Can rural electrification jump-start employment? Evidence from KwaZulu-Natal, South Africa *

Taryn Dinkelman
Department of Economics
University of Michigan

PRELIMINARY. DO NOT QUOTE OR CITE.

September 8, 2007

Abstract

Access to modern fuels is often argued to have gender-specific effects on labor outcomes. However, there is little empirical evidence on whether or how electricity affects employment of women engaged in time intensive home production. This paper investigates the impact of domestic electrification on employment in rural South Africa, where more than 60% of households still rely on wood for basic energy needs. Between 1995 and 2001, electricity infrastructure was rolled out to over two million households. I assess the impact of electrification on community level employment using two waves of aggregate Census data matched with technical, administrative and geographic data on the electricity grid. I exploit community-level variation in project timing to estimate district fixed effects models of changes in employment rates, controlling for baseline variables. To address endogenous placement of infrastructure and measurement error in the treatment variable, I instrument for treatment with electricity using land gradient that affects the cost of grid expansion but is unlikely to directly affect changes in employment outcomes. IV results indicate that female employment rates increase by 13.5 percentage points in treated areas, but there are no significant effects for men. A series of checks show that results are not driven by selective migration, spatial spill-overs or pre-existing trends in employment growth. In probing the channels through which electricity has large female effects, I find that women in poorer areas, women in their thirties and women in skilled and semi-skilled jobs experience the largest increases in employment. In this reduced form approach, I cannot conclusively rule out that electricity creates new demand for labor. However, the characteristics of this technology (very low levels of power), the lack of spill-overs between areas and the fact that treatment effects are only significant for women suggests that a net labor supply interpretation of the employment results is not unreasonable.

*Email tdinkelm@umich.edu. Eskom data and institutional background kindly provided by Jenny Barnard, Sheila Brown, Ed Bunge, Jim Stephenson, Steven Tait, Innocent Nxele and Amos Zuma. Important comments from and conversations with Matias Busso, HwaJung Choi, John DiNardo, David Lam, Claudia Martinez, Justin McCrary, Dori Posel, Samara Potter and Jeff Smith 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 an NSF 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 has budgeted for $20 billion to expand generation capacity and the world's largest hydro electric power plant which would generate enough power for Africa is planned for the Democratic Republic of Congo.[2] While these initiatives are predominantly geared towards industry, they may have the potential to affect outcomes for women and children if they enable 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 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

[1] World Bank, $http://ppi.worldbank.org/explore/ppi_{e}xploreSector.aspx?sectorID = 2$

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

[3] Behaviors for children are not analyzed in this paper; this is because school enrollment in South Africa is almost universal up to the legal school-leaving age of 15.

grid. A key feature of this roll-out was that activity was concentrated on household connections and not 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, understanding technological constraints on infrastructure roll-out can provide legitimate exogenous variation in allocation of infrastructure. In this paper, I use data on key constraints driving roll-out to construct causal estimates of the effect of electrification on employment. I compare changes in employment rates in areas that have had electrification projects (treated areas) to those that have not (control areas) and instrument for treatment status using land gradient, a key variable driving costs of electrification.[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 additionally adjusts for some differences in growth paths across communities.

Defining treatment status with project data is certainly preferable to defining treatment based on use. 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 areas that are lagging behind (pulling OLS estimates down again). To address this indeterminate bias I instrument for allocation of an electricity project using average community land gradient.

[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]This is similar in spirit to Duflo and Pande (2006) who look at the impact of large scale dam construction on farming productivity and poverty. Their IV strategy also relies on geography. In that paper, dam infrastructure had significant public goods features and substantial spill-overs into non-dam areas. Domestic electricity has more features of a private good and is less likely to have spill-over effects across districts unless electrification stimulates labor demand.

Gradient directly affects the average cost per connection. During the period, cost was the primary factor in prioritizing areas for electrification since Eskom was financing all roll-out and had to meet annual household connections targets. Gradient is also an empirically strong predictor of treatment status and of the change in the proportion of households with electric lighting over the period. My identification assumption is that conditional on district fixed effects, baseline controls and detailed measures of proximity to local labor market opportunities, land gradient is unlikely to directly affect *changes* in employment rates. Conditional on instrument relevance and validity, this strategy identifies the local average treatment effect: the average treatment effect for those areas induced to be treated by the instrument.

I argue that this IV strategy identifies responses to exogenous changes in electrification– that is, that electrification on the basis of lowest average cost does not follow demand. However, this research design is unable to definitively establish whether electricity generates an employment response by directly stimulating new demand for labor, or whether it affects labor supply. The labor supply channel is, however, more plausible for several reasons. First, roll-out was driven by household level targets (not firm targets), and the level of service 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 something that is apparent in my data. Finally, the types of firms stimulated by electricity would have 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.

My results indicate that female employment rates are sensitive to the presence of electrification infrastructure in rural areas, but that 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 employment rates are falling.

Some of the effect of electrification may operate through differential in-migration in response to treatment. However, in a simple bounding exercise, I find that migration of employed or employable individuals cannot explain the female employment effect. Furthermore, a false experiment provides no evidence that areas to be treated in future have faster employment growth rates prior to treatment– in fact, they appear to have lower employment growth before treatment.

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 heterogeneity in treatment effects and probe 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, and that women with better market opportunities are also more likely to experience the short run gains. In addition, poorer areas where households are initially more constrained in terms of alternatives to electricity exhibit the largest employment gains.

This paper contributes to a growing microeconomic literature on the effects of physical infrastructure in developing countries in two ways.[8] First, I estimate the effects of electrification using actual project data that captures changes in access only rather than changes in access and 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 mixed, and varies by type of infrastructure. [Duflo, 2001] finds that school construction in Indonesia adds an extra 0.12 to 0.19 years of schooling with an associated 6.8 to 10.8 wage return; [Banerjee et al.,] find that Chinese wages are 37% higher in areas transected by railroads; and [Akee, 2006] estimates the effect of road construction on wage and agricultural employment in the island economy of Palau at a substantial 27 percentage points and -14.7 percentage points respectively.

Since very little work has been done on the time use effects of electrification, I begin by describing the context in which we might expect electricity to have effects on time use. I use data from a panel of households in rural KwaZulu-Natal (KZN) to present suggestive evidence that electrification can plausibly reduce time spent in fuel wood collection and lead to households obtaining time saving appliances complementary to electricity. Section 3 describes conceptually

[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.

[8]There is an established macroeconomic literature which estimates the effects of public infrastructure on total factor productivity using time series data. Estimates of the elasticity of total factor productivity to public infrastructure are typically large and positive, ranging from 0.2 to 0.5, with electricity having the average elasticity. See [Aschauer, 1989] and [Canning, 1998]. [Fedderke and Bogetic, 2005] find that electricity has the same total factor productivity elasticity as average infrastructure investment in South Africa (0.2).

[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.

how a positive shock to home production technology may have ambiguous effects on labor supply, and suggests for which groups effects may be larger. Sections 4 and 5 describe the data and the context of South Africa's electrification; while section 6 outlines the empirical strategy and section 7 presents main results and specification checks. Section 8 probes the channels through which electrification affects employment, and section 9 concludes with a discussion of the paper's main findings.

2 Effects of electricity

2.1 Energy poverty in South Africa

The KwaZulu-Natal Income Dynamics Study (KIDS) provides suggestive evidence that domestic electrification may alter time use and the structure of home production. KIDS interviewed 713 African households in rural areas of the province in 1993 and 1998.[10] I use the question "Are you connected to the electricity grid?" to identify which households did not have access in 1993 and of those, which households gained access by 1998. Additional questions in each wave capture the average amount of household time (in minutes per week) spent in collecting fuel-wood, ownership of electrical appliances and market work. The data illustrate what happens to time use, appliance ownership and employment in households that did and did not get access to electricity between 1993 and 1998.

Table 1 describes features of the matched sample of households in 1993 and 1998. Between 35% and 66% of households must fetch wood on a weekly basis for own use. In both types of households– those that gained access by 1998 and those still without– females are primary fuel wood collectors. In 1993, they spend at least five hours per week collecting fuel. Restricting to the set of households that collect wood at all, fuel wood collection absorbs between 14 and 15 hours per week, or at least 2 person-work days. Figure 1 plots the density of log total weekly minutes spent on fuel-wood collection. The set of households contributing to all three densities are those without access to the grid in 1993. By 1998, both densities have shifted leftwards. Households gaining access to the grid experience a larger shift in this distribution.[11]

Appliances complementary to electricity enhance the labor-saving effects of this new technology. In Table 2, columns (1) to (3) present household fixed effects models of appliance ownership. The coefficient on the interaction between year and connected to the grid is identified by households

[10]Further description of the survey appears in the data appendix.

[11]Each of the 1998 distributions is significantly different from the 1993 distribution: the p-values from the Kolmogorov-Smirnov test for differences between empirical distribution functions are 0.05 (for households that do not get connected) and 0.000 (for those that do get connected). The two distributions in 1998 are not significantly different from each other (p-value = 0.138); however, sample sizes for each group in 1998 are relatively small to detect significant shifts.

that switch from no access in 1993 to access by 1998. In these households, ownership of fridges increases by 34% while kettle ownership rises by 37%; stove ownership also increases by a smaller and insignificant 6.5%. All of these appliances are potentially time-saving, particularly for households previously reliant on wood.[12]

The final two columns of Table 2 present household fixed effects estimates of male and female employment. Over time, employment is increasing for both men and women, but male employment probabilities are lower and female employment probabilities are higher in households that gain a connection. Since electricity is not expected to reduce the probability of being employed, these results suggest that the households getting access to the grid are selectively choosing a connection or are in areas where the decline in male work opportunities and increases in female labor market opportunities confound the effects of electrification. There may also be something about the set of men and women who remain in the panel (i.e. do not attrit) that drives this result. Regardless, KIDS data do not necessarily reveal the causal effects of getting access to the grid.

Since there is no geographic data for KIDS, I cannot identify which areas had an Eskom electrification project nor can I implement the IV strategy– hence the need to use the Census. What these data do suggest is that home production technology is likely to change after a household is electrified and that we should expect to see larger changes in time use for women than for men, since women are disproportionately involved in collection of and food preparation with traditional fuels.

2.2 Conceptual framework: linking electricity to labor supply

In Becker's (1965) canonical model of time use and Willis' (1979) model of fertility, home production and female labor supply, households use time and market goods to produce time and goods intensive commodities. Women supply labor to the market if their market wage is at least as large as the shadow price of time in home production. The shadow price of time in the home– also described as the marginal cost of time– depends on household preferences for time and goods intensive commodities and on existing home production technologies. In equilibrium, women supply labor up to the point where these two wages are equal, or do not work at all when the shadow price of time is higher than the market wage. Conditional on a market wage, women with a higher value of marginal product in the home (the marginal product of her labor in home production multiplied by the household's valuation of the output of this production) are less likely to work. Any factor affecting the technology of home production may therefore alter labor supply on the extensive and intensive margins.

[12]It is not possible to distinguish electric and gas stoves (or TV and radio). Anecdotal evidence suggests that televisions were widespread in rural areas even before electricity arrived. Televisions can run on car batteries that are charged on a weekly basis and many furniture stores sell car batteries for this purpose.

The arrival of infrastructure for domestic electricity may be characterized as a positive shock to time productivity.[13] 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 shadow prices of time- and goods-intensive commodities fall, but asymmetrically so. At a give time intensity of production, the shadow price of the time intensive commodity falls further.

This change induces an income and substitution effect. The increase in effective time raises real income, and demand for leisure and all normal commodities rises. The fall in the shadow price of the time intensive commodity relative to the goods intensive commodity causes the household to substitute towards the 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. As [Gronau, 1986] points out, the net effect on time supplied to home and market production is ambiguous, since changes in home productivity interact with household preferences to determine relative shadow prices and the marginal cost of time. In cases where 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.

Differences in labor supply effects of this technology shock are therefore linked to heterogenous preferences for time- and market-intensive commodities (which we can't measure), differences in the value of female time in the home and market and differences in initial home production technology (some of which we can measure). Women are more likely to increase their market labor supply when the shadow price of time is initially close to the market wage. For example, women with good market opportunities are more likely to meet the participation constraint when an improved technology reduces the shadow price of their marginal hour in home production. Women who have 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. Finally, households relying on highly inefficient production technologies will experience more labor savings with the advent of electricity than other households, and so the cost of time in the home will fall more dramatically for these women. The work of [Becker, 1965], [Willis, 1973] and [Gronau, 1986] indicates that whether electrification leads to increases in short run labor supply is an empirical question and that we should expect differential effects for different types of households.

Note that electrification may also affect the technology of market production, although this is less likely with small scale household electrification. If market work is self-generated as in the case of craft work and small scale service workers, then domestic electrification may create more

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

jobs in addition to releasing labor time into the economy. If this demand effect was dominant, we might expect to see (i) larger employment growth in low skill occupations, (ii) no differential response for men and women or for women of different ages and (iii) different employment effects when comparing treated areas to adjacent versus non-adjacent control (i.e. control areas at different risk of experiencing labor demand spill-overs). In what follows, I present some evidence that these three effects are not observable in the data.

3 Details of the electrification program

Eskom, South Africa's national electricity utility, is entirely responsible for electricity generation and transmission and is the sole distributor for most rural areas.[14] 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, however, was denied to many African households, particularly those in homelands. Addressing this backlog in domestic connections became a development priority under the National Electrification Programme (NEP).[15]

As part of the NEP, Eskom committed to electrify 300,000 households annually from 1995 onwards. 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* 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. The 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

[14]The following detail is 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.

[15]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.

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

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."

Under the NEP, Eskom had committed to an annual connections target that they regarded as "firm and non-negotiable" [Eskom, 1996] and fully subsidized all new grid connections [Gaunt, 2003].[17] These 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 in South Africa:
Distance: The bulk of electrification cost is in laying distribution lines. Hence, how far communities are from the existing substation infrastructure and high voltage lines 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].
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.
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 build transmission poles [Eskom, 1996].

All three measures are observable in the data sets I have assembled. However, 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 form the basis of my IV strategy. I provide further arguments for the validity of this exclusion restriction in the empirical strategy below.

Since project incidence is my measure of treatment, not all households in an area may actually be treated for electricity and, in addition, not all households treated with a connection may be able to use this electricity.[18] Figure 2 verifies that getting new access to infrastructure did indeed 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, the only community Census variable that measures use of electricity.[19] Treated areas do begin with

[17] In early years, connection fees were charged to consumers but never collected.
[18] To use electricity, customers had to purchase pre-paid electricity cards from local stores.
[19] Interview with Gisela Prasad: "Electric lighting was synonymous with the roll-out".

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] Being treated with an Eskom project clearly increases the proportion of households using electricity relative to the control group.

The fact that not all households may be able to afford electricity cautions us against inferring causal effects by comparing electricity using versus non-using households in a cross section, or even over time. Infrastructure roll-out provides a far cleaner measure of access. Before the roll-out, few households had the choice to use electricity while after the roll-out, many more households had this new opportunity. The effects of electrification are, however, still estimated for those groups that are able to use this service in the short run.

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 in 1996 indicates that 2.7% of urban African households used wood for cooking while 63.4% of rural households did so. Second, there are potentially fewer confounders in rural areas than in urban areas. In the empirical work, I am able to control for changes in access to other development services which are a likely source of confounding during the period. Finally, although urban electrification expanded fastest in the early 1990s, as the grid expanded and the technology of domestic electrification developed to supply smaller power loads, more rural areas were connected. 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.[21]

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. A detailed description of this exercise is provided in the appendix.

[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]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.

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 1,994 areas in rural KZN; their boundaries are illustrated in Figure 3. Notice that the former homeland KwaZulu was broken up into many pieces across the province. This was to accommodate Africans in areas that were deemed inhospitable for white settlement.[Christopher, 2001]

The community aggregate data provides full (weighted) population totals for each year for different combinations of variables. Key variables that can be constructed include the proportion of households with electricity in each year and the proportion of African adults in different age groups in different labor market states. In 1996, the question 'Does the person work?' as asked of all adults, where work was defined as including 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, the question was '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). These questions are relatively comparable over time, so under-counting of informal work which is typical in developing country contexts should be similar over time.

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 the access each community has to local economies.

4.2 Infrastructure data

I collected technical data on the location of Eskom's distribution network in 1996 from Eskom planning engineers. I combine the infrastructure database with the Census spatial data and construct a measure of straight line distance (in kilometers) from each community to substations that are part of the 1996 grid. This variable captures one of the key cost variables determining placement of electrification projects, since substations are necessary pieces of infrastructure required for stepping down electrical power to lower voltages appropriate for domestic use. Figure 3 illustrates the position of these sub-stations as triangles. Note that substations are not all 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 at the end of the 1980s, which ensured that grid infrastructure extended into rural areas.

4.3 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 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).

4.4 Project data

To assign treatment status to each community, I collected administrative data from Eskom on the number of new household connections per year by location for the period 1990 to 2007. In most locations, there is a spike in household connections in one year, indicating an Eskom electrification project. I define this year as the treatment year. One critical issue is that boundaries of Eskom regions do not line up with Census boundaries. I overlay Eskom's spatial infrastructure data linked to the project information on top of the Census boundaries to assign treatment status to communities: for any community that lies inside a project area, all of the information from that project was assigned to that community. This introduces some error in treatment assignment: some communities will be assigned full treatment status when only a small percentage of households in the area were treated.

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 analysis; there are 407 of these out of the total 2407 tribal areas in the sample (17%). 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 for between 1996 and 2001 ($T_{connect}$).

[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.

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 use OLS to estimate the following for the average treatment effect α_2:

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

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 possible. Specific terms affecting selection can be written in an error components framework:

$$\epsilon_{jdt} = \mu_j + \delta_{jt} + \lambda_{dt} + \nu_{jdt} \tag{2}$$

where μ_j is a community fixed effect, δ_{jt} is a community trend term, λ_{dt} 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:

$$(y_{jdt+1} - y_{jdt}) = \alpha_1 + \alpha_2 \Delta T_{jdt} + \Delta\delta_{jt} + \Delta\lambda_{dt} + \Delta\nu_{jdt} \tag{3}$$

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

There are three reasons to suspect that even in a 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 unobservables may occur if electrification projects are allocated to areas that are growing faster for unobservable reasons ($\text{cov}(\Delta T_{jdt}, \Delta\delta_{jt}) > 0$ or $\text{cov}(\Delta T_{jdt}, \Delta\lambda_{dt}) > 0$), and $\alpha_{2,OLS}$ will be biased upwards. This is the usual concern with estimating the returns to infrastructure development.

Second, negative selection on unobservables may occur if projects are targeted to the more disadvantaged areas conditional on being low cost (that is, $\text{cov}(\Delta T_{jdt}, \Delta\delta_{jt}) < 0$ or $\text{cov}(\Delta T_{jdt}, \Delta\lambda_{dt} < 0)$). In this case $\alpha_{2,OLS}$ will be biased downwards since the control group will overcompensate for trend in the treatment group. This second source may be more likely where electrification was driven by a socio-political compact between Eskom and the newly elected government.

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 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.[23]

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.[24] First, I control for baseline X-variables (X_{jd0}) that should affect a community's growth path ($\Delta\delta_{jt}$). 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 and measures of adult educational attainment in the area. Female headed households and adult sex ratios are included as additional indicators of community poverty.[25] I also include district fixed effects in this first differenced regression to take out common differences across local labor markets over time ($\Delta\lambda_{dt}$). However, the Census does not provide a rich set of covariates to control for differences in growth paths across communities. Hence, I also instrument for program placement using mean community land gradient (Z_j). The first stage regression for ΔT_{jdt} is:

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

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. I also assume that gradient is not correlated with the measurement error in the treatment variable. This latter seems a reasonable assumption, given that the source of the error mainly arises from mismatch in administrative boundaries.

Gradient is a plausible candidate for an instrument for several reasons: first, gradient is theoretically one of the three main cost drivers of electrification as described earlier. Empirically, gradient is also a good predictor of treatment assignment. Second, gradient predicts community level changes in use of electricity 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, gradient is unlikely to be directly correlated with unobservables related to changes in outcome variables. The typical concern with gradient is that it directly affects agricultural productivity: it alters the ability of the land to retain rainfall (as opposed to run-off) and the

[23]See [Lewbel, 2007] for a discussion of the effects of measurement error in a binary treatment variable.

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

[25]Standing, Sender and Weeks (1996) argue that both measures are good indicators of an area's poverty status in South Africa.

extent of soil erosion (Lal, 1998) and hence soil quality for crop yields.[26] Although my study area is rural, it is not traditionally agricultural. The baseline occupational distribution for men and women is shown in Figures 5a and 5b. Particularly for women, these areas support a bimodal distribution of occupations: public sector work (teaching, nursing, municipal officials) and elementary occupations (crafters, domestic workers, street vendors, machine operators), with almost no formal or informal agriculture. While the Census undoubtedly under-counts some subsistence agriculture across the country, the former homeland area is not regarded as a major source of subsistence farming.[27] Hence, gradient here is highly unlikely to be affecting real opportunities for work.

The only case in which gradient would be considered a questionable instrument would be if gradient directly lowered the costs of developing factories and places of work - or, if flatter places were growing faster anyway. As a specification check, I perform a false experiment using only the set of areas that are never treated during the period and those that are treated after 2001. The question this experiment attempts to answer is: compared to never treated areas, are there any signs of employment growth between 1996 and 2001 in areas that are only treated post-2001? If the false experiment rejects a zero coefficient on the instrumented treatment variable, this would be evidence against the assumption that gradient only affects allocation of electricity projects. As a further check, I also test for whether the locations of the main female employers (schools and white and Indian households) are correlated with the gradient. If these "places of work" are not related to gradient, then we have more confidence in the exclusion restriction in (15).

Note that both OLS and IV results for employment encapsulate any effects of electrification on migration.[28] 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.[29] Nevertheless, a migration response is also an outcome of interest. 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

[26] Land gradient has been used in different contexts as a key control in estimating agricultural production functions [Udry, 2000].

[27] This is primarily because Africans were moved onto homeland areas on the worst pieces of land in the province. See Simkins (1976) 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 Standing, Sender and Weeks (1996) chapter 6 and Aliber (2002) for an outline of how homeland areas do not generate the majority of income from agricultural produce.

[28] This issue confronts Black et al (2004) 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.

[29] Strauss and Thomas (1995) provide a good discussion of selective migration concerns in program placement studies in developing economies.

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 Figure 4. All communities in this sample are rural, former tribal areas of KZN province. The map highlights several relevant placement issues. First, not all treated areas are positioned close to 1996 grid infrastructure and many areas adjacent to the grid are control areas. Hence, being close to the original grid is neither necessary nor sufficient for subsequent electrification. Second, not all treated areas are close to towns, so proximity to a local economic center is not a prerequisite for having an Eskom project. Third, there is a good distribution of treated areas across the entire province.

Table 3 presents descriptive statistics for the sample in 1996. Means and standard deviations are provided for the 1,994 communities and separately for each of the treatment and control groups. On average, over 50% of households are female headed and the female/male adult sex ratio is well over 1.[30] This area is very poor: 61% of households live on less than 6,000ZAR per year. Some differences in the baseline year observables for treatment and control areas suggest program placement may be targeted towards wealthier communities. The adult sex ratio is significantly higher in treated areas, as are the proportions of high school educated men and women. Treated areas are on average 2.8 kilometers (1.7 miles) closer to the nearest road and town compared to control areas. However, there is no significant difference in the proportion of female headed households by treatment area, and the difference in poverty rates is small and significant only at the 10% level. Treatment and control areas are also not significantly different in terms of the proportion of White and Indian adults in the population. This variable proxies for potential employers in the area.[31]

Given that Eskom was targeting low average cost areas, 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

[30] Sex ratios in these areas are typically very skewed towards females since homelands were historically migrant labor reservoirs for mines in the interior of the country.

[31] 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.

"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".[32] Figure 6 illustrates the overlap in mean gradient distributions for treatment and control areas.

Since average gradient may mask substantial variation in terrain within a community, the lower part of table 3 provides additional information on gradient. Treated areas have a lower value of major gradient points than control areas, 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 4 shows unadjusted difference-in-difference estimates for a range of basic services across treatment and control areas. Average electrification rates rise from 13% to 42% in treated areas– a statistically significant difference of 26 percentage points over the period. Measurement error in the treatment is also apparent: control communities also experience a rise in electrification rates, although this is a much smaller change. The rest of Table 4 provides evidence that Eskom projects are not correlated with expansion of other development related services. Access to sanitation did not change differentially across treatment and control areas and access to water close actually *degenerated* in treatment relative to control areas. Access to a telephone in the home rose by a significant three percentage points in treated areas. This increase is likely driven by changing access to cell-phones within the home, which require electricity for charging.

To uncover initial evidence of employment effects of electrification, Table 5 compares changes in outcome variables across treatment and control groups. The main outcome variable is the employment to population rate of Africans aged 15 to 59 inclusive. Over the period, employment rates fall by 3.6 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 1.9%. These results are unusual; electrification is not expected to reduce employment. However, a combination of measurement error in the treatment variable, negative selection into treatment and differential in-migration may drive these unadjusted differences-in-differences results.

The second part of table 5 suggests that migration may contribute towards the fall in male employment and little change in female employment. Treated areas begin with higher populations in 1996 but they 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. As the last two panels of the table indicate, treatment areas appear to have a slower reduction in in-migration rates than control areas. In-migrants are adults who report moving in to their current 'area' in the five years before the relevant Census. Although this question is not

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

as specific as it could be, the information on in-migrants does tell us something about differential trends across treatment and control areas.

Overall, descriptive statistics indicate that treated areas are better on some baseline outcomes although not all of the key poverty variables; there is some measurement error in the treatment variable since some communities are getting electricity in other ways too, although the vast majority are getting access through this program; there is faster population growth into treated areas in the five years before the Census than into control areas. This growth in the number of people in these areas will be captured in both OLS and IV results below.

6.2 OLS and IV main results

6.2.1 First stage

First stage estimates for assignment to treatment are presented in Table 6.[33] A one standard deviation increase in gradient (about 10 degrees) reduces the probability of being treated between 1996 and 2001 by 4%. The F-statistic on this coefficient begins in the low range without other controls, but increases to a reasonably high number once including other relevant controls measured in 1996.[34] The other 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%, while a one standard deviation increase in household density (30 households) increases the probability of treatment by 3.9%.

Including district fixed effects improves the precision of the gradient coefficient, but does not alter the point estimate on gradient. Distance from the grid appears to matter more across districts, as the coefficient falls substantially with the inclusion of district fixed effects. Interestingly, proximity to the nearest town and road are not significant predictors of assignment to treatment. Having a higher fraction of women with at least a high school education raises the probability of treatment substantially: at the mean of this variable (0.068), predicted treatment probability is 5.3% points higher. Adding in controls for the change in access to other household services (water, sanitation) has no impact on the gradient coefficient.

The first stage provide some mixed evidence that treated areas may be negatively selected on wealth or income. First, the community poverty rate variables have positive coefficients, although these are never significantly different from zero. Second, while adult sex ratio and female headed households are significantly different from zero, their signs provide opposing messages: in areas with a higher proportion of female headed households (i.e. poorer), assignment to treatment is less likely to occur; but in areas with higher female to male sex ratios (i.e. poorer), treatment is more likely

[33]Results from a logit model of the treatment are very similar to these linear probability model results. Using modal gradient as an instrument and both modal and mean gradient produces very similar results.

[34]I implement inference tests that are robust to potentially weak instruments in the second stage results.

to occur. If positive selection was driving project priorities, we would have expected negative and significant coefficients on all three poverty related variables.

Columns (5) and (6) report first stage coefficients for two alternative specifications of the treatment variable: an exposure measure counting the number of years since treatment (0 for never treated and 1,2,3,4 or 5 for those treated between 1996 and 2001), and the fraction of households connected during the period. A one standard deviation increase in gradient reduces the year of treatment by 0.12 (i.e. leads to projects happening later or not at all), and reduces the fraction of households connected by 2.2%. Although the strength of the first stage varies across dependent variables, gradient clearly 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.

6.2.2 Employment: second stage

Tables 7 and 8 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 the mean differences presented earlier: 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. 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.

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 3.9 percentage points and is not significantly different from zero. To address concerns about incorrect 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 test for women can strongly reject zero, although the confidence interval is wide, between 10 and 45 percentage points. The male test cannot reject zero. [35] In a community with a median number of adult women (N=264), this 13.5 percentage point increase in female employment translates into an increase from 21 women working to about 57 women working.

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 example, 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

[35]See Cruz and Moreira (2005), Mikusheva and Poi (2006) and Chernosukov and Hansen (2007) for a description of Anderson-Rubin tests and a motivation for why they are robust to weak identification even under heteroscedasticity.

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: hence, the employment effect is not being driven by access to other services that may also affect the technology of home production.

7 Specification checks

In this section, I present several different pieces of evidence that together bolster the research design and results.

7.1 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 9 reproduces OLS results for the full sample and for successive sample limitations.[36] For women, OLS results become larger and positive as (i) largest areas are dropped from the analysis, and as we restrict to (ii) areas that have at least 80% of households connected or (iii) with large changes in electrification rates. Coefficients for men also rise above -0.019 as the sample size falls, but treatment is never statistically significant across samples.

The rise in the OLS coefficients suggests that some measurement error is present in the treatment variable. However, even with a cleaner measure of treatment, the treatment effect is under 2 percentage points for female employment. Admittedly, this is a fairly crude way to test the implications of measurement error, but it is unlikely that measurement error alone explains the entire gap between OLS and IV results. In addition, measurement error in a binary explanatory variable would bias coefficients towards zero ([Lewbel, 2007])– it could not account for the negative male OLS coefficient. The OLS results are much more likely confounded by a community level effect not adequately controlled for by the X's and district fixed effects. The fact that we saw treatment more likely in poorer places reinforces the idea that the bias in OLS is probably downwards (i.e.

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

negative selection).

7.2 False experiment

My identifying assumption is that gradient only influences employment growth through its indirect effect on electrification. This includes being uncorrelated with underlying differences in unobservable economic growth (positive or negative). If this assumption is valid, we should not only see positive IV estimates of employment growth in period t to $t+1$ but also no relationship between instrumented treatment and employment growth in periods $t+1$ to $t+2$– even though gradient is a relevant cost factor in both periods. If there was such a relationship, this would caution us against interpreting the treatment effect between t and $t+1$ as the causal effect of electrification.

Although there are only two waves of Census data, having administrative data on electricity projects from 1990-2007 is helpful for implementing this false experiment. I check for evidence of a pre-program effect between 1996 and 2001 for areas treated only after 2001. I redefine the treatment dummy = 1 if an area was treated in 2002 and = 0 if never treated or never treated until after 2002. I do not define areas with projects occurring after 2002 as treated. In late 2002, responsibility for electrification passed from Eskom to the Department of Minerals and Energy, hence criteria for project allocation under this new regime was subject to a different set of pressures.

Table 10 presents OLS and IV results for the false treatment variable and the smaller sample. OLS results suggest that areas that are to be treated in future have larger employment growth in the pre-period for men and women (between 1.2% and 1.4%) compared to areas never treated. These are not statistically significantly different from zero. IV results suggest that employment growth during this period in areas was substantially *lower* in areas induced to get the future treatment by gradient. These results provide no evidence that the instrument is only picking up pre-existing trends in employment growth in treated areas. In fact, they suggest some negative selection of areas in which treatment is going to occur because of lower gradient.

7.3 Spill-overs

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 may have positive spill-over effects on neighboring untreated areas. This would dampen estimates of employment growth in response to electrification. Alternatively, if employed or employable individuals move households across space towards treated areas, electrification projects may reduce employment in control areas and hence 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.

To rule out the possibility that large treatment effects are being observed simply because control areas are close to treatment areas, 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. The OLS results are never significantly different from zero while IV coefficients are large, positive and close to the main IV estimate. Areas close to prior electricity projects exhibit significant differences in employment over time and these are slightly larger than the point estimates for areas further away. The fact 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.

The similarity of the point estimates across successive sample restrictions gives us further confidence that labor demand is not the primary channel though which electrification operates. If it were, we would expect to see either increasing (in the case of positive spill-overs when households do not move) or decreasing coefficients (negative spill-overs when households do move) as comparisons are made over ever more distant control areas.

7.4 Migration

Any migration effects of electrification are captured in both OLS and IV coefficients. Recall from Table 5 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 11 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. For men, the bounding procedure reduces the size of the OLS coefficients and increases the IV point estimate somewhat. Male employment is 9.4 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.

With these three specification checks, there is little evidence to suggest that the instrument is picking up differences in employment growth related to unobservables or affected by spatial spillovers between communities. Measurement error only partially explains the difference in OLS and IV coefficients. Since the first stage indicated that poorer places are more likely to be treated, a negative selection of areas could explain more of this gap. In the next section, I further ask whether a local average treatment effect interpretation can reconcile the OLS and IV results and in the process highlight some of the channels through which electrification works to raise female employment growth.

8 Mechanisms

In a heterogenous treatment effects world, IV estimates the local average treatment effect (LATE): the treatment effect on communities induced to be treated by their flatter slope. For these marginal communities, the employment effect may be larger than the average treatment effect. As suggested by the work on home production models, sub-groups most likely to be affected by the new home production technology are those who relied heavily on fuel wood for energy needs, who have fewer other constraints in home production and who have better market opportunities. To probe each of these channels, I investigate whether treatment effects are larger for (i) women living in poorer areas where dependence on wood for fuel is likely higher; (ii) women of different age groups and with different dependency rates of small children; and (iii) women in different skill groups.

8.1 Heterogenous treatment effects related to poverty

Data from the 10% Census micro-data indicates that in rural KZN, 60% of African households earning below 6,000ZAR rely on wood for cooking, while only 40% of households earning above this annual threshold primarily use wood. Although the Census aggregate data does not provide information on community level use of wood, I investigate whether treatment effects are different for communities with different poverty rates. I interact the 1996 poverty rate of the community with the treatment dummy and re-estimate the models in Tables 7 and 8.

Table 13 presents coefficients for baseline poverty rate, the treatment dummy and the interaction. For both OLS and IV results, the interaction term has a positive coefficient. At the mean community poverty rate, female employment is 17 percentage points higher and male employment is 9.8 percentage points higher. Neither are significantly different from zero. The male and female coefficients move in the same direction, but it is not possible to say whether the female interaction coefficient is larger than the male coefficient. Once the interaction term is included, the coefficient on treatment becomes negative for men and women. Differences in responses across very poor and less poor communities can help to explain why the OLS coefficient is close to zero for women and negative for men in the basic specification. When we compare treatment and control areas in the basic OLS model, poverty rates were fairly similar across groups (see Table 3). If employment growth is slower in richer areas and faster in poorer areas, the combined effect would average out in the treatment-control comparison. Employment growth for women is indeed faster in the poorer half of communities (0.001) and slower in the richer half (-0.01). In contrast, IV results estimate the LATE for those groups induced to be treated by virtue of their slope. These areas are mainly poorer areas to begin with conditional on other controls, as shown in the first stage. Communities with low density and further from the grid (i.e. high cost on other variables) are the ones moved around by the gradient measure. These poorer communities are also more likely to respond to access to electricity. They will have been the most constrained in prior access to alternative fuels, relying most heavily on wood for energy needs.

8.2 Age effects and interactions with dependency ratios

Even within poor communities, there is potential heterogeneity in how women respond to the arrival of household electricity. 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 14 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.[37]

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. Employment grows by 3.9 percentage points for women between the ages of 30 and 34, by 2.6 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).[38]

[37] Results for men are suppressed as the treatment coefficient was never significant for any cohort.

[38] 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.

8.3 Effects for different skill groups

To investigate whether household electrification has larger impacts for women with higher market wages, I would ideally like to know something about actual or potential wages. Since the Census does not ask about wages, I use the skill distribution across occupations to proxy for high, medium and low wage jobs. All adults aged 15 to 59 who were classified as employed within the Census are asked to report their current occupation. I collapse occupations into three skill groups: skilled (managers, professionals and associate professionals who represent mainly teachers, nurses and public servants), semi-skilled (service sector workers, crafters, machine operators, skilled and subsistence agriculture) and unskilled (elementary occupations). I create an employment-type to population ratio for each skill-type and re-run the analysis for the three outcomes.

Table 16 provides OLS and IV results for women and men in each skill group. Each cell reports coefficients on the treatment dummy from a separate regression. OLS results are again not statistically different from zero. Skilled and semi-skilled occupations exhibit the largest growth rates in the IV specification. Female semi-skilled work increases by 6.2 percentage points in electrified areas, while skilled work increases by 4.9 percentage points. Unskilled work also increases but by a smaller and not significant percentage. Although male IV coefficients are positive, none of them are different from zero.

That employment for women with higher skill levels is being affected by electrification is consistent with the intuition from a simple model of labor supply and home production. In such a model, women with the same VMP of time in the home who experience the same technology shock will be more likely to shift labor into the market if their market wages are higher.[39]

8.4 Testing further threats to validity

The identification assumption used to justify gradient as a valid instrument not only requires gradient to be directly uncorrelated with labor market trends or shocks but also directly uncorrelated with changes in the demand for labor. The false experiment provided some evidence that gradient is not picking up local labor market trends. Given that we are finding large female employment responses to electrification, a final check to test the validity of the instrument involves asking: does a flatter land gradient make it easier to establish places of work that particularly favor female employment? If so, we might be concerned that gradient directly influences employment growth over time, independent of new electricity.

Recall that women are primarily employed in public sector work, some semi-skilled work and elementary jobs. The *places* of work for each of these groups include (mainly) schools, small scale

[39][Greenwood et al., 2005] assume this when modeling changing female labor force participation rates in response to falling appliance prices in the USA.

manufacturing firms, road sides and tourist venues, and households employing domestic workers. Data on the spatial distribution of businesses over time would be ideal for this test. In the absence of such data, I provide two pieces of indirect evidence that test whether gradient is correlated with establishment of businesses over time.

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.[40] Column (1) of Table 17 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 education. In table 17, I check to see whether gradient (average and modal) is correlated with this change. Column (2) indicates no significant relationship between gradient and the change in potential employer households over time.

Given the constraints of my data, it is not possible to decisively reject that electricity does not enable new firms to open (i.e. does not create new jobs). However, the fact that the instrument is uncorrelated with changes in typical employers bolsters the argument that results are not being driven by places that are growing *because* they are on flatter land and hence easier to establish businesses on.

9 Discussion and conclusion

This paper provides new evidence that rural household electrification can jump-start employment, particularly for women. Using variation in the timing and location of domestic electrification projects in KwaZulu-Natal and instrumenting for treatment to deal with measurement error and endogenous placement of infrastructure, I show that female employment increases by a statistically significant 13.5 percentage points (bounded below at 5 and above at 45 percentage points). Male employment rates are not statistically significantly different from zero. I use project data after 2001 to show that areas to be treated in future are not experiencing faster employment growth in the

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

period before treatment. This lends support to the assumption that the instrument does deal with selection on unobservable time trends.

One of the contributions of this paper is to use several new data sets in combination with spatially matched Census data to identify areas that experience infrastructure expansion. This allows me to separate access to electricity from use of electricity, which is something previous studies have been unable to do.

However, even using project data to define treatment does not allow me to separate whether electrification affects labor supply or labor demand. Within the limits of my data, I cannot decisively rule out that electrification did not directly create jobs. For example, electricity may facilitate craft work at night or small shops and pubs operated from inside the home. Given that the South African roll-out supplied small, non-commercial amounts of electricity, and given the absence of spill-over effects across space and lack of correlation between the instrument and key sources of female labor demand, the effects of electrification may be plausibly interpreted as a net increase in labor supply.

The large female employment results are concentrated in poorer areas. Communities with higher poverty rates are both more likely to receive an Eskom electrification project as evidenced in the first stage, and exhibit a larger response upon electrification. This is most likely because poorer areas are heavily reliant on time consuming wood collection and wood cooking, as the Census micro data suggest. Within these areas, women in their thirties and in skilled and semi-skilled groups experience larger employment effects.

One caveat to interpreting my results is that electrification may have general equilibrium effects. Since domestic electrification has more features of a private than a public good, the extent of household production spill-overs across households is limited. The largest GE effect we might expect is through migration in to (or slower migration out of) areas that become more attractive to live in. Population is indeed growing faster in treated compared to control areas. However, a simple bounding exercise illustrates that even if all in-migrants are counted as employed, their presence could not explain the majority of the female employment effect.

Rural electrification is unlikely to be the solution to South Africa's large unemployment problem.[41] First, the electrification roll-out occurred in very rural parts of the country where reliance on time-consuming fuel wood was high. In other, less rural parts of South Africa, alternative modern fuels (kerosene, natural gas) have undoubtedly already relaxed the time burden of home production on women. Second, although effects are large percentage wise, they do not translate into a large number of newly employed, simply because populations in these rural areas are very small. To roughly estimate how many newly employed women resulted from this rural electrification, we can add up 13.5% of the 1996 female African population in each treated community. Imposing the extreme assumption that employment response is the same in all treated areas, this would imply

[41] South Africa has broad unemployment rates in excess of 30%.

an increase of 22,487 workers, which is only 1.1% of the total increase in new jobs calculated by Posel and Casale (2003) for this period across the country. What my results do indicate is that there is a role for rural electrification to play in changing the nature of rural labor markets in ways that differentially affect women. Employment effects of public infrastructure may be as important in helping households escape from poverty as the more typical effects on health and education.

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.

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

[Becker, 1965] Becker, G. (1965). A theory of the allocation of time. *Economic Journal*, (299).

[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, 2001] Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4).

[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.

[Lewbel, 2007] Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica*, 75:537–551.

[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*.

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

[Udry, 2000] Udry, C. (2000). *Gender, Agricultural production and the theory of the household*, chapter 4. MIT Press.

[Willis, 1973] Willis, R. (1973). A new approach to the economic theory of fertility behavior. *Journal of Political Economy*, 81(2):S14–S64.

10 Data Appendix

10.1 KIDS data 1993 and 1998

Panel data from the KwaZulu-Natal Income Dynamics Study in 1993 and 1998 available at [http://sds.ukzn.
A survey of 713 households and 5923 African individuals followed over time in rural areas of the province. The entire study covered African and Indian individuals in urban and rural areas too, but I exclude urban areas and Indians from my analysis. Eighty percent of the households and individuals in the rural sample were found in the same household and cluster in 1993 and 1998 (May et al, 1999).

10.2 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.

10.3 Census panel of communities

As in most countries, boundaries in South Africa have shifted over time.[42] 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".[43] 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".[44]

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.

[42] [Christopher, 2001]
[43] [Africa,]
[44] [Africa,]: pages 21, 26.

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 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.

10.4 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.

10.5 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.

10.6 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 (eventually).

Table 1: Characteristics of KIDS households without a grid connection in 1993

1993 Characteristics	Connected to the grid by 1998	Not connected to the grid by 1998	Difference	P-value
Anyone fetches wood?	0.353 (0.479)	0.662 (0.474)	-0.309	0.000
Minutes per week spent collecting fuel wood (incl. zeros)	300.030 (864.352)	607.782 (848.256)	-307.752	0.000
Minutes per week spent collecting fuel wood (excl. zeros)	859.407 (1294.219)	918.580 (895.803)	-59.173	0.692
Prop wood collectors female	0.961 (0.017)	0.956 (0.010)	0.005	0.811
Household per capita income (Rands)	158.378 (10.320)	122.452 (6.872)	35.926	0.002
N	170	266		

Table 1: Sample is rural African households captured in 1993 KIDS data and found at follow up in 1998. Per capita income is measured in Rands.

Table 2: Appliance uptake and employment - KIDS fixed effects

	Does anyone own a:			Employment	
	Kettle?	Fridge?	Stove?	Male	Female
	(1)	(2)	(3)	(4)	(5)
Year	0.054	0.009	0.376	0.012	0.011
	(0.174)	(0.205)	(0.244)	(0.040)	(0.039)
Connected to grid∗Year	0.372	0.344	0.065	-0.061	0.042
	(0.045)	(0.060)	(0.053)	(0.063)	(0.046)
N observations	872	872	872	1780	2340
N groups	436	436	436	500	541
R^2	0.363	0.254	0.021	0.178	0.098

Table 2: Data are from KwaZulu-Natal Income Dynamics Study 1993 and 1998, sample of African rural households interviewed in both years in the same cluster. Household fixed effects models for household j in time period $t = 1993, 1998$ estimated by OLS for each binary outcome variable y_{jt}. Additional control variables measured in each year include: female head of household, household size, number of children under age 6, any male or female pensioner. Also included is the 1993 household monthly expenditure and the interaction of this variable with year. Robust standard errors in brackets, clustered at the community level.

Table 3: Baseline X variables by treatment status

X-variables: 1996	Sample mean	Treatment Mean	Control Mean	Δ_{T-C}	p-value
Poverty rate	0.607	0.590	0.611	-0.021	0.057
	(0.194)	(0.169)	(0.200)		
Female headed households	0.550	0.547	0.551	-0.004	0.575
	(0.127)	(0.120)	(0.129)		
Adult sex ratio (f/m)	1.475	1.409	1.491	-0.081	0.000
	(0.289)	(0.249)	(0.296)		
Proportion Indian, white adults	0.002	0.001	0.003	-0.002	0.190
	(0.031)	(0.003)	(0.034)		
Kilometers to road	38.335	36.065	38.889	-2.824	0.042
	(24.613)	(24.103)	(24.712)		
Kilometers to town	39.028	36.796	39.572	-2.776	0.007
	(18.298)	(15.319)	(18.918)		
Adult men with high school	0.063	0.076	0.060	0.016	0.000
	(0.046)	(0.048)	(0.045)		
Adult women with high school	0.068	0.085	0.064	0.021	0.000
	(0.052)	(0.056)	(0.050)		
Variables affecting cost					
Household density	20.660	30.756	18.198	12.559	0.000
	(29.488)	(48.146)	(22.057)		
Kilometers from grid	19.348	15.681	20.243	-4.562	0.000
	(13.476)	(9.964)	(14.059)		
Land gradient - mean	22.255	20.326	22.725	-2.400	0.000
	(9.903)	(8.558)	(10.151)		
Gradient - mode	19.103	16.606	19.712	-3.106	0.000
	(12.762)	(10.546)	(13.177)		
Gradient - std. dev.	10.840	10.316	10.967	-0.651	0.003
	(3.924)	(3.824)	(3.938)		
Gradient - range	52.502	50.263	53.047	-2.784	0.002
	(15.763)	(15.337)	(15.822)		
Gradient - median	21.335	19.153	21.867	-2.714	0.000
	(10.594)	(8.906)	(10.903)		
N communities	1994	391	1603		

Table 3: 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.

Table 4: Treatment and changes in access to basic services: unadjusted difference in differences

Outcomes	Year	All	Treatment	Control	Δ_{T-C}	p-value
Electric lighting	1996	0.088	0.133	0.077	0.06	0.00
		(0.201)	(0.219)	(0.195)		
	2001	0.172	0.424	0.111	0.31	0.00
		(0.295)	(0.361)	(0.239)		
	Δ_t	0.08	0.29	0.03	**0.26**	0.00
Water nearby	1996	0.164	0.284	0.135	0.15	0.00
		(0.257)	(0.313)	(0.233)		
	2001	0.169	0.265	0.145	0.12	0.00
		(0.246)	(0.282)	(0.231)		
	Δ_t	0.01	-0.02	0.01	**-0.03**	0.05
Flush toilet	1996	0.014	0.015	0.014	0.00	0.77
		(0.066)	(0.060)	(0.067)		
	2001	0.042	0.045	0.041	0.00	0.24
		(0.061)	(0.074)	(0.058)		
	Δ_t	0.03	0.03	0.03	**0.00**	0.62
Telephone in household	1996	0.008	0.010	0.007	0.00	0.04
		(0.023)	(0.023)	(0.023)		
	2001	0.142	0.169	0.135	0.03	0.00
		(0.109)	(0.104)	(0.109)		
	Δ_t	0.13	0.16	0.13	**0.03**	0.00
		1994	391	1603		

Table 4: Standard deviations in brackets below. Community sample consists of sub-places in former KwaZulu tribal areas. All means are calculated over communities. Treatment is one if the first Eskom project occurred between 1996 and 2001, otherwise 0. Excluded from the sample are areas first treated prior to 1996. Service access is calculated as follows: (1) proportion of households with electricity as main source of lighting; (2) proportion of households with a water source in the house or within 200 meters of house; (3) proportion of households with a flush toilet; (4) proportion of households with a phone (landline or cellphone) in the house.

Table 5: Means of community level outcomes by treatment status

Outcomes	Year	All	Treatment	Control	Δ_{T-C}	p-value
Female employment	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 employment	1996	0.137	0.162	0.130	0.032	0.00
		(0.118)	(0.115)	(0.118)		
	2001	0.100	0.111	0.098	0.013	0.02
		(0.099)	(0.089)	(0.101)		
	Δ_t	-0.036	-0.051	-0.032	-0.019	0.00
Log population	1996	6.777	6.924	6.741	0.184	0.001
		(0.977)	(1.041)	(0.957)		
	2001	6.964	7.277	6.887	0.389	0.000
		(0.846)	(0.772)	(0.846)		
	Δ_t	0.187	0.353	0.147	0.206	0.000
Female in-migrants (N)	1996	9.189	14.243	7.956	6.287	0.00
		(21.130)	(24.271)	(20.107)		
	2001	6.895	13.309	5.330	7.979	0.00
		(20.397)	(33.348)	(15.308)		
	Δ_t	-2.294	-0.934	-2.626	1.693	0.00
Male in-migrants (N)	1996	5.737	9.268	4.876	4.392	0.00
		(14.823)	(17.649)	(13.919)		
	2001	5.281	10.240	4.072	6.169	0.00
		(16.132)	(27.458)	(11.521)		
	Δ_t	-0.456	0.972	-0.804	1.776	0.00
N		1994	391	1603		

Table 5: 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 6: Assignment to treatment first stage OLS

	Treatment variable measured as:					
		Dummy			Year	Fraction
	(1)	(2)	(3)	(4)	(5)	(6)
Gradient*10	-0.0386	-0.0386	-0.0405	-0.0400	-0.120	-0.022
	(0.020)	(0.017)	(0.013)	(0.013)	(0.054)	(0.009)
Distance from grid*10		-0.047	-0.020	-0.020	-0.009	-0.017
		(0.020)	(0.022)	(0.022)	(0.065)	(0.012)
HH density*10		0.017	0.012	0.013	0.049	0.005
		(0.004)	(0.005)	(0.005)	(0.018)	(0.004)
Poverty rate		0.033	0.032	0.027	0.012	0.046
		(0.066)	(0.067)	(0.066)	(0.212)	(0.044)
Adult sex ratio (f/m)		0.353	0.136	0.126	-0.076	0.056
		(0.118)	(0.104)	(0.104)	(0.365)	(0.075)
Proportion female headed hh's		-0.163	-0.115	-0.109	-0.338	-0.028
		(0.048)	(0.038)	(0.038)	(0.143)	(0.025)
Proportion Indian/white adults		-0.444	-0.367	-0.389	-1.386	-0.196
		(0.161)	(0.140)	(0.128)	(0.390)	(0.086)
Distance to road*10		0.005	-0.006	-0.006	0.021	-0.003
		(0.008)	(0.009)	(0.009)	(0.034)	(0.005)
Distance to town*10		0.017	0.010	0.009	-0.045	0.002
		(0.014)	(0.015)	(0.015)	(0.047)	(0.008)
Adult men with high school		-0.031	0.101	0.096	-0.238	0.278
		(0.453)	(0.401)	(0.389)	(1.256)	(0.222)
Adult women with high school		0.834	0.724	0.782	2.492	0.218
		(0.422)	(0.397)	(0.376)	(1.284)	(0.210)
Δ hh's with water close by				0.015	-0.266	0.020
				(0.045)	(0.159)	(0.041)
Δ hh's with flush toilets				0.182	0.943	0.085
				(0.086)	(0.381)	(0.047)
District FE	No	No	Yes	Yes	Yes	Yes
Obs	1994	1994	1994	1994	1994	1994
R squared	0.009	0.075	0.167	0.169	0.150	0.199
F-stat: gradient	3.590	5.440	9.180	8.980	5.000	6.000
Prob>F:	0.060	0.020	0.000	0.000	0.030	0.010

Table 6: 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 7: Employment effects for women - OLS and IV

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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)
Distance to grid *10		0.004	0.004	0.004		0.008	0.006	0.007
		(0.003)	(0.003)	(0.003)		(0.004)	(0.004)	(0.004)
HH density *10		0.000	0.000	0.000		-0.001	-0.002	-0.002
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
Poverty rate		0.032	0.035	0.031		0.028	0.031	0.028
		(0.011)	(0.011)	(0.011)		(0.013)	(0.016)	(0.015)
Proportion female headed households		0.036	0.039	0.034		0.008	0.022	0.019
		(0.023)	(0.023)	(0.023)		(0.032)	(0.030)	(0.030)
Adult sex ratio (F/M)		0.020	0.020	0.024		0.036	0.038	0.040
		(0.010)	(0.010)	(0.009)		(0.015)	(0.013)	(0.013)
Indian/white adults		-0.495	-0.485	-0.482		-0.433	-0.413	-0.410
		(0.269)	(0.269)	(0.256)		(0.271)	(0.263)	(0.255)
Distance to road*10		0.001	0.001	0.001		0.000	0.001	0.002
		(0.001)	(0.001)	(0.001)		(0.001)	(0.002)	(0.002)
Distance to town*10		-0.004	-0.003	-0.004		-0.005	-0.004	-0.005
		(0.002)	(0.002)	(0.002)		(0.003)	(0.003)	(0.003)
Adult men with high school		0.150	0.161	0.159		0.146	0.139	0.137
		(0.104)	(0.104)	(0.092)		(0.102)	(0.101)	(0.094)
Adult women with high school		-0.180	-0.195	-0.153		-0.257	-0.290	-0.257
		(0.115)	(0.116)	(0.100)		(0.120)	(0.114)	(0.108)
Δ proportion with water close by				0.028				0.026
				(0.008)				(0.010)
Δ proportion with flush toilet				0.111				0.085
				(0.058)				(0.058)
District FE?	N	N	Y	Y	N	N	Y	Y
Additional variables?	N	N	Y	Y	N	N	Y	Y
N	1992	1992	1992	1992	1992	1992	1992	1992
R^2	0.000	0.067	0.075	0.100				
Standard confidence interval							[0.02-0.26]	[0.015 - 0.28]
AR confidence interval							[0.05-0.45]	[0.05 - 0.45]

Table 7: 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 8: Employment effects for men - OLS and IV

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.019	-0.015	-0.009	-0.010	-0.027	0.057	0.039	0.039
	(0.008)	(0.006)	(0.006)	(0.006)	(0.085)	(0.081)	(0.068)	(0.068)
Kms to grid *10		0.010	0.007	0.007		0.013	0.008	0.008
		(0.004)	(0.004)	(0.004)		(0.005)	(0.005)	(0.004)
HH density *10		0.002	0.002	0.002		0.000	0.001	0.001
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Poverty rate		0.064	0.066	0.062		0.061	0.065	0.061
		(0.018)	(0.017)	(0.017)		(0.021)	(0.018)	(0.018)
Female-head hh		0.242	0.249	0.245		0.219	0.243	0.240
		(0.032)	(0.034)	(0.033)		(0.041)	(0.037)	(0.036)
Adult sex ratio (F/M)		0.004	0.001	0.003		0.017	0.007	0.009
		(0.011)	(0.011)	(0.012)		(0.019)	(0.015)	(0.016)
Indian/white adults		0.530	0.543	0.532		0.560	0.559	0.550
		(0.318)	(0.313)	(0.307)		(0.313)	(0.310)	(0.304)
Log kms to road		0.000	0.000	0.000		0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Log kms to town		-0.001	0.000	0.000		-0.001	0.000	0.000
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Adult men with h/s				-0.117				-0.125
				(0.122)				(0.130)
Adult women with h/s				0.134				0.097
				(0.128)				(0.146)
Δ proportion with water close by				0.027				0.027
				(0.009)				(0.009)
Δ proportion with flush toilet				0.093				0.083
				(0.075)				(0.074)
District FE?	N	N	Y	Y	N	N	Y	Y
N	1993	1993	1993	1993	1993	1993	1993	1993
R-squared	0.005	0.162	0.179	0.191				
C.I.							[-0.09-0.18]	[-0.09 - 0.2]
AR C.I.							[-0.05-0.25]	[-0.05 - 0.25]

Table 8: 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 9: Assessing measurement error in the treatment variable

	Smaller communities: < X households?				Fraction connected >0.8		Δ electricity > 0.1	
Type of restriction:	<300	<300	<100	<100				
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Treatment coefficient	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
for Female employment	0.001	0.206	0.018	0.319	0.012	0.136	0.010	0.126
	(0.005)	(0.102)	(0.010)	(0.179)	(0.009)	(0.092)	(0.007)	(0.065)
N	1622	1622	625	625	1420	1420	1620	1620
R^2	0.105		0.206		0.109		0.188	
for Male employment	-0.011	0.073	-0.017	0.051	-0.011	0.008	-0.003	0.061
	(0.008)	(0.090)	(0.011)	(0.145)	(0.013)	(0.112)	(0.009)	(0.084)
N	1623	1623	626	626	1421	1421	1620	1620
R^2	0.186		0.202		0.191		0.188	

Table 9: Notes: 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. Treatment= 1 if community had first Eskom project between 1996 and 2001. All other variables (except treatment) are measured in 1996. Columns (1)-(4) restrict to successively smaller community sizes, columns (5) and (6) exclude communities where less than 80% of the community was treated, and the last two columns exclude all communities that did not have at least a 10% increase in the proportion of households with access to electric lighting.

Table 10: False experiment

X variables	Δ Female employment OLS (1)	IV (2)	Δ Male employment OLS (3)	IV (4)
False treatment variable	0.011	-1.334	0.013	-0.641
	(0.009)	(5.779)	(0.012)	(3.224)
Distance to substation 1996*10	0.006	0.004	0.011	0.010
	(0.005)	(0.025)	(0.005)	(0.013)
Household density 1996*10	-0.001	0.015	-0.001	0.006
	(0.002)	(0.071)	(0.002)	(0.040)
Poverty rate 1996	0.023	0.024	0.039	0.040
	(0.014)	(0.061)	(0.024)	(0.036)
District FE?	Y	Y	Y	Y
Additional variables?	Y	Y	Y	Y
N	1095	1095	1096	1096
R^2	0.122		0.190	

Table 10: Robust standard errors in parentheses, clustered at main place level. False treatment variable= 1 if community had Eskom project in 2002, = 0 if not. Sample is restricted to those areas that had projects in 2002 or not at all. 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 treatment variable and change in other services; each regression includes a full set of district fixed effects. The excluded instrument is average community land gradient (measured in deviation form). F-statistic for gradient in the first stage is 0.08

Table 11: Restricting to control areas further from prior treatment communities

Coeff. on treatment dummy for women	(1) OLS	(2) OLS	(3) IV	(4) IV
Full sample	-0.004	0.001	0.045	**0.135**
N=1992	(0.005)	(0.005)	(0.055)	**(0.062)**
Control areas >1km from an area	-0.009	-0.005	-0.002	0.104
treated before 1996	(0.005)	(0.006)	(0.042)	(0.061)
N=1656				
Control areas >5km from an area	-0.008	-0.005	0.079	0.114
treated before 1996	(0.006)	(0.008)	(1.294)	(0.097)
N=1374				
Additional variables?	N	Y	N	Y

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 1996.

Table 12: Bounding results for women and men

	Women				Men			
variables	OLS		IV		OLS		IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment=T	0.000	0.000	0.110	0.135	-0.013	-0.010	0.106	0.083
	(0.005)	(0.005)	(0.068)	(0.071)	(0.007)	(0.006)	(0.090)	(0.069)
Distance to grid 1996*10		0.002		0.005		0.004		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)		(0.001)		(0.002)
Poverty rate 1996		0.024		0.021		0.050		0.048
		(0.009)		(0.014)		(0.014)		(0.016)
N	1992	1992	1992	1992	1993	1993	1993	1993
R-squared	0.000	0.025			0.003	0.175		
Standard C.I.				[0-0.27]				[-0.05-0.22]
AR confidence interval				[0.05-0.4]				[0.05-0.25]

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 8: Heterogenous treatment effects related to poverty

	(1)	(2)	(3)	(4)
X variables in 1996	Women: OLS	Women: IV	Men: OLS	Men: IV
Treatment=T	-0.027	-0.052	-0.049	-0.069
	(0.016)	(0.304)	(0.024)	(0.412)
Treatment*poverty rate	0.048	0.280	0.065	0.161
	(0.024)	(0.444)	(0.034)	(0.592)
Poverty rate 1996	0.025	-0.007	0.054	0.041
	(0.011)	(0.056)	(0.019)	(0.076)
District FE?	Y	Y	Y	Y
Additional variables?	Y	Y	Y	Y
Interaction effect at mean poverty rate	0.029	0.170	0.039	0.098
Main + interaction effect	0.002	0.118	-0.010	0.029
pval	0.71	0.07	0.11	0.722
N	1992	1992	1993	1993
R-squared	0.102		0.192	

Table 13: Notes: Robust standard errors in parentheses, clustered at main place level. Outcome variable is change in proportion of African females/males aged 15-59 who are employed. 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, number 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 instruments are the average land gradient and the interaction of average land gradient and poverty rate in 1996. To calculate the additional effect of the treatment in poorer areas, I multiply the interaction coefficient by the mean poverty rate in the sample (0.75).

Table 14: 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	0.005	0.000	0.009	-0.001	0.020	0.000	**0.039**	0.001	**0.026**	0.001	0.013	0.001	**0.019**	-0.001	0.003	0.000	0.004
	(0.000)	(0.007)	(0.001)	(0.014)	(0.001)	(0.014)	(0.001)	**(0.018)**	(0.001)	**(0.014)**	(0.001)	(0.012)	(0.001)	**(0.009)**	(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		0.076		0.053		0.028	

Table 14: 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 15: Differences in occupations by gradient in control areas only - OLS

	OLS
Outcome variable in 1996	
Proportion of skilled men	-0.001
	(0.000)
Proportion of semi skilled men	-0.012
	(0.002)
Proportion of unskilled men	-0.002
	(0.001)
Proportion of skilled women	-0.001
	(0.000)
Proportion of semi skilled women	-0.002
	(0.001)
Proportion of unskilled women	-0.004
	(0.001)
N	1603

Table 15: Differences in 1996 means for communities by gradient*10. Outcome variables are the proportion of adults aged 15-59 who report occupation as skilled (including managers, teaching and life sciences professionals and associate professionals), semi-skilled (services, crafters, machine operators, clerks, skilled agriculture) and unskilled (subsistence agriculture, elementary workers).

Table 16: Results for occupations of different skills - OLS and IV

	(1)	(2)	(3)	(4)
	Women		Men	
Δ prop. people in each occupation	OLS	IV	OLS	IV
Skilled occupations	0.000	**0.049**	-0.001	**0.040**
	(0.001)	**(0.023)**	(0.001)	**(0.021)**
Semi-skilled occupations	0.002	**0.062**	-0.004	0.041
	(0.002)	**(0.028)**	(0.004)	(0.044)
Unskilled occupations	-0.001	0.027	-0.004	0.026
	(0.003)	(0.029)	(0.003)	(0.034)
N	1992	1992	1993	1992

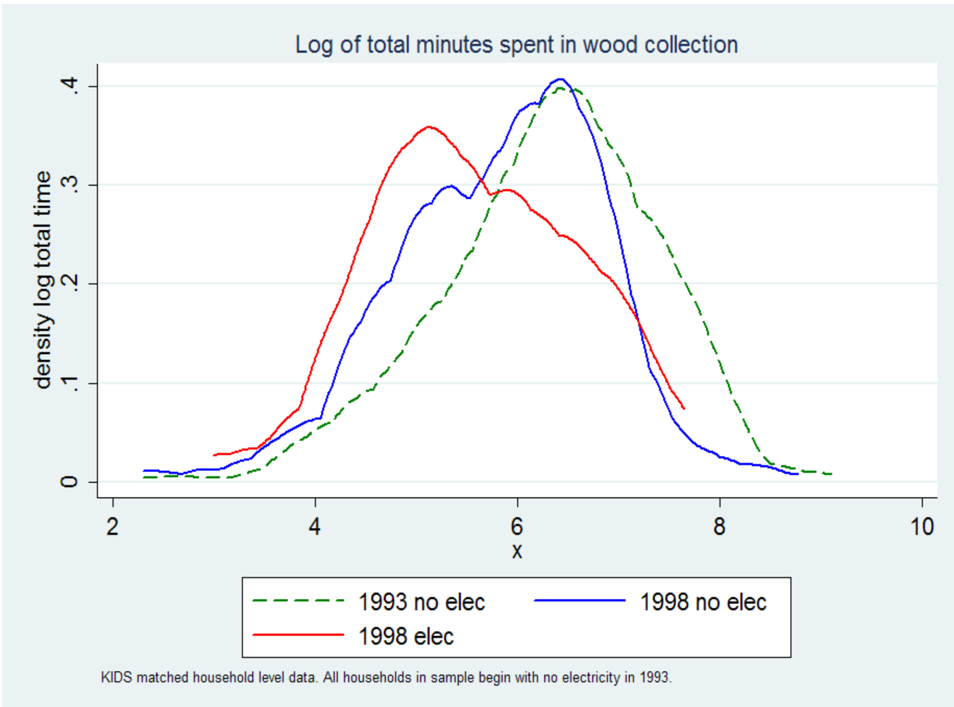
Table 16: Robust standard errors in brackets, clustered at main place level. Each cell presents the coefficient on treatment from a separate regression. Each regression contains the same set of control variables as the full model. Treatment=1 if treated between 1996 and 2001 for the first time; otherwise zero. Outcome variables are measured as proportion of all African adults (15-59) who work in occupation type Y.

Table 17: 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	1994	1994

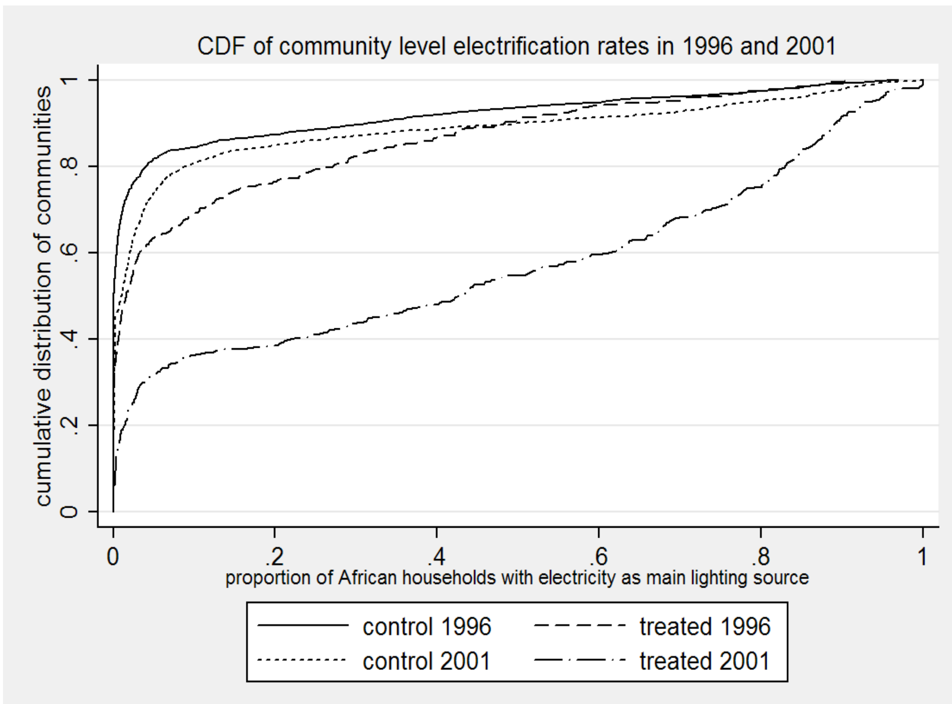
Table 17: Robust standard errors in brackets, clustered at main place level. Outcome variable is the change in the number of schools in the community (range=0 to 5), and the change in the proportion of Indian and white adults ages 20 and over with at least grade 8 education. All other variables also included.

Figure 1



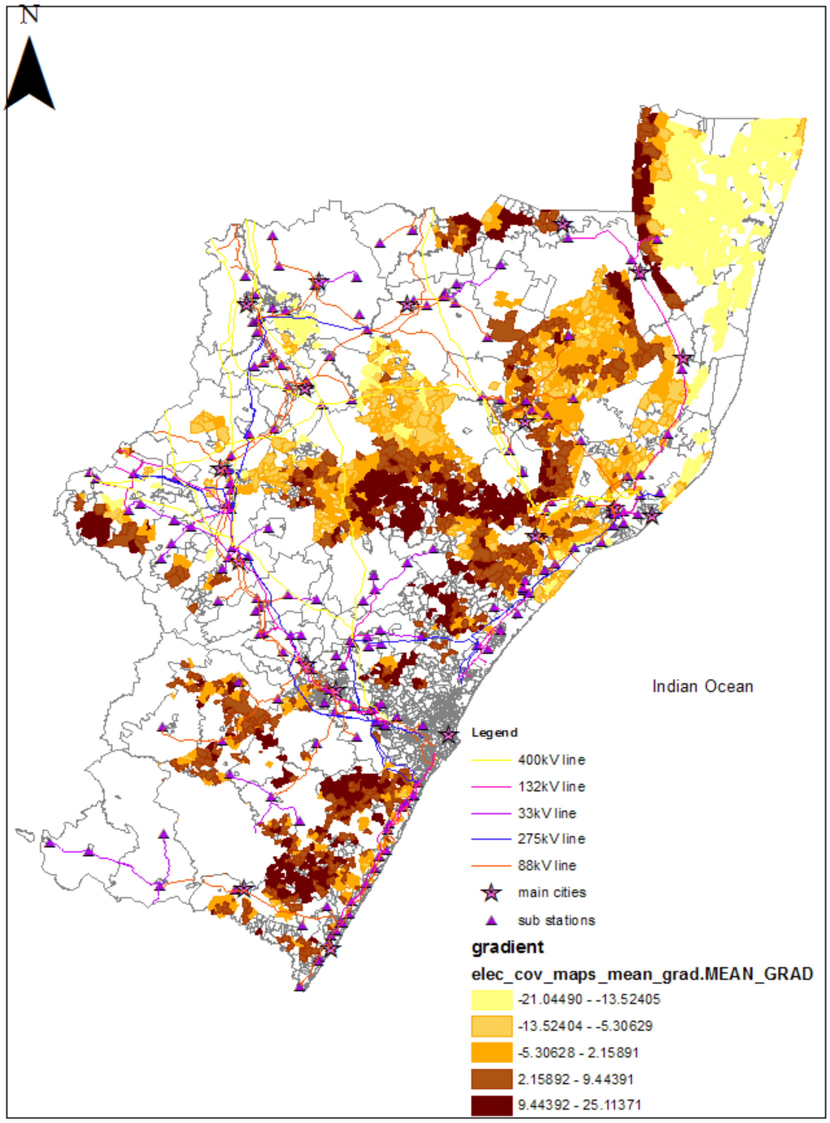
Notes: Data are from KIDS.

Figure 2



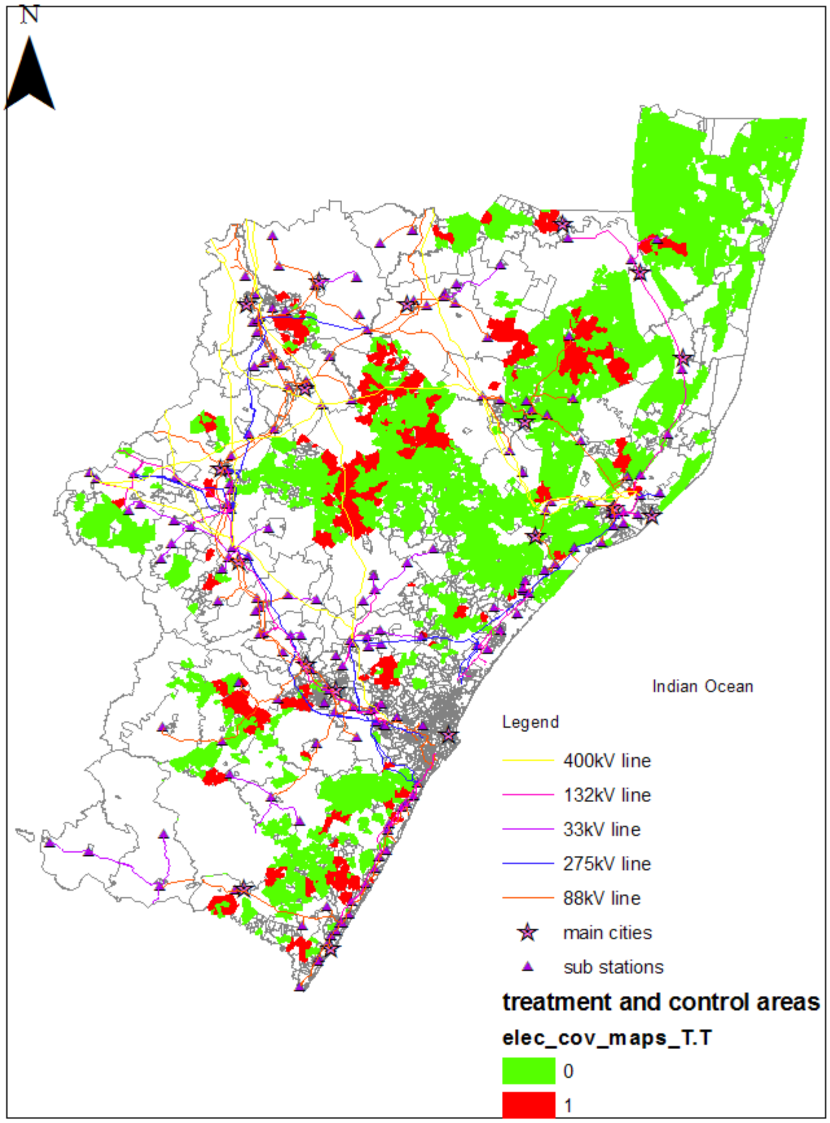
Notes: Empirical cumulative distribution functions of the proportion of households with electricity as main source of lighting in Census 1996 and 2001. Treatment group is the set of communities that had an Eskom electrification project between 1996 and 2001; control group is the set of communities that 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 3



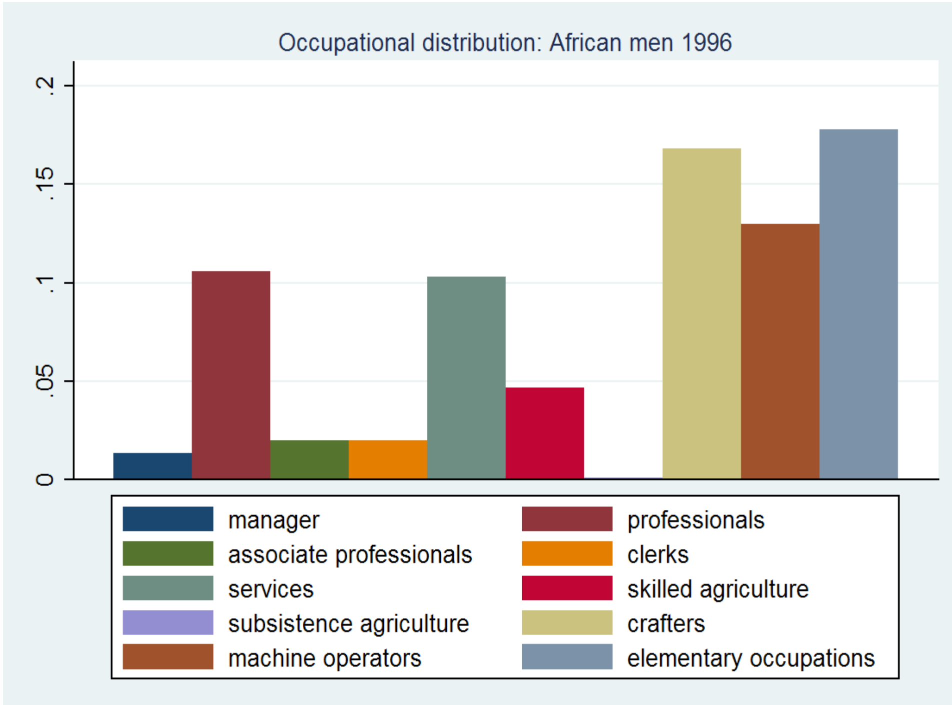
Notes: Map of KwaZulu-Natal region. Shaded areas are in the sample. Areas of steeper average gradient are in darker brown shading; areas of flatter average gradient are pale yellow. Lines represent electricity grid lines in 1996, triangles are electricity substations in 1996 and stars represent towns.

Figure 4



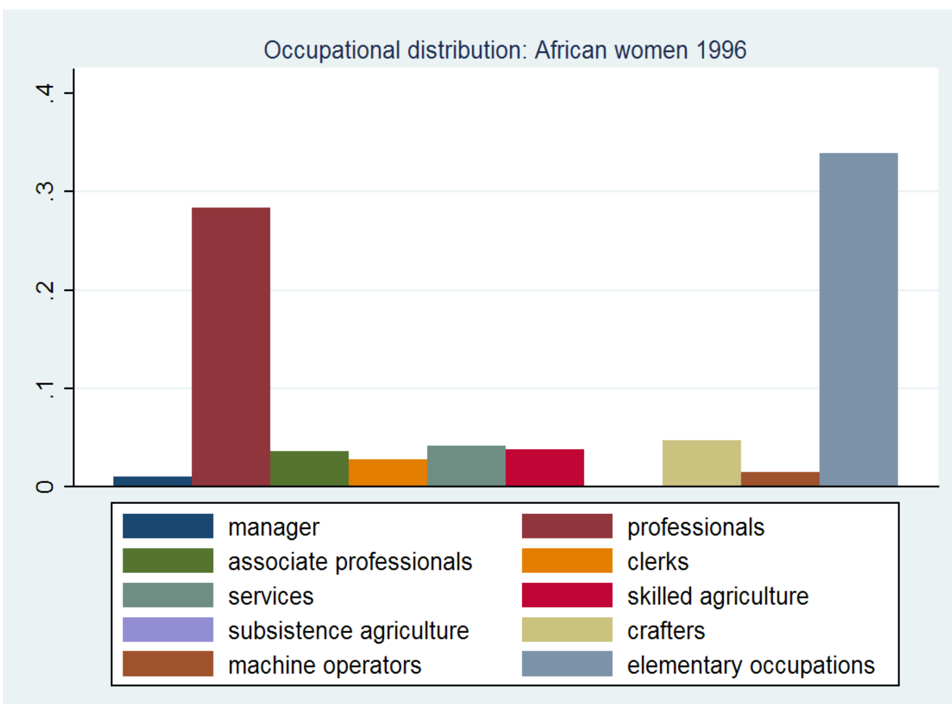
Notes: Map of KwaZulu-Natal region. Shaded areas are in the sample: red areas are treated with an Eskom project between 1996 and 2001, green areas are treated after 2001 or not at all. Lines represent electricity grid lines in 1996, triangles are electricity substations in 1996 and stars represent towns.

Figure 5a



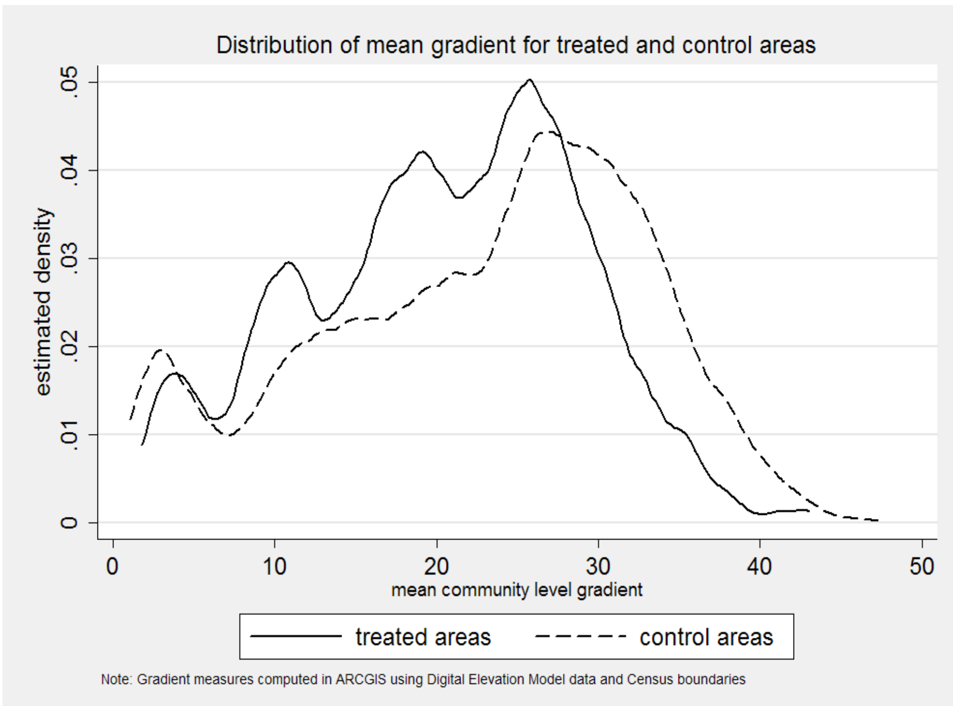
Notes: Distribution of occupation groups for men in Census 1996. Groups are defined as the number of men employed in each occupation over all employed men.

Figure 6b



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.

Figure 6


Distribution of mean gradient for treated and control areas
estimated density
0
.01
.02
.03
.04
.05
0
10
20
30
40
50
mean community level gradient
treated areas
control areas
Note: Gradient measures computed in ARCGIS using Digital Elevation Model data and Census boundaries