differential migration cannot explain the entire female employment effect.

7.2 Age effects for women

With other constraints on home production, the technology shock may be less effective at shifting labor out of the home. To investigate for which age groups these employment effects are largest, I redefine the outcome variable to be the change in employment to population rates for five year age-groups. Table 13 provides OLS and IV coefficients on the treatment dummy for each of 9 five-year age cohorts. Each column presents the results from a separate regression.[42]

None of the OLS results indicate any response to treatment. IV results are larger and positive for each age group but significant only for women in their thirties and late forties.

[42]Results for men are not shown as the treatment coefficient was never significant for any cohort.

Employment grows by 3.9 percentage points for women between the ages of 30 and 34, by 3 percentage points for the 35 to 39 year old group and by a smaller but still statistically significant 1.9 percentage points for the older age group. A distinguishing feature of women in their thirties is that their youngest children are more likely to be of school going age than the youngest children of women in their twenties. Hence, this group is less constrained by another time-intensive household duty (child-care).[43]

7.3 Examining male employment results

Gender differences in the IV results are consistent with a model in which women are primarily responsible for home production and so more likely to be the ones to respond to home production technology shocks than men. In the context of rural South Africa, there is an additional reason to expect differential effects by gender. One legacy of the apartheid migrant labor system is continued high concentrations of women in these areas. With many men permanently away from these rural areas, it is useful to consider whether any male response to domestic electrification is possible if the types of men who remain have poor labor market prospects. In these areas, women may have better market options than men and so may be even more likely to respond to a reduction in the time demands of home production.

One easy-to-interpret measure of gender imbalance in a community is the fraction of adult men in the total adult population for that community. The mean fraction of men in the sample is 0.4, with a lower bound of 0.25 and an upper bound of 0.65 in 1996. Areas where the proportion of adult men is closer to 0.4 are more representative of a typical community outside the homelands than areas with very low fractions of men.[44] I use this variable measured in 1996 interacted with treatment and with gradient to look for differential treatment effects across areas with a higher fraction of men.

Table 14 presents these results. The IV results show both female *and* male employment rates are higher in treated areas when there is a higher percentage of men. At the mean proportion of men (0.41), the female treatment effect is 0.66 larger while the male treatment effect is 1.62 larger. These coefficients are arguably implausibly large and neither are significant. However, decomposing the treatment effect in this way is useful. Although the total treatment effect for women is positive and significant, much of this effect is being picked up in areas with a larger percentage of men. Similarly for men: although the overall treatment coefficient is smaller than for women, the interaction term is large and positive. The point estimate for men is more than double that for women, suggesting that men are much more

[43] Census micro data indicate that the proportion of women in their thirties with a youngest child of school going age is substantially higher than for women in their twenties. In a rough test of this channel, I interacted the treatment variable with the ratio of children under age 6 to women adult women, and find a negative but insignificant coefficient on the instrumented interaction term. Results are not precisely estimated, since dependency rates are unlikely to vary widely across communities.

[44] In 1996, the proportion of adult men in the entire KZN province is 0.46.

responsive to treatment in areas with more representative sex distributions or in other words, less responsive to treatment in areas with many more women than men.

Why men might be more responsive to treatment in areas with a higher percentage of men is linked to the type of men who are in these communities. The Census does not contain very detailed information on individuals, but we can examine the education and age distribution by the percent male variable. In results that are not shown, I find that areas with a lower fraction of men have a significantly lower fraction of both educated men (and women) in 1996 and in 2001. These areas also have significantly more men in the very young (15-19) and old (over 55) age brackets, who are likely to be less responsive to any supply-side intervention. There are large differences in the proportion of men in prime working age categories across low and high percent male communities: high percent male areas have a much higher proportion of men in their twenties, thirties and forties. In contrast, differences in the fraction of women in each age group are much less pronounced for different values of percent male– females of all ages are more likely to be found in each community than men of all ages. This is a peculiarity of the history of South African homeland areas which cautions against a quick extrapolation of my results to other contexts where sex ratios and age distributions are more even.

8 Discussion

Expansion of public infrastructure for water and basic energy services is often argued to ease the burden of necessary home production and release women's time in particular for more market work. However, very little evidence for this effect is available, partly because of the difficulty of empirically separating infrastructure expansion from general increases in economic growth or from politically motivated targeting. The ideal experiment of randomizing the allocation of infrastructure to households is unlikely to be available, given the networked nature of much infrastructure. Instead, a more feasible experiment would contain some natural variation in access to infrastructure unrelated to increases in economic growth or political influence. With this design in hand, there is more hope of estimating the effects of providing households with access to improved home production technology.

This paper provides new evidence on the effects of rural household electrification on employment using such a research design. One of the contributions of the paper is to use several new data sources in combination with spatially matched Census data to identify areas that experience infrastructure expansion. Using variation in the timing and location of electrification projects in KwaZulu-Natal and instrumenting for treatment to deal with measurement error and endogenous placement of infrastructure, I estimate large increases in female employment and no significant increases in male employment. Large female responses are evident for women aged 30-49: ages at which child care responsibilities are fewer. Treated areas also experience large and significant increases in electric lighting and reductions in the use of wood as main

source of cooking fuel. The plausibility of using gradient as an IV for treatment is supported by a series of checks: similar results are found using the interaction of community gradient and the position of a community in relation to straight lines connecting key grid points and there is no reduced form relationship between gradient and outcomes for the set of areas treated before 1996.

During this period, the national unemployment rate rises to over 30%. More jobs are indeed being created, but more people are also entering the labor market. Casale and Posel (2004) estimate that just over 2 million new jobs were created in the period 1995-2003, and that most of the female jobs are in the informal economy and domestic worker sector. In addition, school construction over this period does creating new opportunities for teachers in the public sector. Within the limits of my data, I cannot definitively rule out that electrification directly created any of these new jobs. However, I provide three pieces of evidence that weigh against a labor demand explanation for the employment effects I find. First, the South African roll-out supplied small, non-commercial amounts of electricity that were not suitable even for informal businesses. Second, there is no evidence of any spill-over effects across space: estimated treatment effects are not larger (or smaller) when comparing treated areas to non-adjacent control areas. Third, there is no correlation between gradient and the key sources of female labor demand that I can measure: schools and domestic worker employer households.

A caveat about external validity of these results is important. First, roll-out in my sample occurred in very rural parts of the country where reliance on time-consuming fuel wood was high. This context is possibly still relevant for other rural parts of Africa, but less so for more urban parts of South Africa where the transition to more modern fuels is already well under way. Second, since homeland areas were labor-sending regions under *apartheid*, historical patterns of migration have contributed to skewed sex ratios in these communities. There is some evidence that male-female differences in estimated treatment effects are specific to this context: in areas with a higher fraction of men, the interaction term on treatment is very large and positive for men, although not statistically different from zero. Low percent male communities tend to obscure treatment effects for men.

It is also worth bearing in mind that although the IV results are large, they translate into relatively small changes in total number of jobs, simply because the unit of analysis in my sample is a small community. In a community with the median number of adult women in 1996 (N=264), a 13.5 percentage point increase in female employment translates into an increase from 21 women working to about 57 women working (35 more employed women). This research does indicate a role for rural electrification to play in changing the nature of rural labor markets. More broadly, my results bring a new emphasis to the direct employment effects of public infrastructure.

References

[Africa,] Africa, S. S. *Census Mapping Manual 2000*.

[Akee, 2006] Akee, R. (2006). Road to palau. *IZA Working Paper*.

[Aschauer, 1989] Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics*, (23):177–200.

[Bailey and Collins, 2006] Bailey, M. and Collins, W. (2006). The economic and demographic effects of household electrification,1920 to 1960. Manuscript.

[Banerjee et al.,] Banerjee, A., Duflo, E., and Qian, N. Railroad to success: The effect of infrastructure on economic growth. 2007.

[Banerjee and Somanathan, 2007] Banerjee, A. and Somanathan, R. (2007). The political economy of public goods: some evidence from india. *Journal of Development Economics*, 82.

[Budlender et al., 2001] Budlender, D., Chobokoane, N., and Mpetsheni, Y. (2001). A survey of time use: How south africans spend their time. Technical report, Statistics South Africa.

[Canning, 1998] Canning, D. (1998). A database of world infrastructure stocks. *World Bank Policy Research Working paper*, (1929).

[Cattaneo et al., 2007] Cattaneo, M., Galiani, S., Gertler, P., Martinez, S., and Titiunik, R. (2007). Housing, health and happiness. Technical Report 4214, World Bank Policy Research Working Paper.

[Charmes, 2005] Charmes, J. (2005). *Gender, Time Use and Poverty in Sub Saharan Africa*, chapter 3. The World Bank.

[Christopher, 2001] Christopher (2001). *The Geography of Apartheid*. check.

[Cutler and Miller, 2005] Cutler, G. and Miller, D. (2005). The role of public health improvements in health advances: the twentieth-century united statess. *Demography*, 42(1):1–22.

[Duflo and Pande, 2005] Duflo, E. and Pande, R. (2005). Dams. NBER Working PAper 11711.

[Eskom, 1996] Eskom (1996). Eskom annual report. Technical report, Eskom.

[Fedderke and Bogetic, 2005] Fedderke, J. and Bogetic, Z. (2005). Infrastructure and growth in south africa: Direct and indirect productivity impacts of nineteen infrastructure measures.

[Gaunt, 2003] Gaunt, T. (2003). *Electrification technology and processes to meet economic and social objectives in Southern Africa*. Department of electrical engineering, University of Cape Town.

[Greenwood et al., 2005] Greenwood, J., Seshadri, A., and Yorukoglu, M. (2005). Engines of liberation. *Review of Economic Studies*, 72:109–133.

[Gronau, 1986] Gronau, R. (1986). *Handbook of Labor Economics*, volume I, chapter Home Production: A Survey. Elsevier Science Publishers.

[Jalan and Ravallion, 1999] Jalan, J. and Ravallion, M. (1999). Does piped water reduce diarrhea for children in rural india? World Bank Policy Research Working Paper No. 2664, World Bank.

[Jimenez, 1995] Jimenez, E. (1995). *Handbook of Development Economics*, volume 3B, chapter 43: Human and physical infrastructure: public investment and pricing policies in developing countries, pages 2774–2836. Elsevier Science.

[Loshkin and Yemtsov, 2005] Loshkin, M. and Yemtsov, R. (2005). Has rural infrastructure rehabilitation in georgia helped the poor? *World Bank Economic Review*, 19(2):311–333.

[Michael, 1973] Michael, R. T. (1973). Education in non-market production. *Journal of Political Economy*.

[Rosenzweig and Wolpin, 1986] Rosenzweig, M. R. and Wolpin, K. I. (1986). Evaluating the effects of optimally distributed public programs: Child health and family planning interventions. *American Economic Review*, 76(3):470–482.

[Saghir, 2005] Saghir, J. (2005). Energy and poverty: Myths, links and poverty issues. Technical report, The World Bank. Energy Working Notes No 4.

9 Data Appendix

9.1 Census community data 1996 and 2001

These data are released with proprietary software (Supertable) by StatsSA. The software allows one to extract community totals for various combinations of variables at enumeration area (in 1996) or sub-place (2001) level. The software is obtainable at http:
www.statssa.gov.za.

9.2 Census panel of communities

As in most countries, boundaries in South Africa have shifted over time.[45] There are two aspects of these boundaries that make working with the Census data challenging. First, the 1996 data is available at the Enumeration Area (EA) level, which is smaller than a US Census tract. These areas contain up to about 250 households. The 2001 Census data is not available at the EA level for confidentiality reasons - the data is only released at the Sub-Place level (SP) which is an aggregation of 2001 EA's (and more like a US Census tract). In order to create the panel, one approach is to aggregate 1996 EA's up to the 2001 SP's and conduct the analysis at this larger level of aggregation.

However, a second boundary issue makes this approach impossible. Between 1996 and 2001, some EA boundaries were re-drawn. Hence, some of the 1996 EA's span the 2001 EA boundaries. Statistics South Africa notes that EA boundaries should never cut across existing administrative boundaries, and all "social boundaries should be respected".[46] In most cases, re-demarcation involved the following real changes to 1996 EA's: "splits" that occurred when obstacles or boundaries divided the EA naturally, and "merges" that occurred between EA's that were small or that were legally, socially or naturally a geographical entity. Changes were made only when "absolutely necessary".[47]

This suggests that the 2001 EA's are more appropriate settlement areas than the 1996 EA's. Since I aggregate up to sub-place level anyway, any 1996 EA's that were merged together to make a 2001 EA do not pose a problem. Rather, it is the split EA's that may lie partially within a sub-place that could be problematic. I create the panel in the following ways, using spatial software (ArcGIS 9.2): I assign to each 2001 SP all of the 1996 EA's with which it intersects. This is a many-to-many mapping, as some SP's will contain more than 1 EA and some EA's will fall into multiple SP's. For each EA, I calculate the proportion of the EA polygon are that falls inside each SP. I use this proportion as a weight to assign some of the 1996 EA data to the 2001 SP for EA's that span 2001 boundaries. In order for this matching exercise to yield correct measures of sub-place aggregates, I must assume a uniform distribution of people over

[45] [Christopher, 2001]
[46] [Africa,]
[47] [Africa,]: pages 21, 26.

the 1996 EA. Once the panel of areas has been created, I use the matched identifiers to create Census aggregate data in 1996 and 2001.

9.3 Creating measures of land gradient

I used digital elevation model data to construct measures of average land gradient using GIS software (ArcMap 9.1). The procedure works in roughly the following manner: for each pixel on the image representing a 90m interval, there is an associated elevation (above sea level) point. The elevation data are captured digitally by a radar system that flew onboard the Space Shuttle Endeavour in February of 2000. For each pixel, the maximum rate of change is calculated between itself and its 8 adjacent neighbors. Mean gradient per community is created by averaging over these measures across all pixels falling inside each Census community. I also calculate the variance of gradient points for each community, the range and the majority of points in each area.

The source for these data is the 90-meter Shuttle Radar Topography Mission (SRTM) Global Digital Elevation Model (http://glcf.umiacs.umd.edu/data/srtm/) available at www.landcover.org.

9.4 Creating other measures of proximity

Eskoms 1996 grid network was provided to me by Steven Tait. I observe the geographic location of all power lines from the highest voltage (400kV) to the lowest voltage (33kV) in this year. I also observe the position of each sub-station, a necessary piece of infrastructure for stepping down electrical current to domestic-use voltage. I spatially merge the grid information with the Census geography to calculate straight line distances between Census centroids and the nearest electricity substation.

Census 1996 spatial data were used to generate straight line distances from each community centroid to the nearest road and town. These distances are then merged with the aggregate Census data.

9.5 Creating the treatment variable

Sheila Brown at Eskom provided me with a list documenting the number of pre-paid electricity connections per Eskom area by year from 1990 to 2007. Areas were referenced by name and village code. Eskom's planning units do not line up accurately with Census regions. to match project data to Census regions, I first mapped the project data to a physical location (using a spatial database of transformer codes that corresponded to project codes) and then matched these locations back to Census regions.

A list of Census sub-places containing these generated treatment variables will be available on my web site (soon).

Table 1: Covariates (measured at baseline) by treatment status and correlation with gradient

X-variables: 1996	Sample mean	Treatment Mean	Control Mean	Δ_{T-C}	Coeff. on gradient
Poverty rate	0.607	0.590	0.611	-0.021	0.000
	(0.194)	(0.169)	(0.199)		(0.001)
Female headed households	1.475	0.547	0.551	-0.004***	0.003
	(0.288)	(0.120)	(0.129)		(0.001)***
Adult sex ratio (f/m)	0.550	1.409	1.491	-0.082	0.005
	(0.127)	(0.249)	(0.295)		(0.001)***
Fraction Indian,	0.002	0.001	0.002	-0.002	0.000
White adults	(0.021)	(0.003)	(0.023)		0.000
Kms to road	38.307	36.065	38.854	-2.789*	0.084
	(24.599)	(24.103)	(24.695)		(0.206)
Kms to town	39.008	36.796	39.548	-2.751*	-0.001
	(18.293)	(15.319)	(18.914)		(0.213)
Fraction adult men	0.063	0.076	0.060	0.016***	-0.001
with high school	(0.046)	(0.048)	(0.045)		(0.000)***
Fraction adult women	0.068	0.085	0.064	0.021***	-0.001
with high school	(0.052)	(0.056)	(0.050)		(0.000)***
Cost drivers					
Household density	20.671	30.756	18.208	12.549***	-0.224
	(29.499)	(48.146)	(22.066)		(0.151)
Kms from the grid	19.322	15.681	20.211	-4.53***	-0.132
	(13.456)	(9.964)	(14.038)		(0.206)
Land Gradient - mean	22.261	20.326	22.734	-2.408***	
	(9.895)	(8.558)	(10.142)		
Gradient - mode	19.115	16.606	19.728	-3.122***	1.113
	(12.761)	(10.546)	(13.176)		(0.028)***
Gradient - std. dev.	10.841	10.316	10.970	-0.653**	0.334
	(3.919)	(3.824)	(3.932)		(0.013)***
Gradient - range	52.509	50.263	53.057	-2.793**	1.319
	(15.753)	(15.337)	(15.809)		(0.065)***
Gradient - median	21.341	19.153	21.875	-2.722***	1.059
	(10.588)	(8.906)	(10.896)		(0.009)***
N communities	1992	391	1601		1992

Table 1: Standard deviations in brackets below. Community sample consists of sub-places in former KwaZulu tribal areas. All means are calculated over communities, all variables measured in 1996. Treatment is 1 if the first Eskom project occurred between 1996 and 2001, otherwise 0. Excluded from the sample are areas first treated prior to 1996. Household poverty rate is proportion of African households in the sub-place earning <R6,000 per year. Sex ratio is the number of African adult females (ages 15-59) over the number of adult males (ages 15-59). Distances (to nearest main road, nearest town, nearest Eskom sub-station) are measured as straight line distances from the centroid of the sub-place to the nearest object. Household density is number of households per square kilometer. Various land gradient statistics are created in ARCMAP and provided at the sub-place level. Final column presents coefficients from regressions of each covariate on mean gradient, including district fixed effects. Robust standard errors are clustered at main place level.

Table 2: Changes in access to basic services

	OLS		IV		Mean $\bar{y}$
Outcome $\bar{y}$ is:	No controls	Controls	No controls	Controls	
Δ electric lighting	0.258	0.233	0.661	0.713	0.08
	(0.031)***	(0.031)***	(0.233)***	(0.232)***	
Δ wood for cooking	-0.049	-0.042	-0.305	-0.283	-0.03
	(0.012)***	(0.012)***	(0.197)	(0.148)*	
Δ nearby water sources	-0.03	0.009	-0.449	-0.287	0.01
	(0.028)	(0.023)	(0.251)*	(0.231)	
Δ access to flush toilets	0.003	0.01	0.042	0.095	0.09
	(0.006)	(0.005)**	(0.080)	(0.067)	

Table 2: Cells contain coefficients (robust standard errors clustered at main place level) from regressions of dependent variable on all explanatory variables. Treatment=1 if first Eskom project occurred between 1996 and 2001, otherwise =0. Excluded from the sample are areas first treated prior to 1996. Instrument is mean community gradient. Community sample consists of those sub-places in former KwaZulu tribal areas. Service access in each year is calculated as follows: (1) proportion of households with electricity as main source of lighting; (2) proportion of households using wood as main source of cooking fuel; (3) proportion of households with a water source in the house or within 200m; (4) proportion of households with a flush toilet; (5) proportion of households with a phone (landline or cellphone) in the house. Sample mean (standard deviation) of the dependent variable presented in final column. Controls include: XXXX

Table 3: Means of community level outcomes by treatment status

Outcomes	Year	All	Treatment	Control	Δ_{T-C}	p-value
Female e/pop	1996	0.070	0.085	0.066	0.020	0.00
		(0.082)	(0.072)	(0.084)		
	2001	0.069	0.081	0.065	0.016	0.00
		(0.074)	(0.065)	(0.076)		
	Δ_t	-0.001	-0.004	0.000	-0.004	0.25
Male e/pop	1996	0.137	0.162	0.131	0.032	0.00
		(0.118)	(0.115)	(0.118)		
	2001	0.100	0.111	0.097	0.013	0.02
		(0.097)	(0.089)	(0.098)		
	Δ_t	-0.037	-0.051	-0.033	-0.018	0.00
Log population	1996	6.778	6.924	6.742	0.182	0.00
		(0.975)	(1.041)	(0.955)		
	2001	6.969	7.277	6.894	0.383	0.00
		(0.830)	(0.772)	(0.827)		
	Δ_t	0.191	0.353	0.152	0.201	0.00
N		1992	391	1601		

Table 3: Standard deviations in brackets below. Community sample consists of sub-places in the former KwaZulu tribal areas. Treatment is 1 if the first Eskom project occurred between 1996 and 2001 , otherwise 0. Excluded from the sample are areas first treated prior to 1996. All variables are constructed for Africans only. Employment proportions are calculated over the proportion of African adults aged 15-59 inclusive. In-migrants are the number of people who report moving to the area sometime in the five years before each respective Census.

Table 4: Assignment to treatment first stage OLS

	Treatment indicator variable				Year treated	Fraction treated
	(1)	(2)	(3)	(4)	(5)	(6)
Gradient*10	-0.040	-0.040	-0.040	-0.040	-0.120	-0.020
	(0.020)*	(0.020)**	(0.010)***	(0.010)***	(0.050)**	(0.010)**
Distance from grid*10		-0.050	-0.020	-0.020	-0.010	-0.020
		(0.020)**	(0.020)	(0.020)	(0.060)	(0.010)
HH density*10		0.020	0.010	0.010	0.050	0.000
		(0.000)***	(0.010)**	(0.010)**	(0.020)***	(0.000)
Poverty rate		0.034	0.034	0.029	0.016	0.046
		(0.066)	(0.067)	(0.066)	(0.212)	(0.044)
Adult sex ratio (f/m)		0.350	0.134	0.124	-0.080	0.055
		(0.118)***	(0.104)	(0.104)	(0.365)	(0.075)
Proportion female headed hh's		-0.164	-0.117	-0.111	-0.344	-0.029
		(0.048)***	(0.038)***	(0.038)***	(0.416)*	(0.025)
Proportion Indian/white adults		-0.693	-0.576	-0.571	-1.789	-0.320
		(0.256)***	(0.250)**	(0.230)**	(0.168)***	(0.151)**
Distance to road*10		0.000	-0.010	-0.010	0.200	0.000
		(0.010)	(0.010)	(0.010)	(0.300)	0.000
Distance to town*10		0.020	0.010	0.010	-0.400	0.000
		(0.010)	(0.010)	(0.010)	(0.500)	(0.010)
Adult men with high school		-0.041	0.090	0.087	-0.265	0.273
		(0.451)	(0.399)	(0.387)	(1.253)	(0.221)
Adult women with high school		0.836	0.726	0.783	2.493	0.219
		(0.419)**	(0.395)*	(0.375)**	(1.280)*	(0.209)
Δ hh's with water close by				0.016	-0.265	0.020
				(0.045)	(0.159)*	(0.041)
Δ hh's with flush toilets				0.178	0.931	0.083
				(0.086)**	(0.382)**	(0.047)*
District FE	No	No	Yes	Yes	Yes	Yes
Mean of treatment variable	0.196	0.196	0.196	0.196	0.332	0.096
Obs	1992	1992	1992	1992	1992	1992
R^2	0.009	0.075	0.168	0.169	0.150	0.199
F-stat: gradient	3.610	5.400	9.060	8.870	4.920	6.000
Prob >F:	0.060	0.020	0.000	0.000	0.030	0.010

Table 4: Robust standard errors clustered at main place level: hierarchy of geography is from smallest (sub-place) to main place to largest (district). Land gradient is measured in degrees and all distances are measured in kilometers. All distances are measured from centroid of polygon to nearest object (road, town or substation). Ten district fixed effects dummies included in columns (3) to (6). Adults are aged 15 and up. Household poverty, density, adult sex ratio, proportion of female headed households, proportion of Indian and white adults, proportion of adult African men and women with at least a high school qualification are all measured in 1996. The change in the proportion of households with access to water close by (in the house or no more than 200 meters away) and with a flush toilet is the change from 1996 to 2001. The outcome variable is a dummy in columns (1)-(4) where 1 indicates the area had an Eskom project in between 1996 and 2001, otherwise 0. In column (5) the outcome measures how many years ago the project was completed: values are 1,2,3,4 and 5 for up 5 years before 2001, and 0 if no project occurred during this period. In column (6), the outcome measures the fraction of 1996 households that have been connected under Eskom electrification projects occurring between 1996 and 2001.

Table 5: Employment effects for women - OLS and IV

X variables	OLS	OLS	OLS	OLS	Reduced form	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.004	0.000	0.002	0.001		0.045	0.091	0.136	0.135
	(0.005)	(0.005)	(0.005)	(0.005)		(0.055)	(0.062)	(0.064)**	(0.062)**
Kms to grid *10		0.004	0.004	0.004	0.004		0.001	0.001	0.001
		(0.003)	(0.003)	(0.003)	(0.003)		(0.004)*	(0.000)	(0.000)
HH density *10		0.000	0.000	0.000	0.000		-0.001	-0.002	-0.002
		(0.001)	(0.001)	(0.001)	(0.001)		(0.002)	(0.001)	(0.001)
Poverty rate		0.032	0.035	0.031	0.036		0.028	0.031	0.028
		(0.011)***	(0.011)***	(0.011)***	(0.012)***		(0.013)**	(0.016)**	(0.015)*
Prop. female headed		0.036	0.039	0.034	0.041		0.008	0.022	0.019
households		(0.023)	(0.023)	(0.023)	(0.023)*		(0.032)	(0.030)	(0.030)
Adult sex ratio (F/M)		0.020	0.020	0.024	0.022		0.036	0.038	0.040
		(0.010)**	(0.010)**	(0.009)**	(0.010)**		(0.015)**	(0.013)***	(0.013)***
Prop. Indian/whites		-0.495	-0.485	-0.482	-0.491		-0.433	-0.413	-0.410
		(0.270)*	(0.269)*	(0.256)*	(0.265)*		(0.271)	(0.263)	(0.255)
Kms to road*10		0.001	0.001	0.001	0.000		0.000	0.001	0.002
		(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.002)	(0.002)
Kms to town*10		-0.004	-0.003	-0.004	-0.003		-0.006	-0.004	-0.005
		(0.002)**	(0.002)	(0.002)*	(0.002)		(0.003)**	(0.000)	(0.003)*
Men with high school		0.150	0.161	0.159	0.152		0.146	0.139	0.137
		(0.104)	(0.105)	(0.092)*	(0.104)		(0.102)	(0.101)	(0.094)
Women with high school		-0.180	-0.195	-0.153	-0.192		-0.257	-0.290	-0.257
		(0.115)	(0.116)*	(0.100)	(0.115)*		(0.120)**	(0.114)**	(0.108)**
Δ hh's with water				0.028					0.026
nearby				(0.007)***					(0.010)***
Δ hh's with flush				0.111					0.085
toilet				(0.058)*					(0.058)
Gradient*10					-0.006				
					(0.002)***				
District FE?	N	N	Y	Y	Y	N	N	Y	Y
N	1992	1992	1992	1992	1992	1992	1992	1992	1992
R^2	0.000	0.067	0.075	0.100	0.078				
Standard C.I.								[0.01-0.26]	[0.01-0.26]
AR C.I.								[0.05-0.4]	[0.05-0.4]

Table 5: Robust standard errors in parentheses, clustered at main place level. Outcome variable is change in proportion of African females aged 15-59 who are employed (2001-1996). All other variables (except treatment) are measured in 1996. Treatment is 1 if community had the first Eskom project between 1996 and 2001, otherwise 0. Standard confidence intervals are provided for IV results as well as confidence intervals from the Anderson-Rubin test. The AR test is robust to weak instruments and was implemented to be robust to heteroscedasticity.

Table 6: Employment effects for men - OLS and IV

X variables	OLS	OLS	OLS	OLS	Reduced form	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	-0.018	-0.016	-0.010	-0.011		-0.053	0.053	0.041	0.042
	(0.008)**	(0.006)**	(0.006)	(0.006)*		(0.080)	(0.080)	(0.068)	(0.068)
Kms to grid *10		0.009	0.006	0.006	0.006		0.012	0.007	0.008
		(0.004)**	(0.004)	(0.004)	(0.004)		(0.005)**	(0.005)	(0.005)
HH density *10		0.002	0.002	0.002	0.002		0.000	0.001	0.001
		(0.002)	(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Poverty rate		0.066	0.068	0.064	0.068		0.063	0.066	0.063
		(0.018)***	(0.017)***	(0.016)***	(0.017)***		(0.020)***	(0.018)***	(0.018)***
Female-head hh		0.235	0.243	0.240	0.242		0.213	0.237	0.234
		(0.031)***	(0.033)***	(0.033)***	(0.033)***		(0.041)***	(0.036)***	(0.036)***
Adult sex ratio (F/M)		0.002	-0.001	0.001	0.001		0.015	0.005	0.007
		(0.011)	(0.011)	(0.012)	(0.011)		(0.019)	(0.015)	(0.016)
Prop.Indian/whites		-0.077	-0.055	-0.052	-0.051		-0.030	-0.027	-0.024
		(0.275)	(0.270)	(0.257)	(0.269)		(0.280)	(0.273)	(0.262)
Kms to road		0.002	0.000	0.000	0.000		0.001	0.000	0.001
		(0.001)	(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Kms to town		-0.008	-0.003	-0.003	-0.003		-0.009	-0.003	-0.004
		(0.002)***	(0.003)	(0.003)	(0.003)		(0.003)***	(0.003)	(0.003)
Men with high school				-1.407	-1.431		-1.399	-1.468	-1.491
				(1.258)	(1.168)		(1.258)	(1.241)	(1.343)
Women with high school				1.405	1.009		0.615	0.712	0.997
				(1.314)	(1.258)		(1.551)	(1.451)	(1.506)
Δ hh's with water				0.278					0.273
nearby				(0.009)***					(0.009)***
Δ hh's with flush				0.826					0.723
toilet				(0.747)					(0.737)
Gradient*10					-0.002				
					(0.003)				
District FE?	N	N	Y	Y	Y	N	N	Y	Y
N	1992	1992	1992	1992	1992	1992	1992	1992	1992
R^2	0.005	0.152	0.169	0.179	0.168				
Standard C.I.								[-0.09-0.17]	[-0.09 - 0.17]
AR C.I.								[-0.05-0.25]	[-0.05 - 0.25]

Table 6: Robust standard errors in parentheses, clustered at main place level. Outcome variable is change in proportion of African females aged 15-59 who are employed (2001-1996). All other variables (except treatment) are measured in 1996. Treatment is 1 if community had the first Eskom project between 1996 and 2001, otherwise 0. Standard confidence intervals are provided for IV results as well as confidence intervals from the Anderson-Rubin test. The AR test is robust to weak instruments and was implemented to be robust to heteroscedasticity (see Moreira (2001) and Chernosukov and Hansen (2007)).

Table 7: How much does measurement error in treatment account for gap between OLS and IV?

	Treatment measure is:							
	First project btwn 96-01		+ Δ10% electricity		+ > 0.8 connection rate		Restricted to < 100 hh's	
	(1)	(2)	(3)	(4)	(5)	(6)		
Treatment coefficient for:	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Δ female employment	0.001	0.135	0.010	0.126	0.012	0.136	0.017	0.315
	(0.005)	(0.062)**	(0.007)	(0.065)**	(0.009)	(0.092)	(0.010)*	(0.177)*
N	1992	1992	1619	1619	1420	1420	625	625

Table 7: Each coefficient (standard error) is from a separate regression. Robust standard errors in parentheses, clustered at main place level. Outcome variable is change in proportion of employed African adults aged 15-59. All other variables (except treatment) are measured in 1996. Treatment= 1 if community had first Eskom project between 1996 and 2001. Columns (3) and (4) restrict the sample to all control areas and treated areas with a 10% or higher change in proportion of households using electric lighting (from the Census). Columns (5) and (6) restrict the sample to all control areas and treated areas where Eskom connected at least 80% of households between 1996 and 2001. Columns (7) and (8) restrict the entire sample to communities with fewer than 100 households.

Table 8: Are demand side variables correlated with gradient?

	Δ schools	Δ prop. Indian and White adults
Average gradient*10	0.002	0.001
	(0.013)	(0.001)
Modal gradient*100	0.033	0.008
	(0.088)	(0.008)
N	1992	1992

Table 8: Robust standard errors in brackets, clustered at main place level. Outcome variable is the change (2000-1995) in the number of schools in the community (range=0 to 5), and the change (2001-1996) in the proportion of Indian and white adults ages 20 and over with at least grade 8 education. All other variables also included.

Table 9: Robustness of result to different IV measures, treatment measures and subsamples

		First stage		Female employment		Male employment	
	N	IV coeff.	F-stat	OLS	IV	OLS	IV
Variations in:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T=dummy, IV=mean gradient							
(a) Restricting to places	477	-0.005	9.670	0.004	0.119	-0.023	-0.026
with no electricity		(0.002)***		(0.007)	(0.058)**	(0.009)**	(0.102)
T=alternative measure, IV =mean gradient							
(a) T=year treated	1992	-0.120	4.920	0.000	0.050	-0.003	0.018
		(0.050)**		(0.001)	(0.029)*	(0.002)	(0.025)
(b) T=fraction treated	1992	-0.020	6.000	0.007	0.270	-0.007	0.099
		(0.010)**		(0.008)	(0.137)**	(0.011)	(0.129)
T= dummy, IV = alternative measure							
(a) Gradient*District grad	1992	-0.001	2.800	0.001	0.123	-0.011	0.013
		(0.000)*		(0.005)	(0.056)**	(0.006)*	(0.062)
(b) Gradient	1992	-0.005	7.570	0.001	0.112	-0.011	0.036
		(0.001)***		(0.005)	(0.048)**	(0.006)*	(0.052)
Transect		0.160					
		(0.069)**					
(c) Transect	1992	0.160	5.410	0.000	0.078	-0.011	0.028
(controlling for gradient)		(0.069)**		(0.005)	(0.079)	(0.006)*	(0.083)

Table 9: Each cell is a coefficient from a different regression (OLS or IV). Robust standard errors in parentheses, clustered at main place level. Outcome variable in column (2) is dummy for treatment. Outcome variable in columns (3)-(7) is change in proportion of African adults aged 15-59 who are employed (2001-1996). All other variables (except treatment) are measured in 1996. Treatment is 1 if community had the first Eskom project between 1996 and 2001, otherwise 0. Transect= 1 if community falls on a straight line segment connecting any two substations to each other; otherwise 0. District gradient is the average of community level gradient within each community, for each of 10 districts. Year treated is 0 if never treated between 1996 and 2001, otherwise $= 1, 2, 3, 4, 5$ for up to 5 years before the 2001 Census. Fraction of households treated is the proportion of 1996 households that are connected (cumulatively) over the period 1996-2001.

Table 10: False experiment using sample treated prior to 1996

	Δ female employment	Δ male employment	Δ Log population
Gradient*10	-0.001	-0.001	-0.042
	(0.001)	(0.001)	(0.009)***
Kms to grid 1996*10	-0.005	-0.008	-0.020
	(0.008)	(0.009)	(0.010)**
Household density 1996*10	0.001	0.000	-0.007
	(0.001)	(0.001)	(0.002)***
Household poverty rate 1996	0.080	0.067	-0.495
	(0.033)**	(0.039)*	(0.488)
Prop female headed hh 1996	0.002	0.342	0.844
	(0.059)	(0.078)***	(0.703)
Adult sex ratio (F/M) 1996	0.056	0.073	0.010
	(0.026)**	(0.038)*	(0.345)
Number of Indian/white	-0.158	3.145	-8.163
	(1.776)	(1.481)**	(10.839)
Kms to road	-0.003	-0.002	-0.002
	(0.003)	(0.003)	(0.003)
Kms to town	0.001	-0.007	0.002
	(0.003)	(0.003)*	(0.005)
Adult men with high school	-0.037	-0.218	4.272
	(0.177)	(0.280)	(3.251)
Adult women with high school	0.146	0.426	-2.135
	(0.173)	(0.223)*	(2.766)
Δ hh's with water	-0.004	0.012	0.137
	(0.012)	(0.013)	(0.230)
Δ hh's with toilets	0.376	0.379	-2.173
	(0.142)***	(0.148)**	(0.956)**
N	406	406	406
R-squared	0.162	0.351	0.211

Table 10: Robust standard errors in parentheses, clustered at main place level. Sample is restricted to those areas that had been treated with an Eskom project prior to 1996; they are not in the sample for the main analysis. Employment outcome variables are measured as employment/population proportions for adult Africans. Other included variables are the proportion of Indian/White adults with > grade 8, distance to road and town, proportion of adults with matric, adult sex ratio, proportion of female headed households and household poverty rate, change in proportion of households with access to water close by and to flush toilets. All X variables are measured in 1996 except change in other services. Each regression includes a full set of district fixed effects.

Table 11: Spill-overs to control areas further from prior treatment communities

Coeff. on treatment dummy for women	(1) OLS	(2) IV
Full sample	0.001	0.135
N=1992	(0.005)	(0.062)**
Control areas >1km from an area treated before 2001	-0.005	0.104
N=1656	(0.006)	(0.061)*
Control areas >5km from an area treated before 2001	-0.005	0.114
N=1374	(0.008)	(0.097)

Table 11: Each coefficient (standard error) is from a separate regression. Robust standard errors in parentheses, clustered at main place level. Outcome variable is change in proportion of employed African women aged 15-59. All other variables (except treatment) are measured in 1996. Treatment= 1 if community had first Eskom project between 1996 and 2001. Successive sample restrictions condition on any part of a control community falling outside of an X kilometer radius of an area treated prior to 2001.

Table 12: Excluding in-migrants

	Women				Men			
X- variables	OLS		IV		OLS		IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.000	0.001	0.110	0.135	-0.013	-0.010	0.078	0.084
	(0.005)	(0.005)	(0.068)	(0.071)*	(0.007)*	(0.006)*	(0.080)	(0.069)
Kms to substation 1996*10		0.002		0.005		0.003		0.005
		(0.003)		(0.004)		(0.003)		(0.004)
Household density 1996*10		0.000		-0.002		0.002		0.001
		(0.001)		(0.001)		(0.00013)*		(0.002)
Poverty rate 1996		0.024		0.021		0.051		0.049
		(0.009)***		(0.014)		(0.014)***		(0.016)***
Prop female headed hh 1996		0.047		0.032		0.248		0.238
		(0.019)**		(0.026)		(0.031)***		(0.035)***
Adult sex ratio (F/M) 1996		-0.003		0.013		-0.018		-0.006
		(0.008)		(0.012)		(0.012)		(0.015)
N	1992	1992	1992	1992	1992	1992	1992	1992
R^2	0.000	0.025			0.003	0.138		
Standard C.I.				[0.05-0.3]				[0-0.3]
AR C.I.				[0.05-0.35]				[0-0.3]

Table 12: Robust standard errors in parentheses, clustered at main place level. Outcome variable is change in proportion of African females aged 15-59 who are employed, excluding the count of all recent in-migrants from the numerator. All other variables (except treatment) are measured in 1996. Treatment=1 if community had first Eskom project between 1996 and 2001. Other variables included: number of Indian/White adults, distance to road and town, proportion of men and women with high school, change in proportion of households with water close by and with flush toilets. All regressions include 10 district fixed effects.

Table 13: Age specific employment effects for women

	Age 15-19		Age 20-24		Age 25-29		Age 30-34		Age 35-39		Age 40-44		Age 45-49		Age 50-54		Age 54-59	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
T	0.000	0.004	0.000	0.009	-0.001	0.018	0.000	0.039	0.001	0.030	0.001	0.012	0.001	0.019	-0.001	0.002	0.000	0.003
	(0.000)	(0.007)	(0.001)	(0.014)	(0.001)	(0.013)	(0.001)	(0.019)**	(0.001)	(0.015)*	0.001	(0.013)	(0.001)	(0.008)**	(0.001)	(0.006)	(0.001)	(0.005)
N	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992	1992
R^2	0.013		0.076		0.085		0.135		0.036		0.028		0.000		0.000		0.044	

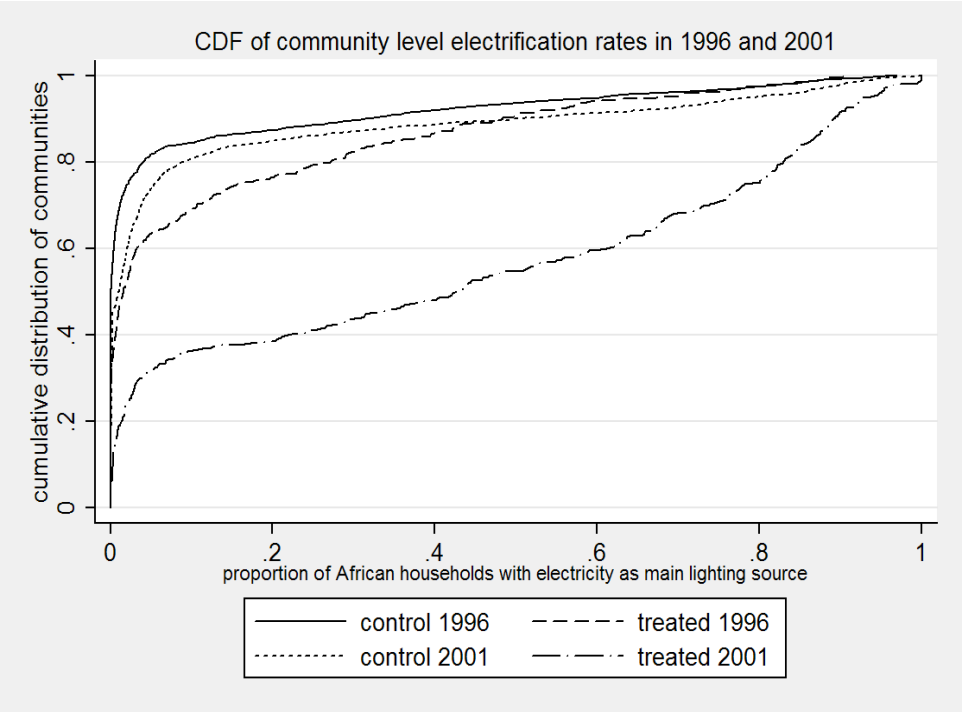
Table 13: Notes: Robust standard errors in parentheses, clustered at main place level. Outcome variable is change in proportion of employed African females in respective 5 year age cohort. All other variables (except treatment) are measured in 1996. Treatment=1 if community had first Eskom project between 1996 and 2001. Other variables included: distance to the grid, household poverty rate, adult sex ratio, proportion of female headed households, proportion of Indian/White adults, distance to road and town, proportion of men and women with high school, change in proportion of households with access to water close by and to flush toilets, 10 district fixed effects. Excluded instrument is average land gradient.

Table 14: Heterogenous effects by percent male adults

X variables	Females		Males	
	OLS	IV	OLS	IV
Treatment	0.060	-0.528	0.022	-1.578
	(0.045)	(0.635)	(0.069)	(0.985)
Treatment*percent men	-0.142	1.620	-0.075	3.971
	(0.112)	(1.586)	(0.175)	(2.452)
Percent men	-0.156	-0.520	-0.053	-0.671
	(0.073)**	(0.250)**	(0.103)	(0.356)*
Interaction effect at mean percent male	-0.058	0.664	-0.031	1.628
(0.41)	(0.046)	(0.650)	(0.072)	(1.005)
Treatment + interaction at mean	0.001	0.136	-0.009	0.050
pvalue	(0.760)	(0.040)	(0.100)	(0.550)
N	1991	1991	1991	1991
R-squared	0.09		0.185	

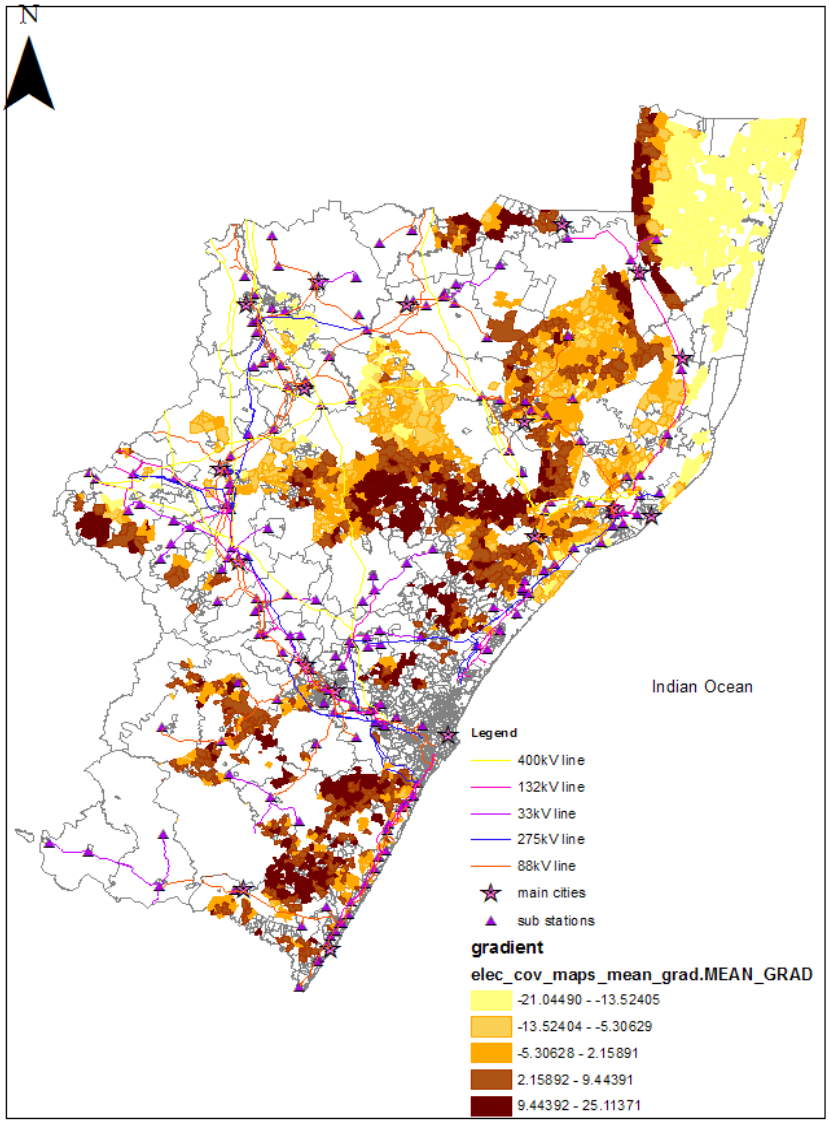
Table 14: Robust standard errors in parentheses, clustered at main place level. Sample restricted to those areas that had been treated with an Eskom project prior to 1996; they are not in the sample for the main analysis. Employment outcome variables are measured as employment/population proportions for adult Africans. Other included variables are the proportion of Indian/White adults with > grade 8, distance to road and town, proportion of adults with matric, adult sex ratio, proportion of female headed households and household poverty rate, change in proportion of households with access to water close by and to flush toilets. All X variables are measured in 1996 except change in other services. Each regression includes a full set of district fixed effects.

Figure 1



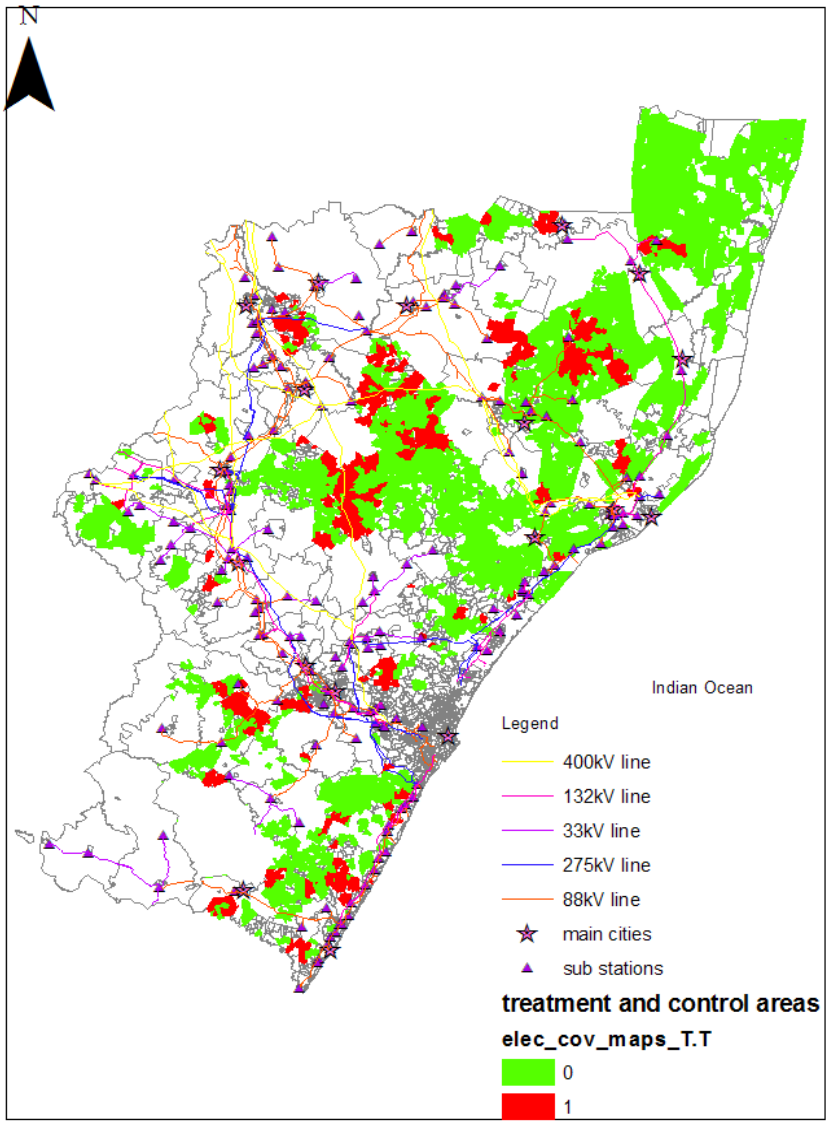
Notes: Empirical cumulative distribution functions of the proportion of households with electricity as main source of lighting in Census 1996 and 2001. Treatment group is the set of communities that had an Eskom electrification project between 1996 and 2001; control group is the set of communities that had no projects before 2001 or no projects at all between 1990 and 2007. Data from 1996 are considered 'before', data from 2001 are considered 'after' the project.

Figure 2



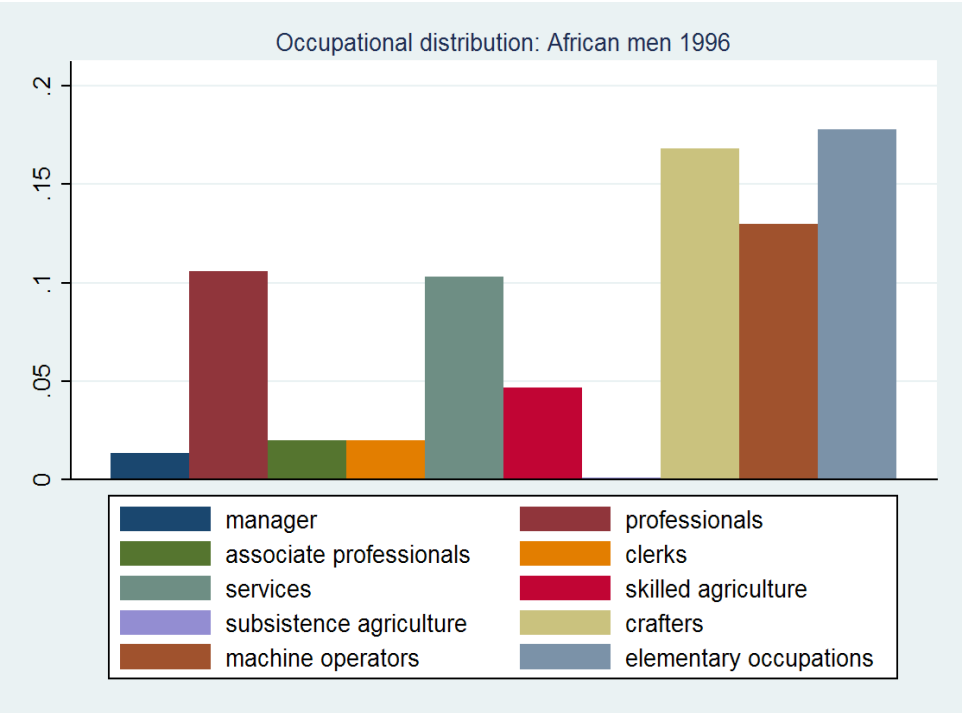
Notes: Map of KwaZulu-Natal region. Shaded areas are in the sample. Areas of steeper average gradient are in darker brown shading; areas of flatter average gradient are pale yellow. Lines represent electricity grid lines in 1996, triangles are electricity substations in 1996 and stars represent towns.

Figure 3



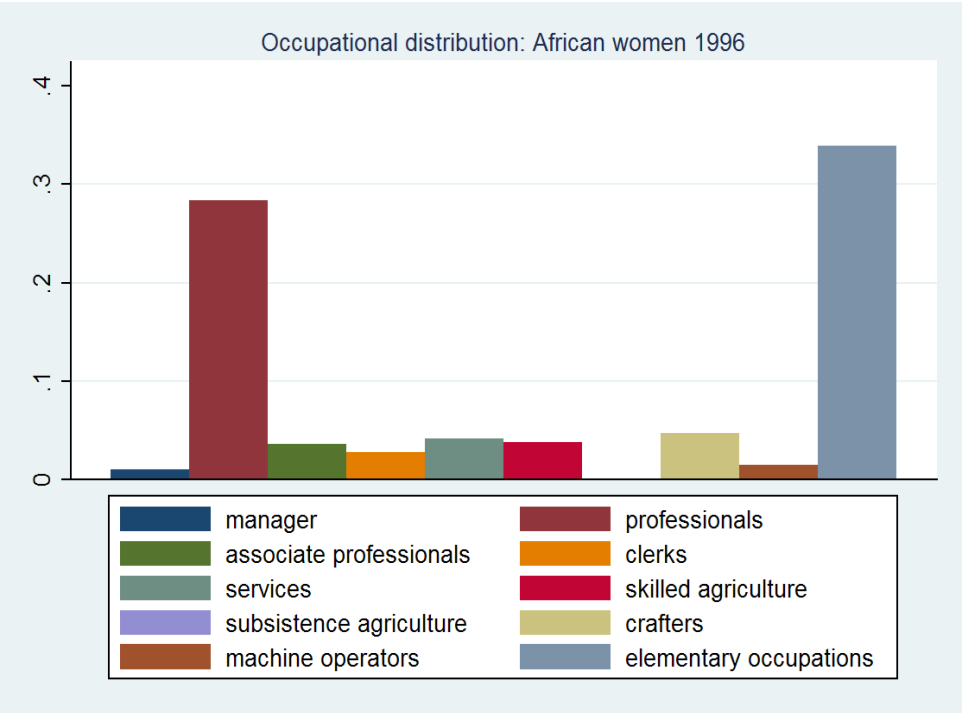
Notes: Map of KwaZulu-Natal region. Shaded areas are in the sample: red areas are treated with an Eskom project between 1996 and 2001, green areas are treated after 2001 or not at all. Lines represent electricity grid lines in 1996, triangles are electricity substations in 1996 and stars represent towns.

Figure 4a



Notes: Distribution of occupation groups for men in Census 1996. Groups are defined as the number of men employed in each occupation over all employed men.

Figure 4b



Notes: Distribution of occupation groups for women in Census 1996. Groups are defined as the number of women employed in each occupation over all employed women.