The Effects of Rural Electrification on Employment: New Evidence from South Africa

Taryn Dinkelman*
Princeton University

This version: August 2010 First version: November 2007

Abstract

This paper estimates the impact of electrification on employment growth by analyzing South Africa's mass roll-out of electricity to rural households. Using several new data sources and two different identification strategies (an instrumental variables strategy and a fixed effects approach), I find that electrification significantly raises female employment within 5 years. This new infrastructure appears to increase hours of work for men and women, while reducing female wages and increasing male earnings. Several pieces of evidence suggest that household electrification raises employment by releasing women from home production and enabling micro-enterprizes. Migration behavior may also be affected.

*Email: tdinkelm@princeton.edu.

Electrification data and institutional background and interviews provided by Jenny Barnard, Sheila Brown, Ed Bunge, Trevor Gaunt (University of Cape Town), Innocent Nxele, Gisela Prasad (Energy Research Development Council at UCT), Jim Stephenson, Steven Tait and Amos Zuma, in May 2006 and 2007. Conversations with and comments from Lori Beaman, Diether Beuermann, Matias Busso, Anne Case, HwaJung Choi, John DiNardo, Ann Ferris, Raymond Fisman, Ben Keyes, David Lam, Jim Levinsohn, Claudia Martinez-A, Justin McCrary, Nancy Qian, Dori Posel, Samara Potter-Gunter, Mark Rosenzweig, Jeff Smith, Gary Solon, James Tybout, Rebecca Thornton and Martin Wittenberg substantially improved the paper. I also thank seminar participants at the following universities, organizations and conferences for their valuable input: Brown, California-Los Angeles, California-Riverside Dartmouth, Duke, the Harris School, the Harvard School of Public Health, the IDB, the International Institute of Economic Studies at Stockholm University, the Kennedy School, KwaZulu-Natal, London School of Economics, Michigan, Michigan State, the Microeconomics of Growth Network, Northwestern, North Carolina-Chapel Hill, Penn State, Princeton, Resources for the Future, Syracuse, Washington-St Louis, Western Ontario, Virginia and Yale as well as suggestions of the editor and referees. Data collection was funded through a National Science Foundation Doctoral Dissertation Grant and an Eva Mueller New Directions in Demography research award. I also acknowledge support from the Fogarty International Center of the U.S. National Institutes of Health (D43TW000657) and the Population Studies Center at the University of Michigan. All remaining errors are my own.

Electricity is pervasive in all industrialized countries and largely absent in the developing world: about 1.6 billion people world-wide lack access to electricity (Saghir, 2005). Even though many would consider electricity to be a "marker" for development, and despite several historical episodes of wide-spread electrification in developed countries (for example, the rural electrification of America in the 1930s), we know little about the direct effects that new access to modern energy infrastructure will have on the process of development.

The primary objective of this paper is to analyze the impact of new access to modern energy on an outcome of considerable interest: the ability of the poor to use their labor resources for market production. In this paper, I estimate the causal impact of household electrification on employment growth in rural communities by analyzing rural electrification roll-out in post-apartheid South Africa. As a second objective, I investigate the mechanisms through which this new infrastructure affects rural labor markets. Since energy infrastructure is likely to expand in poor areas over the next few decades[1], this analysis provides important lessons for many countries as well as for researchers studying the changing nature of developing country labor markets.

The roll-out of grid infrastructure in South Africa provides a particularly good opportunity to evaluate the effects of electrification on market employment. It was rapid, extended into rural areas and targeted low capacity household use rather than industrial users (Gaunt, 2003). In 1993, a year before the end of apartheid, over two-thirds of South African households were without electricity and more than 80% relied on wood for home production.[2] Following the new government's commitment to universal electrification, 2 million households, or almost one quarter of all households across the country, were newly connected to the grid by 2001. This is twice as many households as the number of US farms connected during the first five years of Roosevelt's Rural Electrification Act (Beall, 1940).

Evaluating the effects of this electrification, or of any infrastructure roll-out, is not straightforward. A large literature on the relationship between infrastructure and economic growth acknowledges that infrastructure could be targeted towards growing areas, or towards politically important areas.[3] Such selection biases any comparison of electrified and

[1]World Bank commitments to energy infrastructure in Africa rose from $447 million in 2001 to $790 million in 2007. The World Bank's Lighting Africa initiative aims to provide 250 million Africans with modern sources of energy by 2030.

[2]Charmes (2005) and Saghir (2005) document the time intensity of home production in developing countries. South Africans (mainly women) spend on average two working days per week in fuel-wood collection (Budlender et al, 2002) and rural households spend an average of 3 hours per day on food preparation (own calculations, 1997 October Household Survey).

[3]The tradition in the macroeconomics literature has been to estimate the effects of public infrastructure on total factor productivity using time-series data. Aschauer (1989) is a classic reference; see Canning (1998) for cross-country evidence and Bogotic and Fedderke (2006) for South African evidence. The World Bank (1994) and Jimenez (1995) provide good overviews of the infrastructure literature relevant for developing

non-electrified areas, and in unpredictable ways. Confounding trends in the economy make it even more difficult to tease out the effects of infrastructure on any economic outcomes.

In this paper, I use two empirical strategies to identify the impact of electricity, taking into account endogenous project placement and confounding economic trends. In the main approach, I estimate community-level employment growth rates in communities that do and do not receive an electricity project between 1996 and 2001, instrumenting for project placement. To do this, I collect and match administrative data on roll-out in rural KwaZulu-Natal (KZN) with geographical data and two Census surveys. I use land gradient to generate exogenous variation in electricity project allocation to communities. Higher gradient raises the average cost of a household connection, making gradient an important factor in prioritizing areas for electrification. I argue and provide evidence from a placebo experiment that in the case of rural KZN, an area with poor agricultural prospects, gradient is unlikely to directly affect employment outcomes, conditional on covariates.

As a complement to the main analysis, I use a fixed effects strategy to estimate the impact of electrification on a richer set of labor market outcomes: employment, hours of work, wages and earnings. For this analysis, I construct a four-period panel of magisterial districts (agglomerations of communities) from cross-sectional household survey data in 1995, 1997, 1999 and 2001 and address non-random project placement and confounding economic trends by directly controlling for magisterial district fixed effects and trends, estimating the labor market effects of electrification using only within-district variation in electrification.

Results from both analyses show that employment in rural KZN increases in the wake of electrification. Female employment in the Census rises by a significant 9.5 percentage points in the IV results, which translates into 15,000 more women participating in the labor force, or 0.75 percent of the estimated 2 million new jobs created across the country over the period (Casale and Posel, 2004). The fixed effects analysis using household survey data largely supports these female employment results, although precise inference is more difficult with the small samples in this dataset. Electrification increases work on the intensive margin for women: in districts with the average increase in electrification over the period (15 percent), women work about 8.9 more hours per week, a 3.5 percent increase. Under both analyses, male employment rises (insignificantly) in electrifying areas, although to a lesser extent than for females.

Having established that household electrification increases employment in rural communities, I turn to investigating mechanisms in the second part of the paper. I first explore the impact of electrification on home production activities and find that newly electrified communities experience substantial shifts away from using wood at home, and toward electric

countries.

cooking and lighting. This suggests that household electrification operates as a labor-saving technology shock to home production in rural areas, releasing female time from home to market work.

Second, I rule out the possibility that household electrification stimulated large scale rural industrialization and hence a shift in labor demand by showing the absence of cross-community employment spill-overs. As further evidence that electricity stimulated a net increase in labor supply to the market, the fixed effects analysis indicates that female wages fall (albeit imprecisely) in districts where electrification is expanding more rapidly. This fact is difficult to reconcile with electricity causing large net increases in labor demand.

More plausibly, electricity may have lowered the cost of producing new, home-based services for the market, thereby presenting individuals with alternative ways to use their labor time in self employment and micro enterprizes. I am unable to provide direct evidence on these mechanisms, but I argue that since employment results for men and women are not statistically different from each other, it seems likely that the South African electrification did not exclusively affect rural labor markets through the channel of freeing time from home production. Rather, the reduced-form market employment results capture a combination of increased labor supplied to the market (via the home production channel) as well as increased small-scale labor demand (via new opportunities for producing new goods and services for the market).

A final channel that I investigate relates to migration. I discuss how differential in- and out-migration affect interpretation of the employment results. I show that differential in-migration cannot explain all of the employment effects of electrification, and explain how differential out-migration, while substantial, is also unlikely to account for employment effects, given the profile of out-migrants from rural areas documented in other datasets and by other researchers. Rather, the migration analysis broadly suggests that people may be induced to stay in or to move towards areas in which infrastructure is rolling out.

This paper contributes to two literatures. First, it adds to what we know about the microeconomic effects of physical infrastructure in developing countries, placing new emphasis on labor market effects in an area that has recently focussed on poverty, health and education outcomes.[4] The results here suggest that studies that ignore employment effects could be missing important economic impacts, particularly when the infrastructure has a home production bias. Second, the main result that female employment rises in electrifying areas connects with a large literature on the effects of changing constraints on women's work in the process of economic development.[5]

[4] For example, see Cutler and Miller (2005), Loshkin and Yemtsov (2005), Akee(2006), Duflo and Pande (2006), Banerjee et al (2007), Cattaneo et al (2007).

[5] See, for example, Goldin (1994), Goldin and Katz (2000), Mammen and Paxson (2000), Greenwood et

The paper begins by discussing how household electrification may affect rural labor markets through home and market production. Sections 2, 3 and 4 describe the context of South Africa's electrification, data and empirical strategies. Section 5 presents the main results while section 6 investigates the channels through which electrification affects employment. Section 7 concludes.

1 Theoretical impacts of household electrification

New access to household electrification may change the nature of work in the home as well as the amount and type of work that can be done in the market. Providing new public infrastructure to a location may affect migration of employed and unemployed individuals. Outlining the form each of these changes may take is important for interpreting the empirical results in the paper.

To begin, home production activities are important in my study area. Figure 1 (a) and 1 (b) show the fraction of rural African households in KZN reporting different sources of fuel for cooking and lighting in the 1996 and 2001 Census, separately for communities that get new access to electricity or not during this period. Almost 80 percent of households cook with wood and light their homes using candles in the mid-1990s. In electrified areas, the fraction of households cooking with electricity increases almost three-fold in five years, while the fraction of households using electric lighting more than triples.

The labor supply effect of such a shock to the technology of home production is, however, ambiguous.[6] With this new technology, households become more productive in time-intensive activities like food preparation and storage, and so may substitute more time towards these home-based activities. The same shock also increases the length of the effective day, producing an endowment effect that increases the demand for all normal goods, leading households to supply more labor to market. The more income-elastic the demand for market-intensive goods is, the stronger this endowment effect will be in pushing households to supply more labor to the market. Which effect dominates is theoretically ambiguous; however, the substitution effect is likely to be smaller since the demand for home-produced goods (e.g. meals) is bounded above. Therefore, we expect the advent of household electricity to change the nature of home production and increase labor supplied to the market, particularly for individuals who specialize in home production (i.e. women).[7]

al (2005), Bailey and Collins (2006) and Coen-Pirani et al (2008).

[6]Becker (1965) and Gronau (1986) provide the canonical models of home production, within which the labor supply effects of a shock to home production technology can be shown to be ambiguous.

[7]Responses to the technology shock may also differ across households. If there is heterogeneity across households in initial home production technologies, or in the degree of substitutability of home for market

Electricity may also change work opportunities in rural areas, by stimulating the growth of new firms that create jobs outside the home.[8] Quite apart from this, electricity may directly create jobs within households by enabling the production of new goods and services for the market: for example, food preparation and storage for larger groups becomes easier; operating small appliances to provide market services becomes feasible (e.g. hairdryers, cell phone charging stations, local craft production). In this way, household electrification could unleash previously unrealized demand for labor and an increase in market work, even without the growth of firms.

Household electrification may also affect migration behavior in multiple ways. In- and out-migration could be important responses to electrification, as people gravitate towards areas that are more desirable places to live. However, if in-migrants to electrifying areas already have jobs elsewhere or if out-migrants from non-electrified areas take their jobs with them, we might mistakenly attribute the increase in employment to new household electrification, when the main effect of the roll-out is merely to change the composition of the community.

To isolate how important each of these channels is in explaining the impact of rural electrification on market employment, it would be ideal to show what happens to (1) home production activities (2) market employment (3) the prevalence and size of firms in rural areas (4) the prevalence of home-based micro-enterprizes in rural areas (5) market wages in areas that gain new access to electricity, (6) and migration flows. Data limitations restrict the empirical analysis to (1), (2) and (5) and (6). I investigate whether new access to household electrification increases employment in the market and whether these effects differ by gender, whether changes in methods of home production and changes in wages support a labor supply channel, whether there is any evidence for the labor demand channels, and the extent to which migration into and out of electrifying and non-electrifying areas can account for employment effects. The results of these analyses substantially improve our understanding of the impacts of this infrastructure in a poor, rural setting.

2 South Africa's Electrification Program

By 1990, most economic entities and settlements in white cities and white commercial farms had been electrified. In contrast, one of the legacies of apartheid was that many African households were denied access to basic services, especially if they were living in designated

commodities (for example, meals versus child-care), then the labor supply effects of electrification may differ across these types of households. I present some evidence for this heterogeneity in Web Appendix 2.

[8] Rud (2009) documents the role that rural electrification played in industrializing India.

homeland areas (Gaunt, 2003).[9] At the time of the first democratic elections in 1994, over two-thirds of African households did not have access to electricity. After the elections, all homelands were legally reintegrated into South Africa (Christopher, 2001) and the South African government assumed responsibility for basic service provision for all of its citizens.

As part of a National Electrification Programme (NEP), South Africa's national electricity utility (Eskom) committed to addressing the service delivery backlog and electrifying 300,000 households annually from 1995 onwards. These targets were regarded as "firm and non-negotiable" (Eskom, 1996) and new connections were fully subsidized by the utility (Gaunt, 2003). Since Eskom was a monopolist in electricity generation and distribution during this period, industry commentators describe the support for this roll-out commitment as partly strategic. Eskom was interested in signalling to the government that full access to previously disadvantaged communities could be provided, without introducing competition into the industry.[10] As a result, Eskom met their connections targets in most years. Between 1993 and 2003, about USD1.4 billion was spent on household electrification and about 28 percent of all KZN households, or 470,000 households, were electrified. Almost all of these connections provided households with a minimum level of service, enough to power a few basic appliances at the same time.[11]

Even though all households within an area received the basic connection once the area was selected for electrification, this community-level selection was not random. Almost by definition, networked infrastructure of any kind requires that even identical consumers be connected in some order. And, in the context of the NEP, local political pressures and connections costs each played an important role in prioritizing communities for electrification. Gaunt (2003: 91) comments that although objective criteria were identified for ranking communities, political pressures were part of the "not-easily-identifiable but good reasons for selecting particular target groups". In KZN, both the 1994 provincial elections and the 1995/6 local government elections were hotly contested by the two leading political parties in that province. This political rivalry arguably influenced local public goods allocations. In the rest of this paper, I treat these political factors as omitted variables.[12]

[9] Homelands were pockets of land designated for African settlement and functioning as labor reserves for the economy. Throughout, I retain the use of apartheid-era racial classifications: African for black South Africans, and white and Indian.

[10] Personal communication with Trevor Gaunt, head of Department of Electrical Engineering at the University of Cape Town (May 31 2006).

[11] Service was limited to a power supply that could simultaneously power a few small home appliances e.g. two lights, a small television or radio, a small refrigerator and a water heaters (South African Department of Minerals and Energy, 2004). Newly connected households in my study area report large increases in ownership of electric kettles, refrigerators and lighting (own calculations, KwaZulu-Natal Income Dynamics Study 1993 and 1998).

[12] I use data from local elections in 2000 to shed some light on the importance of political factors in

Annual Eskom reports and interviews with planning engineers also point to the central role of costs in allocating projects to places. The dual pressures of connections targets and internal financing meant that Eskom had strong incentives to prioritize areas with lowest average cost per household connection.[13] These cost factors are central to the main identification strategy in this paper. The bulk of electrification cost is in laying distribution lines out from electricity sub-stations to households. Three factors reduce the cost of these distribution lines: proximity to existing sub-stations and power lines; higher density settlements; and terrain, or land gradient. The less of an incline the land has, the fewer hills and valleys and the softer the soil, the cheaper it is to lay power lines and erect transmission poles (Eskom, 1996; West et al, 1997).

I assemble measures of these three cost factors in my data. Distance from the grid and household density are important control variables, since both are likely to be correlated with economic opportunities that could 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 forms the basis of my instrumental variables strategy that addresses the biases arising from selection on unobservable variables and confounding trends. Further motivation for using gradient as an instrumental variable is postponed to Section 4.

3 Data and sample characteristics

For the main analysis of the employment effects of electrification, I construct a panel data set of community aggregate variables using 1996 and 2001 South African Census data. To this community-level panel, I add in three additional pieces of data: spatial data collected from Eskom on the location of electrification infrastructure in KZN at baseline (1996), administrative data on project placement across the province between 1990 and 2007, and measures of geography at baseline (community land gradient, distances between each community and the nearest electricity substation, road and town).[14] For some parts of the analysis, I also refer to the 10 percent micro Census data for 1996 and 2001.

The unit of analysis for the IV strategy is a community-year. Communities are small, with most having fewer than 900 households. They fall uniquely into 10 districts across

assignment of projects to communities in Web Appendix 3.

[13]Barnard (2006) describes factors affecting network extension to rural communities in KZN: "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."

[14]Details of data sources and data linking procedures are in Web Appendix 1: Data.

the province (on average, there are 181 communities per district) and each district operates much like a local labor market.[15] I restrict the sample to rural ex-homeland communities in KwaZulu-Natal (KZN). This province is home to one-fifth of the population of South Africa and in the early 1990s, contained about 30 percent of the entire African population living in homeland areas. Households in these rural areas are more reliant on traditional fuels than urban households and so are more likely to experience larger effects of electrification. There are also potentially fewer economic confounders in rural than urban areas in the first years after the end of apartheid.

My second empirical strategy uses individual-level data on employment, hours of work, wages, earnings, demographics and households fuel sources from four cross sectional household surveys: the 1995, 1997 and 1999 South African October Household Survey (OHS) and the 2001 September Labor Force Survey (LFS). These micro data are collapsed to magisterial district (MD) aggregates that are larger than communities (38 in my sample) but smaller than Census districts.

3.1 Sample characteristics

Tables 1 and 2 present means and standard deviations of key variables used in the main analysis. All variables are derived from the 100 percent Census sample, so results are not weighted. Table 1 provides descriptive statistics of baseline variables for the full sample of 1,816 communities (column (1)), and separately by Eskom project status of the community (columns (2) and (3)). Communities in the sample are poor: 61 percent of households live on less than 6,000ZAR per year, approximately USD840 at a 2006 USD/ZAR exchange rate. On average, over half of households in a community are female-headed and the female/male adult sex ratio is well over 1. These values underscore the historical function of the homelands as migrant labor communities.

The table also shows values of the three key variables influencing the cost of electrification projects. Average household density is 22 per square kilometer, and communities are on average 19 kilometers away from the nearest electricity substation in 1996. Main roads and towns are further away, at an average distance of 38 kilometers. That communities are closer to the electricity grid than to towns is largely because all white commercial farms were electrified by the end of the 1980s. The final row in the table shows that average community land gradient is 10 degrees. This is "strongly sloping", according to the Food and Agriculture Organisation's gradient classification (FAO, 1998). The first map in Figure

[15] In household survey data, only a handful of people report working outside of their district. In contrast, over half of all women and 60 percent of men work outside of their community (own calculations, Census 2001 micro data, 10 percent sample).

3 shows the spatial distribution of the gradient variable, along with community boundaries of the sample. Shaded areas are communities included in the analysis sample. The geographical fragmentation that characterized former homeland of KwaZulu is evident: the apartheid government forcefully resettled Africans to areas deemed inhospitable for white settlement, wherever those happened to be, with the result that the former homeland areas was not geographically contiguous across the province (Christopher, 2001). Gradient varies widely across the region with dark shaded areas being the steepest.

Administrative data indicate that 20% of communities in the sample area received an Eskom project between 1996 and 2001 (inclusive). The remainder either never received an electricity project or only had a project after 2001, or prior to 1996. The strength of defining electrification status using project data is that new access to infrastructure can be directly identified, rather than inferred from time variation in electricity use, which may be correlated with changes in wealth that are difficult to control for in a two-wave panel.

Several features of project placement are evident in the second map in Figure 3, which shows the distribution of (dark shaded) electrified and (light shaded) non-electrified areas. Being close to the original grid is neither necessary nor sufficient for electrification between 1996 and 2001. Proximity to a town is also not necessary for electrification. Finally, electrified areas are distributed across districts rather than clustered in one area. This important fact makes it possible to include district fixed effects in the main analysis to absorb aggregate differences in employment growth rates across local labor markets.

Stark differences across communities with and without an Eskom project are evident in Table 1, columns (2) to (4). Compared to non-electrified areas, electrified communities are significantly less poor, have fewer adult women relative to men, have higher fractions of high school-educated adults, and are almost 3 kilometers closer to the nearest road and town. Given that low average cost areas were prioritized for projects, it is not surprising that electrified areas have significantly higher household densities, are 4.1 kilometers closer to the nearest substation, and have a 1.2-degree flatter average gradient than areas without an Eskom project. If electricity projects had been randomly assigned to communities, most of these observable characteristics would be balanced across project and non-project areas. Instead, a joint test of the hypotheses that each of these differences is zero can be rejected at the 1% level.[16]

Since the main analysis is based on using gradient to instrument for project placement, I compare values of each covariate across steep and flat areas in the last two columns of Table 1.

[16] I implement this as a Bonferroni test. The relevant p-value for rejection of this joint null at the 1% level of significance, given 10 variables, is $p < 0.01/10 = 0.001$. If at least one p-value is less than 0.001, the null is rejected. In column (4), the null is decisively rejected at the 1% level; while in column (6), this null is not rejected at the 1% level.

I regress each covariate on gradient alone (column (5)) and then include all other covariates and 10 district fixed effects as controls (column (6)). There are no significant differences in poverty rate, the fraction of female-headed households, any of the distance variables or the fraction of females with high school. There are remaining, although small, differences in the adult sex ratio (0.004), household density (0.95 households per square kilometer) and fraction of men with high school (0.003) although a joint test for each difference being zero cannot be rejected at the 1% level. Therefore, column (6) shows that gradient balances more of the community-level variables at baseline, conditional on all other controls.

3.2 Describing community-level employment rates

The main outcome variable this paper analyzes is the employment to population rate of African women and men, ages 15 to 59 (inclusive). Questions about employment in the Census are fairly broad, and similar across years.[17] Table 2 presents average employment rates for men and women, across Eskom project and non-project areas, in each year, as well as the differences in these rates across years (in rows labeled Δ_t) and across areas (in column (4)).

Two striking points emerge from this table: Employment rates are very low for men and women, and are falling– and falling faster– for men in electrified areas between 1996 and 2001. In column (2), female employment remains low (7 percent) and steady across communities between 1996 and 2001 while male employment falls from 14 to 10 percent. Employment is uniformly higher in electrified than in non-electrified communities in 1996. Comparing changes in employment rates in Eskom project areas to the same change in non-project areas (column (4)), the unadjusted difference-in-differences for women is not significantly different from zero while for men it is a statistically significant -1.7 percentage points.

That South Africa has low levels of employment is not a new insight (for example, see Klasen and Woolard (1999) and Banerjee et al (2007)). However, employment rates in Table 2 are extremely low, even for this country. This is partly because there are only a few broad questions on employment in the Census and no probing for work activities as would occur in a labor force survey. Another reasons for these low employment rates is that my sample includes only rural, ex-homeland areas of KwaZulu-Natal. As described by Ardington and Lund (2006: 12), the homelands "consigned millions of people to rural areas with few employment opportunities". These ex-homeland areas are ill-suited for agriculture, so that work opportunities in these areas are concentrated in civil service (mainly teaching) and domestic work, both jobs favoring the employment of women. Many jobs in these areas are

[17] See Web Appendix 1: Data for details on the construction of employment variables.

also marginal, with workers working under 20 hours per week (Ardington and Lund, 2006) and large fractions of households rely on income from welfare grants (old age pensions) and migrant workers.[18] However, even using individual-level data from surveys designed to capture all types of work, employment in rural areas of KZN is very low, and employment in agriculture is almost non-existent.

The large drop in employment for men in Eskom project relative to non-project areas should not be interpreted as the causal effect of electrification. Rather, these changes in employment rates for men and women are confounded by broad changes in the South African labor market during the 1990s. Figure 2a shows trends in male and female employment in rural KZN (including homeland areas) over time using the OHS and LFS household surveys in 1995, 1997, 1999 and 2001. These are the same data used in the fixed effects analysis in section 5.3. Employment rates using these data are higher than in the Census, but still extremely low. Employment for men falls significantly between 1995 and 2001 and falls to a lesser extent for women. Figure 2b shows wage trends using the same data. Over the period, male wages are roughly constant while female wages fall and are lower in 2001 than in 1995. Dissecting overall changes in employment, Banerjee et al (2007) document large shifts in the composition of jobs away from commercial agricultural and mining sectors, and towards service and retail sectors. These trends continued into the 1990s and had a heavy impact on jobs in male-dominated sectors. The types of new jobs created during this time were predominantly low skill and in the informal sector, in sectors that favor female workers (Casale and Posel, 2004) and there is evidence that the number of jobs for self-employed workers and household workers increased substantially between 1995 and 2001 (Banerjee et al, 2007).

A common challenge in evaluating the economic effects of an expansion in infrastructure revolves around how to control for expansions in the economy that may confound the effects of the new infrastructure. The South African case presents a different challenge. Eskom was more likely to be electrifying households in areas that were experiencing longer-term declines in employment and economic activity. This is because grid expansion was constrained by initial network placement, and the network that existed at the end of apartheid had been set up to service commercial farms and previously white towns. Hence, many of the factors that determined whether a community got early access to electricity were the same factors that increased a community's exposure to the industrial restructuring of the 1990s. The results of this type of selection are evidence in the falling male employment rates in Table 2. In the

[18] In Web Appendix 4, I show that the community Census data likely undercounts employment relative to household and labor force survey data, and that this undercount appears somewhat larger for men than for women.

next section, I outline two different empirical strategies that deal with endogenous project placement and these confounding factors in alternative ways.

4 Empirical strategies

Let y_{jdt} be outcome y (for example, the female employment rate) for community j and district d in time period $t = [0, 1]$. T_{jdt} is an indicator variable for whether a community has received an electricity project by time period t. If electrification T_{jdt} was randomly assigned across communities, we could estimate the average treatment effect of electrification (α_2) by ordinary least squares as in (1):

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

where μ_j is a community fixed effect, $\delta_j t$ is a community trend, ρ_d is a district fixed effect, $\lambda_d t$ is a district trend and ϵ_{jdt} is an idiosyncratic error term. To eliminate μ_j and ρ_d, re-write equation (1) in first differences:

$$\Delta y_{jdt} = (y_{jdt+1} - y_{jdt}) = \alpha_1 + \alpha_2 \Delta T_{jdt} + \lambda_d + (\delta_j + \Delta\epsilon_{jdt}) \tag{2}$$

With the two wave Census panel, I can measure Δy_{jdt}, ΔT_{jdt} and λ_d, but not δ_j. OLS estimation of (2) will not identify the causal effects of electrification as long as $\delta_j + \Delta\epsilon_{jdt}$ is correlated with ΔT_{jdt}. If electricity projects are allocated to communities growing faster for unobservable reasons then $\hat{\alpha}_{2,OLS}$ would be biased upwards. However, the results in the previous section suggest that we should be more concerned with negative selection, and a downward bias in $\hat{\alpha}_{2,OLS}$ in the South African case.[19]

To deal with factors that could affect a community's growth path (δ_j), I first control for a vector of community covariates (X_{jd0}) measured in 1996 in estimating equation (2). Covariates include household density; fraction of households living below a poverty line (ZAR 6,000 per household per year); distances to the grid, road and town; fraction of adults that are white or Indian to proxy for local employers; fraction of men and women with a completed high school certificate; and two standard proxies for community poverty, the share of female-headed households and the female/male sex ratio (Standing et al,1996). I also include a set of 10 district fixed effects, so that all comparisons across project and non-project areas occur for areas in the same local labor markets.

Even with these controls, however, confounding trends in community-level employment

[19]Measurement error in ΔT_{jdt} presents another practical challenge for estimating equation (2). See the discussion of this issue in Web Appendix 4.

and unmeasured political factors that could affect project placement are still of concern. To overcome these challenges to identification, I instrument for program placement using average community land gradient (Z_j). The system of equations to be estimated is:

$$\Delta y_{jdt} = (y_{jdt+1} - y_{jdt}) = \alpha_1 + \alpha_2 \Delta T_{jdt} + X_{jd0}\beta + \lambda_d + (\delta_j + \Delta\epsilon_{jdt}) \tag{3}$$

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

where $(\delta_j + \Delta\epsilon_{jdt})$ and τ_{jdt} are unobserved. The identification assumption is that conditional on baseline community characteristics, proximity to local economic centers and grid infrastructure, and district fixed effects, land gradient does not affect employment growth independently of being assigned an electrification project.

One concern with using land gradient as an instrumental variable in a rural setting is that it may directly affect agricultural outcomes. In rural KZN, the direct impact of gradient on agricultural productivity and agricultural employment growth is limited, since most people are not farming. Under 10 percent of employed individuals are involved in agriculture.[20] A second concern is that individuals may sort, non-randomly, across flat and steep areas which could result in differential employment growth, independent of new electrification. While mobility *within* homeland areas during this time is limited by a lack of property titling and the role of tribal authorities in land allocation, in-migration and out-migration do occur, as I describe in the last part of the paper[21]. I show that differential in-migration to flatter areas cannot account for the employment effects of electrification and argue that selective out-migration cannot explain employment effects either, given the profile of rural out-migrants.

Conditional on instrument validity, $\alpha_{2,IV}$ captures the local average treatment effect (LATE) of electricity projects on community-level employment growth. In my results, community composition drives marginal effects. So, if individuals living in flatter areas can better afford electricity once it arrives, or, if individuals living in flatter communities have fewer other home production demands (i.e. child-care), then a larger than average treatment effect may be measured for these areas. Employment returns to electrification may also differ by gradient, leading to larger estimated employment effects for marginal than for average communities. For example, flatter areas always have lower commuting costs, so individuals in flatter areas always face a higher net wage. Since these individuals are initially closer

[20] Farming accounted for only 10 percent of household earnings in homeland areas by the mid-1980s (Vink and Schirmer, 2002). Ardington and Lund (1996: 48) write that "a significant percentage of the income of rural households is sourced outside the household and indeed outside rural areas" and that "land is nowhere the 'main source' of income for the majority of rural households" (Ardington and Lund, 1996: 55). Web Appendix 4 provides more details about the low levels of agricultural employment in rural KZN.

[21] Personal communication, Department of Land Affairs, Pietermaritzburg (June 2006).

to the employment participation margin, they will always be more likely to respond when electricity arrives.[22] These reasons lead us to expect IV estimates to be larger than average treatment effects.

To complement the IV strategy, I present an alternative identification strategy which I refer to as the MD-FE/MD-trends analysis. I pool information from four cross-sections of South Africa household survey data to estimate the impact of electrification on male and female employment, hours of work, wages and earnings. The sample is restricted to African men and women living in rural areas of KZN, for which there are at least 900 respondents per year. The major drawback to using these data is that respondents can only be situated in the magisterial district (MD) in which they reside, which cannot be linked to the Eskom project data.

I regress each of the labor market outcomes on age, age-squared and years of education, obtain the residuals from these regressions and average the residuals within year (t), magisterial district (m) and sex (s) to create up to 304 observations on outcomes (4 years*38 m observations each for males and females). I also construct the fraction of households with electric lighting for each MD-year ($ELEC_{mt}$). This is a reasonable proxy for expanding access to the grid since almost all households getting access to the grid were able to use electric lighting. Then, I estimate regressions of these residuals ($\bar{\hat{\epsilon}}_{mt}$) on $ELEC_{mt}$, a common time trend (t) and a full set of MD fixed effects (λ_m) and MD-specific trends (δ_{mt}):[23]

$$\bar{\hat{\epsilon}}_{mt} = \gamma_0 + \gamma_1 ELEC_{mt} + \gamma_2 t + \lambda_m + \delta_m t + \nu_{mt} \tag{5}$$

Without controlling for MD-FE and MD-specific trends, γ_1 is identified using variation in electric lighting within and across MDs. In the MD-FE/MD-trends specification, γ_1 is identified using variation in electric lighting within the MD over time, after accounting for λ_m and $\delta_m * t$. Including MD-specific trend terms controls for differential trends across MDs with different rates of electrification that could confound the labor market impacts (this is analogous to the correlation between δ_j and project status ΔT_{jdt} in the main empirical strategy). Although these regressions are estimated on a small sample (38 MDs in each wave of data) which makes precise estimation difficult, they do provide useful complementary evidence of the effects of electrification on employment on the extensive and intensive margins

[22]A potential threat to validity arises if gradient is strongly correlated with road access (e.g. Nunn and Puga (2007) discuss the impact of terrain ruggedness on transportation costs). Changing economic activities in distant markets may be more easily accessible for flatter communities, hence making gradient itself a 'treatment'. To test whether employment is only responding to access to roads, I re-estimate results for communities without main roads. Results for female employment, presented in Web Appendix 3, are qualitatively similar.

[23]Instrumenting for $ELEC_{mt}$ is not possible in this framework, as gradient has no predictive power in explaining electrification rates at the more aggregated magisterial district level.

and on earnings and wages. Moreover, given the richer set of labor market outcome variables, these results can be informative about whether electrification affects labor demand or supply in rural areas, or both.

5 Results

5.1 Assignment of electricity projects to communities

First-stage estimates for the allocation of an electricity project to a community are presented in Table 3. The outcome variable is an indicator for whether a community received an electricity project between 1996 and 2001. The coefficient on gradient indicates that for a two standard deviation increase in gradient (about 10 degrees), the probability of receiving an Eskom project falls by about 8 percentage points. Across columns, the size of the coefficient does not change substantially with the addition of more controls while the precision of the estimate improves.

The inclusion of district fixed effects in this first stage is important, as a large amount of the variation in gradient comes from cross-district variation, as evident in Figure 3. This means that without controlling for district, I am comparing project assignment across very different places in terms of gradient and in terms of local labor market conditions. By controlling for district as in columns (3) and (4), I am comparing places that are in the same local labor market, but which are slightly flatter or steeper.

The two other cost variables have coefficients of the expected signs in the first stage results of Table 3: a three-quarter standard deviation increase in distance from the grid (about 10 kilometers) reduces the probability of electrification by 1 percentage point, although this is not significant when all other controls are added. A one-third standard deviation increase in household density (10 households) per square kilometer increases the probability of electrification by about 1.3 percentage points. The influence of household density is robust and strongly significant across specifications.

These project assignment regressions provide mixed evidence on whether newly electrified areas are positively selected on wealth. While areas with more female-headed households (i.e. poorer areas) are significantly less likely to receive an electricity project, areas with more white and Indian adults (i.e. richer areas) are also less likely to be electrified during these years. The community poverty rate and sex ratio variables also have large positive coefficients in all specifications, suggesting that projects may be targeted to poorer areas. This lack of strong evidence for project placement in richer areas and strong predictive power of two of the three cost variables is consistent with the overarching socio-political motivation for the roll-out.

5.2 Employment effects of electrification: OLS and IV results

Coefficients from OLS and IV regressions of employment are presented in Table 4 for women (Panel A) and men (Panel B). The tables provide estimated coefficients and robust standard errors for a subset of control variables, clustered at the sub-district level.[24] The dependent variable in each column is the change in female (or male) employment rate between 1996 and 2001. Columns (1) to (4) present OLS results, column (5) presents the reduced form regression estimates and columns (6) to (9) present the IV results.

The coefficient on Eskom Project in column (1) echoes the descriptive statistics in Table 2: there is no significant change in female employment across project and non-project areas while male employment falls by 1.7 percent. Adding community-level controls and district fixed effects in columns (2) and (3) increases the coefficient on electrification slightly, with the female employment effect still not significantly different from zero and male employment becoming less negative and less statistically significant. The positive, significant coefficients on poverty rate, sex ratio and female-headed households in both tables indicate that female and male employment rises faster in poorer places in the late 1990s.

IV estimates of electrification are substantially larger than OLS estimates and significantly positive for women in Panel A columns (8) and (9). Since Table 1, column (5) indicated that gradient is correlated with some of the control variables and since the F-statistic on the excluded variable in the first stage is larger once other controls absorb residual variation (Table 3), my preferred estimates are in columns (8) and (9) of Table 4.[25] In these columns, female employment increases by 9 to 9.5 percentage points, or between 30 and 35 percent from baseline, in the wake of an electricity project. The Anderson-Rubin (AR) test for whether electrification raises female employment strongly rejects zero and the 5% confidence interval is wider than the standard 5% confidence interval, ranging from 5 to 35 percentage points. Male employment increases by a substantially smaller 3.5 percentage points, and this is not significantly different from zero under either the standard test or the AR test (Panel B column (9)). Although I cannot reject that the male and female employment effects are the same, there is no reduced form for male employment (Panel B, column

[24] The sub-district level is one level of aggregation up from the community level and one level below the district. Inference is robust to estimating standard errors using Conley's spatial error correction methods (Conley, 1999) (see Web Appendix 3).

[25] To address concerns about over-optimistic inference with a possibly weak instrument, heteroscedasticity-robust Anderson-Rubin (AR) confidence intervals are computed for the main Eskom Project parameter estimate in the second stage and shown in Table 4. These AR confidence intervals have correct coverage properties in the presence of weak instruments while standard Wald tests do not (Mikusheva and Poi, 2006; Chernosukov and Hansen, 2007).

(5)).[26] It is therefore difficult to precisely estimate the impact on male employment using these Census data; part of this may be related to the fact that the Census undercounts male employment more than female employment in these areas (see Web Appendix 4 for details).

Another aspect of these results that bears mentioning is the sensitivity of the female employment results to the inclusion of district fixed effects in equation (3). This reflects the fact that differences in gradient are larger across districts than within districts. Excluding district fixed effects means that employment effects are identified off of cross-district comparisons in female employment growth. Since local labor markets differ substantially across districts, including district fixed effects allows me to identify the effect of electrification by comparing slightly steeper to slightly flatter areas within the same local labor market.

The IV results suggest that in a non-electrified community with the median number of adult women in 1996 (N=264), a 9 percentage-point increase in female employment raises the number of women working by 23 women, from 18 to 41. If we assume this 9 percentage point increase applies to the entire group of electrified communities (rather than marginal communities only), this translates into an increase of approximately 15,000 newly employed women out of the baseline female population of 165,637. This is 0.75 percent of the estimated 2 million new jobs created across the country over the period (Casale and Posel, 2004).

5.2.1 Threats to validity in the IV strategy

If employment rates in steep and flat areas evolve differently, in the absence of new electricity, the gradient IV would be invalid. Without more years of data, this is difficult to check directly. Instead, I implement an indirect placebo test using historical administrative data on electricity projects. These data identify areas that are electrified prior to 1996, which were excluded from the main analysis. For these areas, there should be no reduced-form relationship between gradient and employment growth between 1996 and 2001, since they have already received an electricity project. If there is, this would suggest that gradient has a direct effect on employment growth. To test this, I estimate OLS regressions of the change in female employment in areas electrified prior to 1996 ($N = 373$) on gradient and the full set of controls. Column (1) of Table 5 contains the results of this placebo test. The coefficient on gradient is small (-0.001) and insignificant, yet significantly different from the 0.007 reduced form coefficient on gradient in Table 4, Panel A column (5). Thus, there is no evidence of any reduced-form relationship between gradient and female employment (the same is true for males, results not shown) in the set of areas already electrified by 1996. This

[26] I implemented this test by differencing the male and female outcome variables within community and estimating the same OLS and IV regressions using this new dependent variable. This test respects the correlated structure of the error terms ($\Delta\epsilon_{jdt}$) across male and female regressions (see Web Appendix 3).

boosts confidence in the research design.

A second potential threat to the validity of the IV strategy arises if flatter communities received positive labor demand shocks concurrent with electricity projects. Unfortunately, no dataset captures the presence of firms in rural KZN regions. Instead, I test whether there are larger increases in the major sources of female labor demand in flatter communities. Micro Census data suggest that most women in these areas work as teachers or as domestic workers. In columns (2) and (3) of Table 5, I test whether gradient is negatively correlated with growth in new schools (using data from the South African Schools Register of Needs) or with the growth in new employer households (proxied for with the change in fraction of Indian and white adults in the population).

Despite the fact that the number of schools across rural KwaZulu-Natal increases by almost 20 percent between 1995 and 2000, which undoubtedly increases the demand for teachers, column (2) shows this increase is uncorrelated with community gradient. And, although other researchers have documented the growth in low skill, informal sector jobs in the economy during the 1990s (Banerjee et al, 2007; Casale and Posel, 2004), the results in column (3) Table 5 indicate no differential expansion in this source of demand for female workers in flat relative to steep areas of rural KZN.

5.3 Employment and wage effects of electrification: Results from the MD-FE/MD-trends analysis

To provide supporting evidence on the employment effects found using the IV strategy and to shed light on the mechanisms through which electricity raises employment, I turn to results from the MD-FE/MD-trends analysis. Table 6 presents coefficients from OLS and FE regressions of equation (5), for employment rates (Panel A), usual weekly hours of work (Panel B), log wages (Panel C) and log earnings (Panel D). Standard errors are robust to heteroscedasticity and clustered at the MD-level. Recall that the MD fixed effects and MD-specific trends control for the differential economic trends that could confound the impact of electrification on labor market outcomes. The coefficient on electrification is identified off of the variation in electrification rates over time, within an MD, after MD-trends have been accounted for.

Consider first the estimates for employment: in areas where electrification increases, male and female employment increase substantially in the OLS specification. The average increase in electrification over the period (0.15) translates into a 1.3 percentage point increase in employment for men and a 1.8 percentage point increase for women, although male-female differences are not statistically different from zero. Coefficients are similar under OLS and

FE specifications, however, once all fixed effects and trend terms are included, none of the electrification coefficients are precisely estimated in this small sample. Weekly hours of work exhibit the same pattern, with OLS coefficients being estimated more precisely than FE coefficients. Women work 8.9 hours more and men work 13 hours more per week in MDs with higher electrification rates, compared to the same MDs in periods of lower electrification. For the average change in electrification rate (0.15), this amounts to between 1.3 and 1.9 hours more work per week. The male-female differences are again not statistically different from each other. The magnitude of this intensive margin response is consistent with the new work being informal and perhaps in self-employment rather than in full-time formal sector positions.

It is worth comparing the employment results in Table 4 with those of Table 6. Both approaches show female employment rising in electrifying areas, either on the extensive or intensive margins. Male employment effects are never significantly different from zero once selection has been accounted for, but the coefficients on electrification are still generally large and positive. Using variation in project status across steep and flat communities in the same local labor market, Table 4 tells us that in areas that received an Eskom project, female employment increased by 9.5 percentage points; relative to baseline female employment of about 7 percent. Using a different source of variation, household survey results in Table 6 indicate that employment increases by a smaller 1.8 percentage points for women in MDs with the average change in electrification rates. Hours of work increase slightly more, at 3 to 4 percent in electrifying areas.

There are three reasons why these results differ in magnitude. First, while the IV strategy focuses on changes in small communities, the MD-FE/MD-trend analysis examines changes in larger MDs. It is not clear that we should expect analysis at different levels of aggregation to produce the same results. Second, each strategy uses different sources of variation: the IV strategy compares flat to steep areas while the MD-FE/MD-trends analysis uses variation within the same MD over time. Again, it is not clear that we should expect these comparisons to be identical, although it is comforting that they point in the same direction. Finally, new access to electricity is measured in different ways under each strategy: as a binary variable in the IV strategy and as the fraction of households with electric lighting in the MD-FE/MD-trends strategy. We can use information on the change in fraction of households using electric lighting in project versus non-project areas to re-scale the IV results. In Census communities that experience the Eskom-induced increase in electric lighting (65 percent, explained in Table 8 below), female employment rises by 6 percentage points (0.095*0.65). This re-scaled employment result from the IV analysis is much closer to the results from the MD-FE/MD-trend analysis.

Turning to the effects of electrification on wages and earnings in the lower panel of Table 6: wages for women fall in areas where electricity is rolling out (Panel C, columns (1) and (2)), and more so in the MD-FE specification. For the average change in fraction of households with electric lighting, women's wages fall by about 20 percent (1.38*0.15), while for men, the coefficient on electrification rate is positive but not significant. Combining the increase in female hours of work, a large (but insignificant) increase in employment on the extensive margin and the decline in wages, it is not surprising that there are no significant differences in female earnings across electrifying and non-electrifying areas (Panel D, column (5)) or within an MD that sees growing electrification over time (Panel D, column (6)). In contrast, male earnings do rise significantly when electrification rates are higher, by about 16% for the average increase in electrification (0.15*1.10). This also makes sense, given that men appear to be working more hours without any decline in average wages.

The combined results of section 5.2 and 5.3 suggest the following interpretation: when communities get new access to household electricity, employment on the extensive margin increases for women and possibly for men, although male effects are difficult to estimate precisely. On the intensive margin, the best household survey evidence we have indicates that electrification raises hours of work for women and men (although precise estimation of these effects are precluded by the small sample size). And, given the results of the placebo test, there is no strong evidence that contemporaneous expansions in sources of demand for female work confound these employment results. In the next section, I investigate several channels through which electrification may have affected employment in these rural areas.

6 Channels

6.1 Electrification and home production: A labor supply channel

In order for electrification to affect employment through the channel of reduced time in home production, households must switch out of traditional fuels when their communities are connected to the grid and spend less time in home production. There are no data on time use to show the latter effect. However, the simple averages in Figure 1 and results presented in Table 7 illustrate that households do make large adjustments to their home production technologies in the wake of household electrification. Each coefficient reported in the table is from a separate regression, where the outcome variable is the change in fraction of households using electricity for lighting or cooking or using wood for cooking. Columns (1) and (3) do not contain any additional controls while columns (2) and (4) report results from regressions containing all relevant control variables. Robust standard errors are clustered at the sub-district level.

Both OLS and IV regression results illustrate substantial shifts towards using electricity for home production with IV results larger than OLS estimates. Average rates of electric lighting rise by 23 percentage points more in communities with an electricity project than in communities without in the OLS comparison of row (1), column (2). In the same column, reliance on wood for cooking falls by 3.9 percentage points and cooking with electricity rises by 5.6 percentage points. Column (4) indicates that in areas chosen to be electrified because of their flatter gradient, use of electric lighting increases by a substantial and significant 65 percentage points, wood use falls by 27 percentage points and cooking with electricity rises by 23 percentage points.[27]

To check that gradient is not simply picking up easier access to all types of services that could affect home production, rows (4) and (5) of Table 7 present results for two additional outcome variables: the change in fraction of households with access to piped water close to home and the change in fraction of households with a flush toilet at home. There is no evidence that electrified regions experience differential changes in either of these basic services. In fact, the IV results for water services in column (5) and (6) are in the opposite direction to what we would expect if gradient was simply a noisy measure of wealth.

In combination with the main results of the previous section– rising female employment and some indication of falling female wages in electrifying areas– the results on changing home production in Table 7 suggest that one important channel through which electricity affects the rural labor market is by "freeing up" women's time for the market. However, this is unlikely to be the only way in which this infrastructure roll-out affects rural areas. In fact, the similarity of the male and female employment results hint at electricity facilitating new activities for men and women that would allow them to start produce market goods and services at home (e.g. food preparation, personal services requiring electric appliances). However, we would like to be more confident that electrification does not stimulate large net increases in labor demand in these communities. This is what I test for next.

6.2 Electrification and labor demand

Communities as defined in the Census data are small. Hence, any electricity project that generates new firms and new demand for labor should have spatial spillover effects into neighboring areas. If firms create jobs for people living in neighboring areas, positive spillovers in these non-electrified areas would dampen any effects of household electrification. If people move out of neighboring non-electrified areas towards electrified areas to get one of the new

[27]Web Appendix 2 discusses reasons for why the IV results are larger than OLS results.

jobs, a negative spillover would amplify electrification effects. In both cases, the effect is the sum of an incumbents' effect and a spillover effect. In both cases, OLS and IV coefficients should be substantively different when adjacent non-electrified areas most susceptible to these spillovers are excluded from the analysis.

To test this, I re-estimate OLS and IV regressions after excluding non-electrified areas within a one- and five-kilometer radius of an electrified area. Table 8 presents results for each restriction. OLS coefficients are never significantly different from zero, while IV coefficients are large, positive and close to the main IV estimate: neither 0.076 nor 0.069 could be rejected in the full sample. Using this test, there is no evidence of large spillovers across communities.

Combining this lack of spatial spillovers with the facts that the roll-out was driven by household targets, that capacity was too small to stimulate even mid-size manufacturing or service enterprizes (South African Department of Minerals and Energy, 2004), and that female wages are not increasing in electrifying areas, it is implausible that household electrification created jobs by sparking the industrialization of rural KZN.

6.3 Migration and labor market effects of electrification

A final channel through which electrification may affect employment growth is through migration. In Table 9 (Panel A, columns (1) and (2)), I present coefficients from OLS and IV regressions of the log of adult population on an Eskom project indicator and all other control variables. Even after controlling for all other variables, electrified areas have significantly higher population growth rates than non-electrified areas. Population grows by 17 percent more in Eskom project areas, and this growth is 380 percent higher in the IV specification.[28] Given these large differences in population growth, it is important to consider how migration may affect the interpretation of the main employment results.

One possibility is that individuals move towards areas that are electrifying, or away from non-electrifying areas, since the availability of this new infrastructure affects the quality of life across areas. This type of response would be captured as part of the IV employment results. A second possibility is that for reasons unrelated to infrastructure roll-out, flat areas have higher in-migration rates or lower out-migration rates than steep areas. In this case, migration flows could confound IV employment results. In either case, it is differential migration by employed individuals that is relevant for interpreting our employment results.

[28]Clearly, in small communities, numerically small increases in population can translate into large percentage changes. The average number of females (males) in these communities in 1996 is 356 (274). This rises to 446 (319) by 2001. Just considering the raw changes in number of adults over time, electrified areas grow at about 6 percent per year while non-electrified areas grow at about 3 percent.

For example, if individuals who already have jobs elsewhere move in to electrifying areas at higher rates, the direct impact of electricity on employment creation would be inflated. At the same time, if employed adults leave at higher rates from areas that are not being electrified, this would artificially deflate employment in non-project areas. Either type of migration flow would change the composition of the population in electrified relative to non-electrified areas.

In Table 9, Panel A (columns (3) to (6)), I present some evidence that this type of compositional change is present in my sample. I estimate OLS and IV regressions of the change in fraction of men and women with a high school education on all controls (except 1996 education variables) and present coefficient estimates for the Eskom project indicator. While OLS results indicate no differential change in the fraction of skilled females and a falling fraction of skilled men in communities getting access to the grid, the IV results do give us some pause: in columns (4) and (6), the coefficient on Eskom project is similar to the coefficient in the employment regressions of Table 4. A combination of skilled migrants flowing toward flatter areas at higher rates and skilled migrants leaving steeper areas at higher rates account for these compositional changes.

Ideally, it would be possible to estimate employment effects of electrification net of all compositional change. As a first step, differential in-migration can be ruled out as a confounder of the employment results in Table 4. By redefining the employment to population rate to exclude the total number of recent in-migrants from both the numerator and denominator (people who move in to communities in the five years before the Census), I re-estimate the main OLS and IV regressions for the set of incumbents. The new employment variable is therefore the most conservative measure of employment for incumbents. Panel B, columns (3) through (6) demonstrate that electrification effects are still present and, if anything, are larger for incumbent women, and not significant for men. However, Panel B columns (1) and (2) indicate that in-migration is only part of the story: growth of the incumbent population (excluding recent in-migrants in 1996 and in 2001) remains higher in areas that receive an Eskom project by virtue of gradient.

While Census data do not allow me to directly test whether higher out-migration from non-electrifying areas accounts for all of the main employment result, note that out-migrants would need to be employed *before* they migrate, for this to be of concern. If out-migrants are unemployed before migrating, then migration that is higher from non-electrifying areas would work against finding an employment effect of electrification. In fact, although out-migrants from rural KZN do tend to be more educated than those remaining, they are significantly *less* likely to be employed, relative to incumbents.[29] Other researchers have

[29]In Web Appendix 3, I use cross-sectional data from a migration module included in the 2002 Labor

also documented these facts. In an early study, Cross et al (1998) document high rates of rural-to-rural migration in KwaZulu Natal for the purpose of finding work or finding places to live with better infrastructure. Burger et al (2002) use 1996 Census data to show that young men leave rural areas of the former Transkei for urban areas, and that they do so in search of employment (their analysis does not cover women). These men are not initially employed in rural areas, despite having some secondary schooling. Ardington, Case and Hosegood (2009) show that large cash transfers (pensions) to rural households in a former homeland area of KZN facilitate an increase in employment of prime-age adults, particularly of women. They show that this extra household income affects employment through the channel of financing migration for work. Hence, outmigration of people without jobs could be higher from steeper than flatter areas in my sample, but this would not explain the employment effects I estimate in the data.

The results for population growth and composition change in Table 9 hint at two additional ways that electrification of rural households may affect labor markets. Electrification appears to encourage people to relocate and may prevent the outflow of individuals from rural areas. A general equilibrium approach, as well as a richer dataset linking migrants to places of origin and destination, would be required to understand these effects more fully. However, given the profile of out-migrants and the results for incumbent-only employment rates, we can conclude that even this type of migration in response to electrification cannot account for all of the employment effects of electrification documented in earlier sections.

7 Conclusion

This paper uses the mass roll-out of household electrification in South Africa to measure the direct effects of public infrastructure on employment in rural labor markets and to investigate the mechanisms through which these effects operate. Addressing endogenous placement of infrastructure and confounding trends using two different identification strategies, I show that employment grows in places that get new access to electricity. Results from aggregate Census data combined with administrative and spatial data on electricity project roll-out indicate large increases in the use of electric lighting and cooking, and reductions in wood-fueled cooking over a five-year period, as well as a 9.5 percentage point increase in female employment. Further evidence from household-level surveys points towards employment growth on the extensive and intensive margins for women, and possibly for men (although effect sizes are large for men, they are not significant at conventional levels). The fact that

Force Survey to show that out-migrants from rural KZN have significantly higher levels of education than incumbents, yet significantly lower rates of employment than incumbents.

female wages fall, while male earnings rise with no significant change in male wages provides additional evidence that electrification did not spark large increases in the demand for labor through rural industrialization. While electrification of households changed the technology of home production and likely had an effect on female labor supply, the evidence presented here cannot rule out that electricity also altered the types of feasible market activities for all adults. Since similar employment effects for men and women cannot be rejected under either the IV strategy or the MD-FE/MD-trends strategy, it is likely that electrification does not exclusively operate on rural labor markets through the mechanisms of releasing time from home production.

The final result in the paper highlights the challenge that migration presents for research into the effects of infrastructure roll-out. Although migration potentially confounds labor market effects, I showed that electrification raised employment of incumbent women, separately from any in-migration response and argued that the profile of out-migrants works against the outflow of individuals explaining all of the electrification effects. These results raise interesting questions about how infrastructure-building could transform rural communities into more urban entities, either by stimulating in-migration or stemming the tide of out-migration. Addressing such questions successfully is likely to require a general equilibrium approach that is beyond the scope of this paper.

This paper presents some of the first pieces of evidence on the impact of infrastructure for rural electrification on labor markets in a developing country. Regardless of the mechanism, electrification enabled South Africans living in rural areas to increase their participation in modern labor markets. More generally, it highlights the importance of measuring employment effects in infrastructure evaluations. I emphasize the importance of interpreting the measured effects of electrification within the context of existing economic conditions - in the case of South Africa, economic adjustments after the end of apartheid. Using newly linked Census, administrative and spatial data and alternative identification strategies, the paper also provides an example of how we might study other networked infrastructure roll-outs that are inherently difficult to randomize. Collecting project and spatial data from implementing agencies is often feasible, and may generate more variation in actual conditions than legal changes would, in institutionally weak environments. Combining empirical approaches and data sources, each with their own strengths and weaknesses, is also potentially useful for dealing with the multiple biases that make it challenging to identify the effects of infrastructure and for unpacking some of the mechanisms through which infrastructure affects labor markets.

References

Akee, Richard, "Road to Palau," 2006. IZA Working Paper No. 2452.

Ardington, Cally and Frances Lund, "Employment status, security, and the management of risk: a study of workers in Kwamsane, KwaZulu-Natal," 2006. SALDRU Working Paper Series No. 06/02.

___ **, Anne Case, and Vicky Hosegood**, "Labor supply responses to large social transfers: Longitudinal evidence from South Africa," *American Economic Journal: Applied Economics*, 2009, *1* (1).

Ardington, Elizabeth and Frances Lund, *Land, Labour and Livelihoods in Rural South Africa*, Vol. 2: KwaZulu-Natal and Northern Province, Indicator Press, 1996.

Aschauer, David Alan, "Is public expenditure productive?," *Journal of Monetary Economics*, 1989, *23*, 177–200.

Bailey, Martha and William Collins, "The economic and demographic effects of household electrification,1920 to 1960," March 2006. Mimeo.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "Railroad to success: The effect of infrastructure on economic growth," 2007. Mimeo, Department of Economics, MIT.

___ **, Sebastian Galiani, James Levinsohn, and Ingrid Woolard**, "Why has unemployment risen in the new South Africa?," 2007. NBER Working Paper No. 13167.

Barnard, Jenny, "An Investigation into Using GIS in Electrification and Network Planning in Rural Kwazulu-Natal," Technical Report, GISDevelopment.net $http://www.gisdevelopment.net/application/utility/power/maf$0630.$htm$ 2006.

Beall, Robert T., "United States Department of Agriculture Yearbook: Rural Electrification," Technical Report, United States Department of Agriculture 1940.

Becker, Gary, "A Theory of the Allocation of Time," *Economic Journal*, 1965, *75* (1), 493–517.

Budlender, Debbie, Ntebaleng Chobokoane, and Yandiswa Mpetsheni, "A survey of time use: How South Africans spend their time," Technical Report, Statistics South Africa 2001.

Burger, Rulof, Murray Leibbrandt, Cecil Mlatsheni, and Servaas van der Berg, "Migration and the changing rural-urban interface in South Africa: What can we learn from census and survey data?," 2003. South African Labour and Development Research Unit.

Canning, David, "A Database of World Infrastructure Stocks," 1998. World Bank Policy Research Working Paper No. 1929.

Cattaneo, Matias, Sebastian Galiani, Paul Gertler, Sebastian Martinez, and Rocio Titiunik, "Housing, health and happiness," 2007. World Bank Policy Research Working Paper No. 4214.

Charmes, Jacques, "Chapter 3," in Mark Blackden and Quentin Wodon, eds., *Gender, time use and poverty in sub-Saharan Africa*, Washington, D.C.: The World Bank, 2005.

Chernozhukov, Victor and Christian Hansen, "The reduced form: A simple approach to weak inference," 2007. University of Chicago, Graduate School of Business.

Christopher, A.J., *The Geography of Apartheid*, Johannesburg: Witwatersrand University Press, 2001.

Coen-Pirani, Daniel, Alexis Leon, and Steven Lugauer, "The effect of household appliances on female labor force participation: Evidence from micro data," March 2008. Carnegie Mellon University.

Conley, Timothy, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, September 1999, *92* (1), 1–45.

Cross, Catherine, Tobias Mngadi, and Themba Mbhele, "Infrastructure, poverty and development in KwaZulu-Natal," *Development Southern Africa*, Summer 1998, *15* (4).

Cutler, Grant and David Miller, "The role of public health improvements in health advances: The twentieth-century United States," *Demography*, February 2005, *42* (1), 1–22.

Duflo, Esther and Rohini Pande, "Dams," *Quarterly Journal of Economics*, May 2007, *122* (2), 601–646.

EnergyNet Limited, "Rural Energy 21: Bringing Power to Africa," 2004. $http://www.energynet.co.uk/Infosite/images/RE21report2004.pdf$.

Eskom, "Eskom Annual Report," Technical Report, Eskom 1996.

——, "Eskom Annual Report," Technical Report, Eskom 1997.

——, "Eskom Annual Report," Technical Report, Eskom 1998.

——, "Eskom Annual Report," Technical Report, Eskom 1999.

Fedderke, Johannes and Zeljko Bogetic, "Infrastructure and growth in South Africa: Direct and indirect productivity impacts of nineteen infrastructure measures," 2006. World Bank Working Paper No. 3989.

Food and Agriculture Organisation, "Topological classifications," Technical Report, FAO 1998.

Gaunt, Trevor, "Electrification technology and processes to meet economic and social objectives in Southern Africa." PhD dissertation, University of Cape Town 2003.

Goldin, Claudia, "The U-Shaped female labor force function in economic development and economic history," April 1996. NBER Working Paper No. 4707.

—— **and Laurence Katz**, "Career and marriage in the age of the pill," *American Economic Review Papers and Proceedings of the One Hundred Twelfth Annual Meeting of the American Economic Association*, 2000, (XC), 461–465.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu, "Engines of Liberation," *Review of Economic Studies*, 2005, *72*, 109–133.

Gronau, Reuben, "Home production: A survey," in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*, Vol. I, Elsevier Science Publishers, 1986.

Jimenez, Emmanuel, "Human and physical infrastructure: public investment and pricing policies in developing countries," in Jere Behrman and T. N. Srinivasan, eds., *Handbook of Development Economics*, Vol. 3B, Elsevier Science Publishers, 1995, chapter 43, pp. 2774–2836.

Klasen, Stephan and Ingrid Woolard, "Surviving Unemployment Without State Support: Unemployment and Household Formation in South Africa," *Journal of African Economies*, 2009, *18* (1), 1–51.

Loshkin, Michael and Ruslan Yemtsov, "Has rural infrastructure rehabilitation in Georgia helped the poor?," *World Bank Economic Review*, August 2005, *19* (2), 311–333.

Mammen, Kristin and Christina Paxson, "Women's work and economic development," *Journal of Economic Perspectives*, 2000, *14* (4), 141–164.

Mikusheva, Anna and Brian Poi, "Tests and confidence sets with correct size in the simultaneous equations model with potentially weak instruments," *The Stata Journal*, 2006, (1), 1–11.

Nunn, Nathan and Diego Puga, "Ruggedness: The blessing of bad geography in Africa," March 2007. Harvard University.

Posel, Dori and Daniela Casale, "'Two million net new jobs': A reconsideration of the rise in employment in South Africa, 1995-2003," *South African Journal of Economics*, 2004, *72* (5).

___ **, James A. Fairburn, and Frances Lund**, "Labour migration and households: A reconsideration of the effects of the social pension on labour supply in South Africa," *Economic Modelling*, 2006, *23*, 836–853.

Rud, Juan Pablo, "Electricity provision and industrial development: Evidence from India," 2009. Memo, LSE.

Saghir, Jamal, "Energy and Poverty: Myths, Links and Poverty Issues," Technical Report, The World Bank 2005. Energy Working Notes No 4.

Schirmer, Stefan and Nicholas Vink, *The decline of the South African economy*, Edward Elgar, 2002.

South African Department of Minerals and Energy, "Electricity Supply Guide," Technical Report, Department of Minerals and Energy 2004.

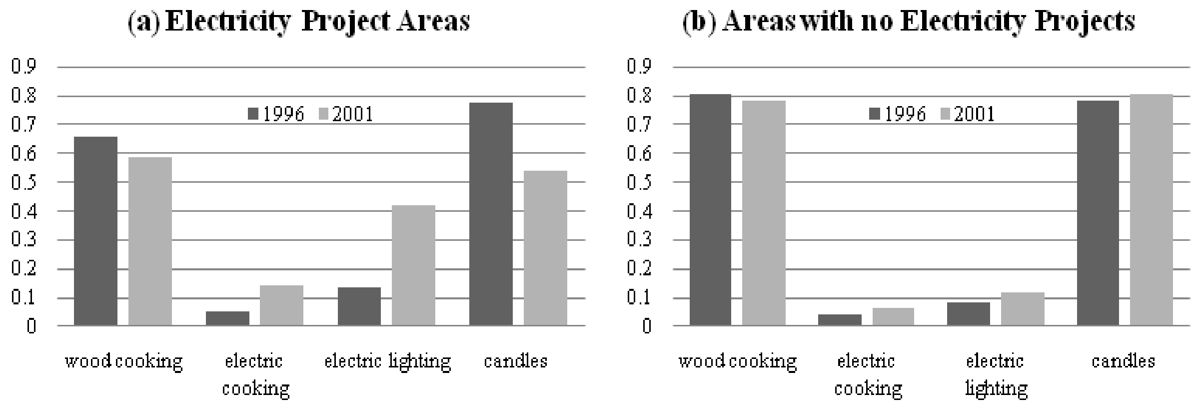
Standing, Guy, John Sender, and Jeremy Weeks, *Restructuring the Labour Market: The South African Challenge*, Geneva: International Labour Office, 1996.

The World Bank, *World Development Report 1994: Infrastructure for Development*, The World Bank, 1994.

___ , "Clean energy for development investment framework: The World Bank Group action plan," Technical Report, The World Bank 2007.

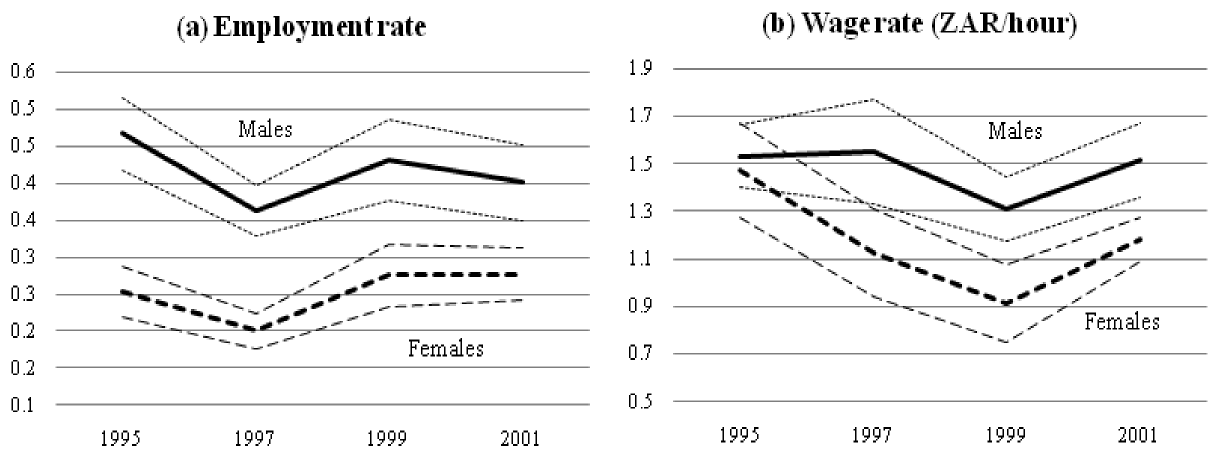
West, Nicola, Barry Dwolatzky, and Alan Meyer, "Terrain-based routing of distribution cables," *IEEE Computer Applications in Power*, January 1997.

Figure 1: Changing home production techniques by electricity project areas



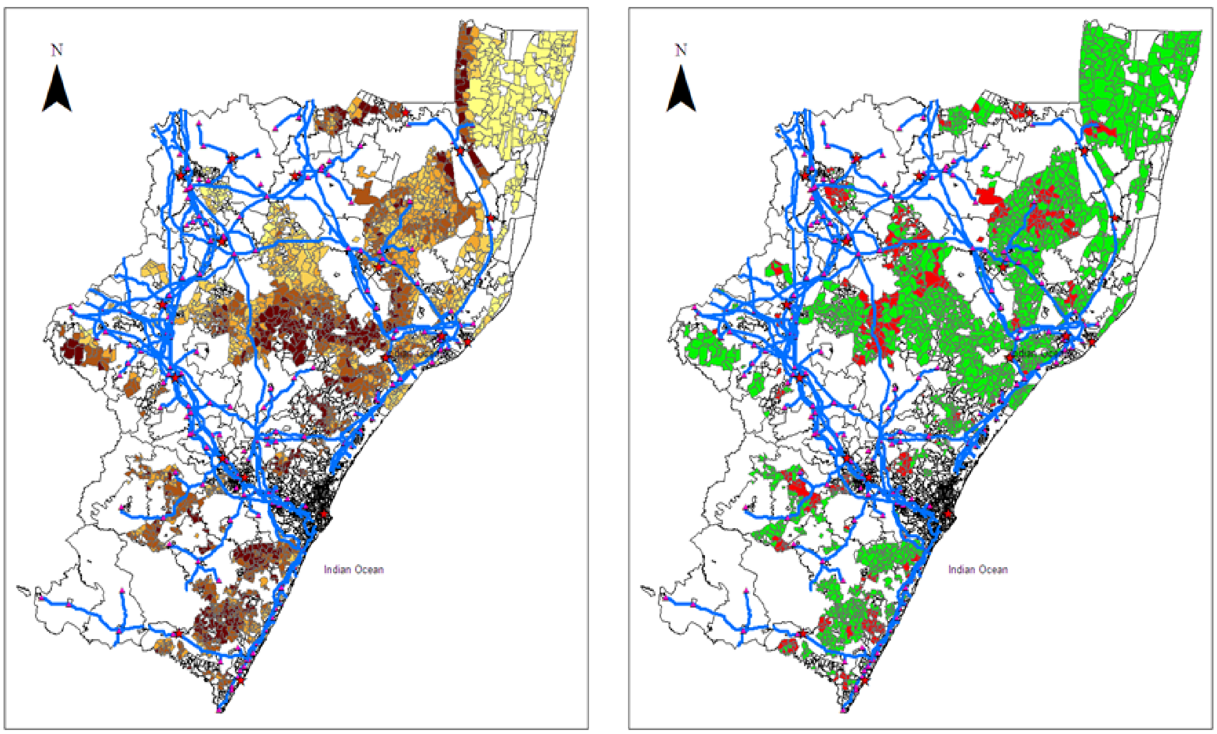
Bar graph shows fraction of households reporting main source of fuel for cooking and lighting as wood, electricity or candles in the 1996 and 2001 Census. Sample includes all households in rural KZN that are included in main analysis sample.

Figure 2: Employment rates and hourly wages over time by gender



Figures show fraction of adult African men and women employed and average hourly wage rate for the employed, using data from October Household Surveys 1995, 1997, 1999 and the September Labor Force Survey 2001. Sample includes individuals living in rural KZN (not just tribal areas). Dashed lines are 95% confidence intervals. The unit of observation is the individual.

Figure 3: Spatial distribution of gradient and of electricity project areas in KwaZulu-Natal, South Africa



Communities included in the sample are shaded (N=1,816). Thick lines depict electricity grid lines in 1996, triangles are electricity substations in 1996 and stars represent towns. Gradient is depicted in the figure on the left: steeper areas are shaded dark, flatter areas are shaded light. Electricity Project areas are depicted in the figure on the right: lighter shaded areas are electrified after 2001 or not at all.

Table 1: Baseline Community Variables by Electrification Project Status and Gradient

Covariates	All	Eskom Project	No Eskom Project	Difference (3) - (2)	Difference by gradient Without controls	With controls
	(1)	(2)	(3)	(4)	(5)	(6)
Poverty rate (<ZAR5,600 annual HH income)	0.61	0.59	0.61	-0.024**	0.00	0.002
	(0.19)	(0.17)	(0.20)	(0.01)	(0.00)	(0.00)
Fracton female-headed households	0.55	0.55	0.55	0.00	0.005***	0.001
	(0.13)	(0.12)	(0.13)	(0.01)	(0.00)	(0.00)
Adult sex ratio ($N_{females}/N_{males}$)	1.48	1.41	1.49	-0.080***	0.011***	0.004**
	(0.28)	(0.25)	(0.29)	(0.02)	(0.00)	(0.00)
Fraction Indian and white adults*10	0.00	0.00	0.00	0.00	0.00	0.000
	(0.01)	(0.00)	(0.01)	(0.00)	(0.000)	(0.00)
Kilometers to road	37.95	35.62	38.54	-2.917**	-0.201*	-0.156
	(24.57)	(24.18)	(24.64)	(1.44)	(0.12)	(0.18)
Kilometers to town	38.57	36.34	39.13	-2.790***	0.278***	0.180
	(18.12)	(15.34)	(18.72)	(1.06)	(0.09)	(0.13)
Fraction men with high school	0.06	0.08	0.06	0.016***	-0.002***	-0.003**
	(0.05)	(0.05)	(0.05)	(0.00)	(0.000)	(0.00)
Fraction women with high school	0.07	0.08	0.06	0.020***	-0.002***	0.000
	(0.05)	(0.05)	(0.05)	(0.00)	(0.000)	(0.00)
Households per km^2	22.05	32.56	19.41	13.152***	-0.523***	-0.945***
	(30.48)	(49.31)	(22.75)	(1.76)	(0.15)	(0.30)
Kilometers from the grid	19.06	15.75	19.89	-4.139***	-0.235***	0.029
	(13.32)	(10.20)	(13.88)	(0.77)	(0.06)	(0.12)
Land gradient - mean	10.10	9.12	10.35	-1.232***		
	(4.89)	(4.21)	(5.02)	(0.29)		
N communities	1,816	365	1,451	1,816	1,816	1,816

Table shows means (s.d.) and differences in means for community covariates. Variables measured in 1996. Columns (5) and (6) show coefficients from regressions of each covariate on gradient, controlling for all other covariates and district fixed effects (in column (6)). Differences significant at p<0.01***, p<0.05** or p<0.1* level. See Data Appendix for details of variable construction. Main control variables are the first ten covariates. For these, the Bonferroni joint test of significance requires p<0.05/10=0.005 to reject the null of all coefficients zero at a 5% level of significance; or p<0.01/10=0.001 to reject at the 1% level.

Table 2: Average community-level employment in 1996 and 2001

	Year	Mean	Eskom project between 1996 and 2001	No Eskom project between 1996 and 2001	Difference: (3)-(2)
		(1)	(2)	(3)	(4)
Female employment rate	1996	0.07	0.09	0.06	0.020***
		(0.08)	(0.07)	(0.08)	(0.00)
	2001	0.07	0.08	0.06	0.017***
		(0.07)	(0.07)	(0.07)	(0.00)
	Δ_t	0.000	-0.003	0.001	-0.004
		(0.002)	(0.005)	(0.00)	(0.00)
Male employment rate	1996	0.14	0.16	0.13	0.031***
		(0.11)	(0.11)	(0.11)	(0.01)
	2001	0.10	0.11	0.10	0.014**
		(0.09)	(0.09)	(0.09)	(0.01)
	Δ_τ	-0.04***	-0.050***	-0.033***	-0.017***
		(0.00)	(0.01)	(0.00)	(0.01)
N		1,816	365	1,451	

Table shows means (s.d.) in columns (1) - (3), and differences in means (and standard errors of the differences) in column (4) for employment rates and population totals at the community level. Differences within communities over time are shown in Δ_t rows. Differences significant at p<0.01***, p<0.05** or p<0.1* level.

Table 3: Assignment to Eskom Project: First stage results

	Binary dependent variable: Eskom Project = [1 or 0]			
	(1)	(2)	(3)	(4)
Gradient*10	-0.083**	-0.075**	-0.078***	-0.077***
	(0.040)	(0.034)	(0.027)	(0.027)
Kilometers to grid*10		-0.040*	-0.012	-0.011
		(0.021)	(0.023)	(0.023)
Household density*10		0.017***	0.012**	0.013**
		(0.004)	(0.006)	(0.006)
Poverty rate		0.023	0.019	0.017
		(0.069)	(0.070)	(0.069)
Adult sex ratio		0.393***	0.165	0.155
		(0.120)	(0.107)	(0.107)
Fraction female-headed hh's		-0.173***	-0.130***	-0.121***
		(0.052)	(0.042)	(0.042)
Fraction Indian and white adults*10		-1.236***	-1.116**	-1.105**
		(0.401)	(0.459)	(0.452)
Kilometers to road*10		0.003	-0.010	-0.010
		(0.009)	(0.010)	(0.010)
Kilometers to town*10		0.016	0.008	0.008
		(0.015)	(0.015)	(0.016)
Fraction men with high school		-0.269	-0.185	-0.152
		(0.500)	(0.411)	(0.417)
Fraction women with high school		1.046**	0.965**	0.984**
		(0.475)	(0.413)	(0.409)
Change in water access				0.012
				(0.048)
Change in toilet access				0.155
				(0.104)
District Fixed Effects	N	N	Y	Y
Sample	All	All	All	All
Mean of outcome variable	0.20	0.20	0.20	0.20
N	1,816	1,816	1,816	1,816
R^2	0.01	0.07	0.18	0.18
F-statistic on instrument	4.20	4.87	8.34	8.26
Probability>F	0.04	0.03	0.00	0.00

Table presents coefficients from OLS regression of Eskom Project indicator on community covariates measured in 1996. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level. Ten district fixed-effects included in columns (3) and (4). Change in fraction of households with access to water and flush toilet measured between 1996 and 2001.

Table 4: Effects of electrification on employment: Census community data

	OLS	OLS	OLS	OLS	OLS	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A: Δ_t Female Employment									
Eskom Project	-0.004	-0.001	0.000	-0.001		0.025	0.074	0.090*	0.095*
	(0.005)	(0.005)	(0.005)	(0.005)		(0.045)	(0.060)	(0.055)	(0.055)
AR 95% C.I.								*[0.05;0.3]*	*[0.05;0.3]*
Poverty rate		0.029***	0.033***	0.031***	0.032***		0.027**	0.032**	0.031**
		(0.011)	(0.010)	(0.010)	(0.010)		(0.012)	(0.013)	(0.013)
Fraction female-headed hhs		0.042**	0.051***	0.047**	0.048**		0.014	0.036	0.033
		(0.019)	(0.019)	(0.020)	(0.020)		(0.031)	(0.026)	(0.026)
Adult sex ratio ($N^{females}/N^{males}$)		0.019**	0.017**	0.020***	0.021***		0.033**	0.029**	0.032***
		(0.009)	(0.008)	(0.007)	(0.007)		(0.014)	(0.012)	(0.012)
Gradient*10					-0.007**				
					(0.003)				
B: Δ_t Male Employment Rate									
Eskom Project	-0.017**	-0.015***	-0.009	-0.010*		-0.063	0.069	0.033	0.035
	(0.007)	(0.006)	(0.006)	(0.006)		(0.073)	(0.082)	(0.064)	(0.066)
AR 95% C.I.								*[-0.05;0.25]*	*[-0.05;0.25]*
Poverty rate		0.062***	0.064***	0.063***	0.063***		0.059***	0.064***	0.062***
		(0.020)	(0.018)	(0.018)	(0.018)		(0.022)	(0.019)	(0.019)
Fraction female-headed hhs		0.217***	0.233***	0.227***	0.225***		0.187***	0.227***	0.220***
		(0.029)	(0.030)	(0.030)	(0.030)		(0.042)	(0.034)	(0.034)
Adult sex ratio ($N^{females}/N^{males}$)		0.018*	0.012	0.017	0.019*		0.034*	0.018	0.023
		(0.011)	(0.011)	(0.011)	(0.011)		(0.019)	(0.015)	(0.015)
Gradient*10					-0.003				
					(0.005)				
Baseline community controls	N	Y	Y	Y	Y	N	Y	Y	Y
District Fixed Effects	N	N	Y	Y	Y	N	N	Y	Y
Controls for Δ_t in other services	N	N	N	Y	Y	N	N	N	Y
N	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816

Table shows OLS and IV regression coefficients for the outcome change in employment rate of African adults. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level. Eskom project is instrumented for using mean community land gradient. See Table 3 for full list of control variables. Ten district fixed effects included in columns (3),(4),(5),(8) and (9). The last two columns provide confidence intervals from the Anderson-Rubin test for the coefficient on Eskom Project. The AR test is robust to weak instruments and is implemented to be robust to heteroscedasticity.

Table 5: Placebo experiment and reduced form for female employers: Census community data

	Placebo experiment: Δ_t in female employment	Growth in major sources of female employment	
		Δ_t Schools	Δ_t Indian and White adults
	OLS	OLS	OLS
	(1)	(2)	(3)
Gradient*10	-0.001	0.007	0.000
	(0.001)	(0.028)	(0.000)
Sample	Areas electrified before 1996	Full sample	Full sample
N	373	1,816	1,816
R^2	0.11	0.06	0.04

Each column shows coefficients from OLS regressions of outcome variables on community gradient and all community-level controls as in Table 3. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level. In column (1), sample is restricted to areas that had electricity projects prior to 1996. In column (2), the outcome variable is the change in the number of schools in a community between 1996 and 2001. In column (3), the outcome is the change in the fraction of Indian and white adults in the community between 1996 and 2001; the fraction of Indian/white adults in the community is excluded from this regression.

Table 6: Employment, Hours of work, Wages and Earnings for Africans in Rural KZN 1995-2001: Household survey data

	Females		Males		Females		Males	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A: Employment [1/0]				*B: Usual weekly hours of work*			
MD Electrification Rate	0.126**	0.128	0.090	0.134	6.646***	8.920	5.671**	13.090
	(0.058)	(0.149)	(0.077)	(0.164)	(1.771)	(6.634)	(2.597)	(12.947)
Trend (1995-2001)	-0.010	0.046**	-0.051***	-0.075***	-0.407	-0.588	-0.322	-1.424
	(0.012)	(0.020)	(0.012)	(0.022)	(0.491)	(0.872)	(0.620)	(1.701)
N	152	152	152	152	151	151	151	151
Mean of outcome	0.25	0.25	0.42	0.42	42.82	42.82	46.94	46.94
R^2	0.06	0.63	0.09	0.76	0.06	0.42	0.03	0.45
	C: Log weekly wages				*D: Log monthly earnings*			
MD Electrification Rate	-0.148	-1.380	0.101	0.171	-0.070	-0.616	0.414**	1.107**
	(0.253)	(1.046)	(0.211)	(0.483)	(0.225)	(0.995)	(0.191)	(0.477)
Trend (1995-2001)	-0.079***	0.132	-0.027	0.077	-0.091**	-0.065	-0.047	-0.085
	(0.030)	(0.137)	(0.032)	(0.063)	(0.037)	(0.131)	(0.033)	(0.063)
N	146	146	148	148	146	146	148	148
Mean of outcome	1.17	1.17	1.49	1.49	6.42	6.42	6.80	6.80
R^2	0.03	0.52	0.00	0.51	0.03	0.52	0.05	0.57

Columns (1), (3), (5) and (7) show coefficients from OLS regressions of magisterial district (MD) residuals on MD electrification rates, a linear time trend and a constant. Columns (2), (4), (6) and (8) show coefficients from the same regressions, including MD Fixed Effects and MD specific trends. Unit of observation is the MD-year. Robust standard errors, clustered at the MD level. Significant at $p<0.01$***, $p<0.05$** or $p<0.1$* level. Panel C and D regressions exclude MDs in which no-one reports positive earnings. Data are from October Household Surveys 1995, 1997 and 1999 and the September Labor Force Survey 2001. Mean MD electrification rate is 0.3 and the average change between 1995-2001 is 0.15.

Table 7: Effects of Electricity Projects on Household Energy Sources and Other Household Services

Outcome is Δ_t in:	OLS No controls (1)	OLS Controls (2)	IV No controls (4)	IV Controls (5)
(1) Lighting with electricity	0.251***	0.239***	0.577***	0.658***
Mean: 0.80	(0.032)	(0.031)	(0.188)	(0.144)
(2) Cooking with wood	-0.045***	-0.039***	-0.266	-0.275*
Mean: -0.035	(0.012)	(0.012)	(0.179)	(0.147)
(3) Cooking with electricity	0.068***	0.056***	0.250**	0.228**
Mean: 0.037	(0.009)	(0.009)	(0.107)	(0.101)
(4) Water nearby	-0.029	0.005	-0.483*	-0.372
Mean: 0.007	(0.029)	(0.024)	(0.249)	(0.248)
(5) Flush toilet	0.003	0.008	0.018	0.067
Mean: 0.03	(0.006)	(0.005)	(0.069)	(0.068)

Each cell in the table presents the Eskom Project coefficient (and s.e.) from an OLS or IV regression of the dependent variable on Eskom Project indicator and (in columns 2 and 4), all control variables described in Table 3. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level. Outcome variables measure the change in fraction of households using different energy sources or with access to basic services. Change in water (toilet) access excluded from the set of controls in rows (4) and (5). Each regression contains N=1,816 except for change in fraction of households using wood; I set 9 observations to missing (rather than to zero).

Table 8: Testing for Spillovers by Excluding Adjacent Areas without Electricity Projects

Outcome is Δ_t Female Employment	Coefficient on Eskom Project indicator		
	OLS	IV	N
	(1)	(2)	(3)
Panel A			
Full sample	-0.001	0.095*	1,816
	(0.005)	(0.055)	
Panel B			
Sample excludes non-project areas within 1 km of project area	-0.004	0.076	1,205
	(0.006)	(0.057)	
Panel C			
Sample excludes non-project areas within 5 km of project area	-0.003	0.069	840
	(0.008)	(0.077)	

Each cell in columns (1) and (2) shows the coefficient (standard error) on the Eskom Project indicator from regressions of the change in female employment rates for different subsamples of the data. All controls as described in Table 3 included. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level. Panel A reproduces the main result from the full sample in Table 5; Panel B and C restrict the sample to exclude non-project communities that are within a 1km or 5km radius of any project community.

Table 9: Effects of Electrification on Population Growth, Skill Composition of Labor Force and Employment of Incumbants

	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	*Δ_t Log Population*		*Δ_t Fraction Females with Matric*		*Δ_t Fraction Males with Matric*	
Eskom Project	0.171***	3.897***	-0.001	0.095*	-0.010*	0.035
	(0.045)	(1.427)	(0.005)	(0.055)	(0.006)	(0.066)
N	1,816	1,816	1,816	1,816	1,816	1,816
Panel B	*Δ_t Log Non-inmigrant Population*		*Δ_t Female Employment: Excluding In-Migrants*		*Δ_t Male Employment: Excluding In-Migrants*	
Eskom Project	0.181***	4.349***	0.000	0.116*	-0.008	0.086
	(0.048)	(1.586)	(0.005)	(0.069)	(0.005)	(0.069)
N	1,816	1,816	1,816	1,816	1,816	1,816

Each cell shows coefficient (standard error) on Eskom Project indicator from OLS or IV regressions of outcomes on all controls as in Table 3. Dependent variable in Panel A columns (1)-(2) is change in log African population. In Panel A columns (3)-(6), it is the change in fraction of women or men that have a completed high school education. In Panel B, columns (1) and (2) is it the change in log African non-inmigrant population where inmigrants have been subtracted from the total number of adults in the community in each year. In columns (3)-(6) of Panel B, the outcomes are change in female and male employment rates where the employment variables exclude the number of inmigrants to each community in each year. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level. Regressions in Panel A columns (3)-(6) do not include controls for baseline fraction of women or men with completed high school.

NOT FOR PUBLICATION

Web Appendix 1: Data description, variable construction and sample selection

This appendix describes all of the data sources used in the paper.

1 Census data

Census community data 1996 and 2001: 100% sample obtained from Statistics South Africa. Census is adjusted for undercount after enumeration.[1]. Data are provided at an aggregated enumeration area level in 1996 and at an aggregated sub-place level in 2001.

Variables in the Census include: counts of employment, population, levels of educational attainment and recent in-migrant status by sex, race and age group; counts of households, female-headed households, and households living below a poverty line (demarcated by annual household income of ZAR 6,000 or less); counts of households using different sources of fuel for lighting and counts of households with access to different types of water and sanitation facilities. Statistics South Africa also provided me with counts of households using different fuels for cooking at the enumeration area (1996) and sub-place level (2001). A limited set of cross-tabulated variable counts are also available in these data.

Employment variables in the Census: As in most Census data, measures of employment are broad. In 1996, adults are asked: 'Does the person work?' Activities listed as work include 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. I define everyone answering yes to this question as employed, else not employed.

In 2001, adults were asked: 'Did the person do any work for pay, profit or family gain for one hour or more?' Possible responses were: yes (formal, registered, non-farming), yes (informal, unregistered, non-farming), yes (farming) and no (did not have work). Everyone who answers yes to this question is defined as employed, else not employed.

Questions about employment are similar across Census waves, although the 2001 employment definition is somewhat broader than the 1996 variable, describing individuals who

[1]Personal Communication with Piet Alberts, Senior Statistician in the Census department of Statistics South Africa, May 2007

work for even one hour per week as employed. Since the main outcome variable is the change in employment rate, these differences will only be problematic if reported part-time work differentially contributes to new employment with lower gradient.

Creating the Census panel of communities: The 2001 Census geography is hierarchically ordered as follows, from largest to smallest unit:

- District: represents a local labor market area in KwaZulu-Natal, containing between 30,000 and 50,000 households.
- Main place or sub-districts: correspond to groupings of towns and surrounding areas.
- Community or sub-places: the lowest unit of observation in the 2001 Census data. Average community size is small: between 200 and 250 households on average.

Boundaries for communities from the 2001 Census define the main unit of analysis. I aggregate the 1996 (smaller) areas up to the (larger) 2001 boundaries.[2] The matched identifiers from this panel of areas are used to extract Census aggregate data in 1996 and 2001. For each 1996 EA, the proportion of the EA polygon area that falls inside each 2001 community is calculated. This proportion is used as a weight to assign a proportion of the 1996 EA data to the 2001 community. The key assumption in this process is that people are uniformly distributed over 1996 EA's.

Selection of communities for sample: Within the set of 4,030 communities in KwaZulu Natal, I restricted the sample to include rural, tribal areas. Communities that were defined as national parks and mines were also excluded. This left 1,992 communities. A final exclusion of communities with fewer than 100 adults in either Census year reduced the sample further by 176 communities, leaving 1,816 in the final analysis sub-sample.

2 Household Surveys: 1995, 1997, 1999, 2001

Obtained from Statistics South Africa. Four waves of household survey data (October Household Surveys for the 1990s and the September Labor Force Survey in 2001) resembling the

[2]Statistics South Africa notes that EA boundaries should never cut across existing administrative boundaries, and all "social boundaries should be respected" (StatsSA, 2000). However, boundaries have shifted over time (Christopher, 2001). 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" (StatsSA, 2000: 21, 26).

World Bank LSMS surveys. Each wave is a nationally representative sample of individuals. The lowest level of geography that can be identified in these household surveys is the Magisterial District, of which there are 38 in rural KZN. I include all Magisterial Districts in the analysis.

Selection of individuals for inclusion in sample: I use the sample of African male and female adults (ages 15-59) living in rural KwaZulu-Natal who report information about employment as well as about hours of work and total monthly earnings. I compute hourly wages using monthly earnings and usual hours of work reports.

3 Schools Register of Needs 1995 and 2000

These data are provided by the South African Department of Education for schools in 1995 and 2000. GPS coordinates for each school are used to assign schools to Census community boundaries. Each community is assigned the total number of schools in each year as well as the change in the total number of schools over the five-year period.

4 Geographic data

Land gradient: The source for these data is the 90-meter Shuttle Radar Topography Mission (SRTM) Global Digital Elevation Model available at www.landcover.org. Digital elevation model data was used to construct measures of average land gradient for each Census community using GIS software (ArcMap 9.1). Gradient is measured in degrees from 0 (perfectly flat) to 90 degrees (perfectly vertical).

Other measures of proximity: Spatial data on Eskom's 1996 grid network (high and medium voltage lines and substations) was provided by Steven Tait at Eskom. These data were used 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 in 1996.

Census 2001 spatial data were used to create measures of the area of the sub-place. I used these area measures in conjunction with total household counts from the Census community data to create household density variables in 1996 and 2001.

5 Electricity project data

Data on Eskom projects in KwaZulu-Natal were provided by Sheila Brown at Eskom. The project list gives the number of pre-paid electricity connections per Eskom-defined area in each year from 1990 to 2007. I define the year of electrification as the year in which a community experienced a spike in household connections (concentrated project activity). Areas are 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 map the project data to a physical location using a spatial database of transformer codes linked to project codes and then merge these locations to Census boundaries.

NOT FOR PUBLICATION

Web Appendix 2: Heterogeneity in electrification effects

In this appendix, I explore the characteristics of communities that contribute the most to the main employment results from the IV strategy.

1 Heterogeneous effects related to income

As part of the South African electrification, once an area had been targeted for new access, each household received a basic connection package: an electric circuit board, a pre-payment meter, three plug points and one light bulb. Households received a default supply of 2.5 amperes or could upgrade to a 20 ampere supply for a fee of about ZAR40 (USD6.00), which most of Eskom's 3 million customers chose to do (Gaunt, 2003). Although industry experts agreed that "Electric lighting was synonymous with the roll-out", and that the NEP did reach poor households, the subsidized roll-out really changed the option to use electricity. Households were still required to pay for using the service by purchasing electricity credits loaded on to pre-paid cards. In 1999, household electricity cost \$0.039 per kilowatt hour (kWh). Estimates of load demand from Eskom reports suggest that most rural households used between 35 and 60 kWh per month, translating into energy expenses of between \$1.37 and \$2.34 per month (Gaunt, 2003), or 1.8 percent of median monthly household income in rural KZN in 1995. Because of this positive marginal cost, the poorest households are likely to have been the least responsive to the new technology in the short-run.

The main IV strategy used in the paper identifies employment effects for communities that are cheaper to electrify by virtue of having a flatter gradient. As is well known, the IV coefficient is a weighted sum of effects for different groups, each of which may be differently affected by the gradient instrument (Kling, 2001). If different groups also experience a different electrification effect, then the IV result will be driven by the groups that are weighted most heavily in the IV parameter estimate. These weights determine which group's effect contributes the most to the total measured effect in the IV regressions.

In communities with flatter gradient, female employment may be more responsive to electrification than in an average newly electrified community. One way in which marginal communities could differ from average communities is in their ability to switch home production technologies when the new service arrives. In creating the IV weights below, I investigate how much of the IV coefficient is driven by changes in communities that look as if they would be in a better position to switch to using electricity once the new connections are made. Using electricity more effectively involves buying complementary appliances so this requires focusing on heterogeneous effects of electrification by some measure of household income.

Since the Census provides only a crude measure of poverty (household income is reported in intervals not consistent over time), I combine the three poverty indicators into a poverty index and consider the characteristics of communities in each quintile of this index. To create the index, I follow Card (1995) and Kling (2001): for the sample of communities in the steepest half of the gradient distribution, I use a logit model to estimate the probability of receiving an electricity project using the baseline poverty rate, the baseline female/male sex ratio and the baseline share of female-headed households. Using coefficients from this regression, a value for every community in the sample is predicted. Each community is then assigned to a quintile of the predicted poverty index, where quintile cut-points are defined on the estimation sample only.

The graph in Web Appendix 2 Figure 1 shows the fraction of communities in each predicted poverty quintile that is electrified between 1996 and 2001, separately for communities in the flattest and steepest halves of the gradient distribution. Both lines slope upwards, indicating that areas with higher predicted values of the poverty index (i.e. richer areas) are more likely to receive an electricity project at all. The gap between the two lines shows that flatter areas are systematically more likely to be electrified than steeper areas. The middle-poorest and second-richest quintiles are most likely to have the probability of a project manipulated by the instrument, which can be seen in the larger gap between the lines occurring at these quintiles.

Some of the same information is provided in Web Appendix 2 Table 1. This table builds up the IV weights for each poverty quintile of the sample. The fraction of the sample falling within each predicted poverty quintile is presented in column (1); the variance of gradient across communities within each quintile is very similar across quintiles, as column (2) indicates. Column (3) echoes Web Appendix 2 Figure 1: there is a larger difference in the fraction of communities electrified across flat and steep areas, in the third poorest and second richest quintiles. In column (4) of that table, I compute the contribution of each quintile to the final IV estimate by calculating the relevant weight (explained in the table notes): we see from the results that middle quintile and the second richest quintile together contribute over 65 percent to the IV result.

Middle quintiles in particular may have larger employment effects because they contain households that experience larger changes in home production technology when electricity arrives compared to richer quintiles, and they are more able to effectively use the new technology than the poorest quintiles. Web Appendix 2 Table 2 shows that middle-poor areas are initially less likely to be using electricity than richer areas and are more reliant on wood for cooking (columns 1 to 3). Columns (4), (5) and (6) of this table present within-quintile reduced-form coefficients from regressions of the change in fuel use on a gradient dummy (1 is flat, 0 is steep). These columns indicate large increases in the use of electricity and large decreases in reliance on wood for cooking in flatter areas for middle-poor, second-richest and richest areas.[1] Finally, column (7) of Web Appendix 2 Table 2 indicates that the female employment result is indeed driven by women living in middle- and second-richest quintile communities: the effects for these communities are large, positive and significant and are weighted most heavily in the final IV results. The coefficients in this table are akin to reduced-form coefficients from a regression of the outcome variable on a binary version of the instrument and all controls. Dividing each coefficient by the corresponding coefficient in column (3) of Web Appendix 2 Table 1 will reproduce the IV coefficient.

2 Heterogeneous effects related to other constraints on women's time

Women who have home-production responsibilities are less likely to be able to respond to new access to electricity, even though productivity at home may be substantially enhanced by the use of electricity. For example, child-care responsibilities raise the value of a woman's time at home and in the absence of pre-school care, this value only falls when children start school. Officially, school-starting age is between ages 6 and 7 in South Africa, but enrollment only reaches 90% by around age 9 (results from 2001 10% Census micro data, not shown). Children also create work at home though, and so the more children in the house that require child-care, the more time can potentially be saved with access to a more efficient power source. It is therefore not clear whether women with younger children will supply more or less of their labor to the market, in response to new household electrification.

Census micro data from 1996 give some indication of which women are more likely to live with a child younger than age 9. Web Appendix 2 Figure 2 is a lowess-smoothed graph of the fraction of women of each age living with at least one child aged 9 or under. The graph is drawn for African women between ages 15 and 59 living in rural areas of KZN and shows a clear distribution of youngest children to households with both younger and older women.[2] After age 30 and up to about age 50, the probability of a woman living with a child who requires constant care falls substantially.

[1] This is related to the point by Greenwood et al (2005) who argue that poorer households are the last to adopt durable goods for home production.

[2] The allocation of young children to households with older women is a common pattern in South Africa, where pension-aged women care for grandchildren in skip-generation households (Case and Deaton, 1998).

To investigate whether the employment effects of household electrification are largest for this latter group of women, I redefine the outcome variable to be $y_{ajdt} = \frac{E_{ajdt}}{P_{jdt}}$, where E_{ajdt} is the number of employed women in age group a for each of nine five-year cohorts and P_{jdt} is the total adult female population in each community in each year. This definition decomposes the employment result into effects for each age cohort: the estimated coefficients sum to the main electrification coefficient in the final column of Table 4 in the main paper. Web Appendix 2 Table 3 presents OLS and IV coefficients (and robust standard errors clustered at the sub-district level) on the electrification dummy for separate regressions.[3] IV results are large and positive for each age group, but significant only for women in their thirties and late forties. Employment grows by 3 percentage points for women between the ages of 30 and 34, by 1.7 percentage points for the 35 to 39 year old group and by a smaller but still significant 1.4 percentage points for women in their late forties. Together, these age groups account for 65 percent of the total female employment result. This indicates that women in age groups in which care of young children is not a significant constraint, are those women most responsive to the arrival of electricity in the home.

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

References

Card, David, "Using geographic variation in college proximity to estimate the returns to schooling," in L.N. Christofides et al, ed., *Aspects of labour market behaviour: Essays in honor of John Vanderkamp*, Toronto: University of Toronto Press, 1995, pp. 201–221.

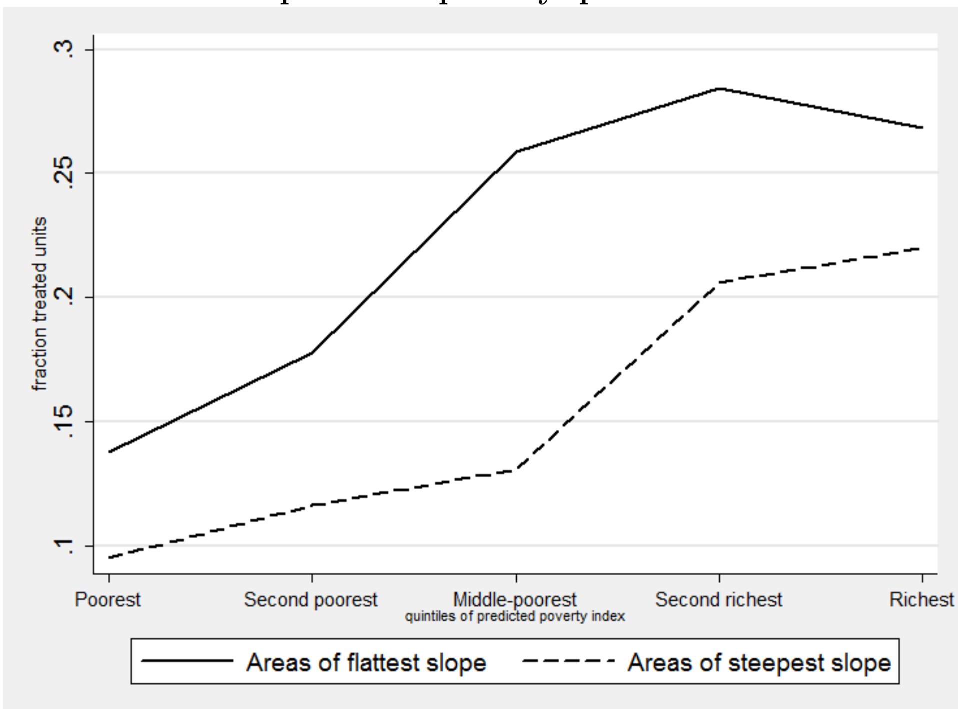
Case, Anne and Angus Deaton, "Large cash tranfers to the elderly in South Africa," *Economic Journal*, September 1998, *108* (450), 1330–1361.

Gaunt, Trevor, "Electrification technology and processes to meet economic and social objectives in Southern Africa." PhD dissertation, University of Cape Town 2003.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu, "Engines of Liberation," *Review of Economic Studies*, 2005, *72*, 109–133.

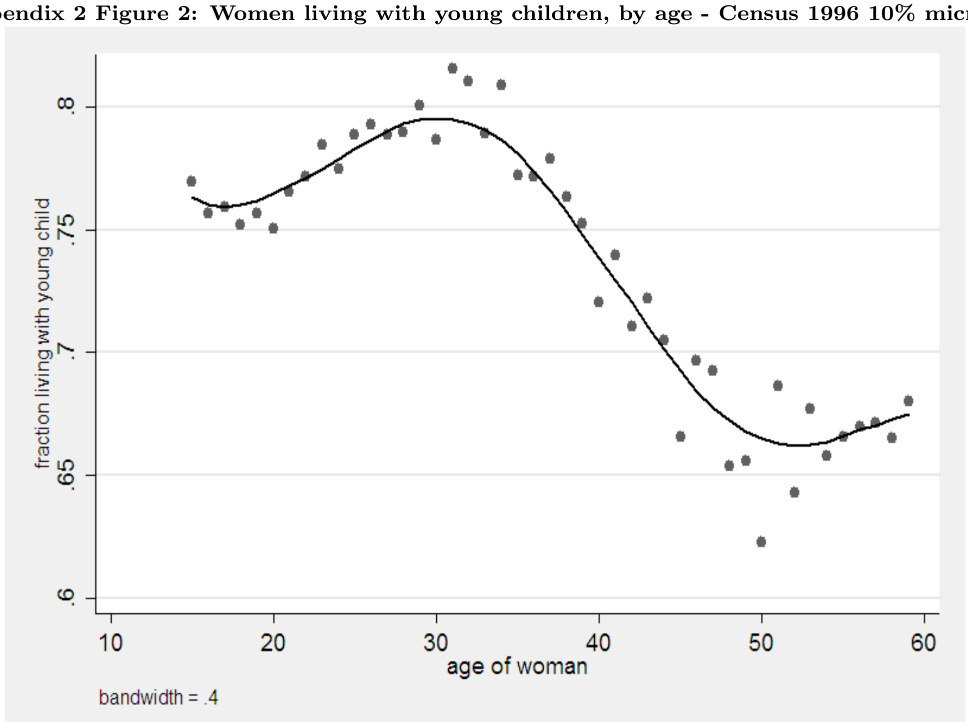
Kling, Jeffrey, "Intepreting instrumental variables estimates of the returns to schooling," *Journal of Business and Economic Statistics*, July 2001.

Web Appendix 2 Figure 1: Effect of gradient on electrification probability, by predicted poverty quintile



Lines show fraction of each predicted poverty quintile that is electrified, by top (steep) and bottom (flat) halves of the gradient distribution. See notes for Table 10 for a description of how poverty index is created. The gap between the two lines indicates at which part of the poverty index the gradient manipulates the probability of electrification the most.

Web Appendix 2 Figure 2: Women living with young children, by age - Census 1996 10% micro sample



Lowess-smoothed graph of the fraction of women of each age living with at least one child under the age of 9. Data are from the 1996 South African Census 10% micro data and include African women aged 15-59 living in rural KwaZulu-Natal. N=116,381 collapsed to 45 age-specific data points.

Web Appendix 2 Table 1: Contribution of each poverty quintile to IV results

Quintiles of predicted poverty index	Fraction of Sample in Quintile	Variance of Gradient by Quintile (λ_q)	E(Δelec|z=1, q,x) - E(Δelec|z=0, q,x)|Q)	IV weight (ω_q)
	(1)	(2)	(3)	(4)
Poorest quintile	0.16	0.20	0.01	0.04
			(0.04)	
Second poorest	0.18	0.21	0.04	0.13
			(0.04)	
Third poorest	0.21	0.21	0.10	0.34
			(0.04)	
Second richest	0.21	0.20	0.10	0.33
			(0.04)	
Richest quintile	0.23	0.20	0.05	0.16
			(0.05)	

This table follows Kling (2001) in building the weights for each poverty quintile of the sample. These weights in column (4) indicate how much each poverty quintile contributes to the overall IV results. First, predicted poverty quintiles are assigned as follows: for communities in the steepest half of the gradient distribution, I project the indicator of electricity project on to community poverty rate, the fraction of female-headed households and the female/male sex ratio. Predicted values are created for every community using these regression coefficients. Communities are assigned to quintiles, where quintile cut-points are defined by the regression sub-sample. Column (1) shows the fraction of the sample that is in each poverty quintile. Column (2) shows λq, the conditional variance of the gradient dummy (1=flat, 0=steep) within each quintile (q): predicted E(P[Z|x,q][1-P(Z|x,q)|q]). Column (3) shows the estimated difference in the fraction of communities with and without Eskom projects by gradient, within each poverty quintile and controlling for covariates: Predicted Δelec|q=predicted E(E((elec|z=1,x,q) - predicted E(elec|z=0,x,q)|q)), Each estimated coefficient in column (3) is on the interaction of the gradient dummy (1=flat, 0=steep) with each predicted quintile dummy. Column (4) assembles these pieces to create the weights: ωq= ([(1)q*(2)q*(3)q])/(Sum of q [(1)q*(2)q*(3)q]).

Web Appendix 2 Table 2: Household energy use by poverty quintile: At baseline and over time, 1996 to 2001

Quintile of predicted poverty index	Fuel Use in Home Production: Fraction using [X] in 1996			Δ_t in Fuel Use for Home Production: Within-quintile difference by gradient			Δ_t in employment by gradient	
	Electric Lighting	Electric Cooking	Wood Cooking	Electric Lighting	Electric Cooking	Wood Cooking	Females	Males
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Poorest quintile	0.02	0.01	0.90	0.00	0.002	0.000	0.001	-0.005
	(0.08)	(0.05)	(0.15)	(0.02)	(0.01)	(0.02)	(0.00)	(0.01)
Second	0.04	0.02	0.85	0.00	0.004	-0.012	0.00806*	0.001
	(0.14)	(0.08)	(0.19)	(0.02)	(0.01)	(0.01)	(0.00)	(0.01)
Third	0.07	0.03	0.81	0.04	0.0152*	-0.0193*	0.00792*	0.002
	(0.17)	(0.10)	(0.22)	(0.02)	(0.01)	(0.01)	(0.00)	(0.01)
Fourth	0.12	0.05	0.72	0.03	0.0214**	-0.0337**	0.0101*	0.002
	(0.23)	(0.12)	(0.26)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Richest quintile	0.18	0.09	0.64	0.04	0.0258**	-0.025	-0.002	-0.012
	(0.27)	(0.16)	(0.30)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)

Columns (1)-(3) present the quintile means of outcome variables in 1996, columns (4)-(8) present coefficients from regression of interactions of gradient dummy and predicted poverty quintile. Significant at p<0.01***, p<0.05** or p<0.1* level.

Web Appendix 2 Table 3: Age-specific effects of Electrification on Female Employment

Δ_t Female Employment	OLS	IV
	(1)	(2)
Ages 15-19	0.000	0.000
	(0.000)	(0.005)
Ages 20-24	0.000	0.009
	(0.001)	(0.013)
Ages 25-29	-0.001	0.015
	(0.001)	(0.012)
Ages 30-34	-0.001	0.030*
	(0.001)	(0.012)
Ages 35-39	0.000	0.017
	(0.001)	(0.013)
Ages 40-44	0.006***	0.001
	(0.002)	(0.001)
Ages 45-49	0.007	0.014*
	(0.012)	(0.008)
Ages 50-54	-0.001	-0.001
	(0.001)	(0.007)
Ages 55-59	0.001	0.004
	(0.001)	(0.006)

Each cell in the table shows the coefficient (standard error) on the Eskom Project indicator from an OLS or IV regression of the change in the age-specific female employment rate on all controls, as in Table 3. Age-specific female employment rate is measured as the fraction of employed African women of age [X] over all females, where [X] is one of nine age groups. Robust standard errors clustered at sub-district level. Significant at $p<0.01$***, $p<0.05$** or $p<0.1$* level. N=1,816 in each regression.

NOT FOR PUBLICATION

Web Appendix 3: Robustness checks

This appendix provides a set of robustness checks and additional statistical tests for the paper.

1 Controlling for political factors

I collected election outcomes data for the KZN municipalities for the first municipal elections in 2000 and matched my sample of Census communities (smaller entities) to the municipality boundaries (larger entities). Using the number of voters voting for each of 9 parties i in each municipality in the 2000 elections, I create a standard measure of political competition (see Banerjee and Somanathan, 2007), assigning to each community j the corresponding value of H_j:

$$H_j = (1 - \sum_{i=1}^{9} voteshare_j) \tag{1}$$

A higher level of H_j indicates more political competition. Results for results, controlling for political heterogeneity are presented in Web Appendix 3 Table 1 and Table 2.

- Table 1 columns (1) and (2) show that the measure of political competition predicts whether a community gets an Eskom project, but only when we do not control for district fixed effects. Once all other controls and district FE are added, the political competition measure has no predictive power in the first stage. More importantly, its inclusion does not change the impact of gradient on the probability of being allocated an Eskom project

- Table 2 columns (1) - (8) show that the inclusion of the political competition variable changes the effects of electrification on female employment only slightly. In areas with more political competition, female employment grows by 3.8 percentage points (going from no to complete competition). In the IV results, female employment is higher by 8.9 percentage points but given the reduction in sample size, this coefficient is not significantly different from zero (not all communities could be mapped to municipal boundaries). Male results are not affected by including the control for political heterogeneity

Although it would be preferable to control for earlier elections outcomes than 2000, this is not possible since the earlier election was "transitional" and those political boundaries were in flux before 2000. The exercise here indicates that while political competition may be important for employment growth, this variable is uncorrelated with gradient after controlling for all other variables and district fixed effects; and so has no substantial effect on the IV employment growth results.

2 Restricting the sample to areas without roads

I do not have access to road-building data in the province over time; only an indicator for whether a major national road runs through a community in 1996. In Web Appendix 3 Table 1 and Table 2, I present results from re-estimating the first stage assignment model and the model for employment on a smaller sample, where I omit communities with a main road running through them. The results for female employment (Table 2, columns (9) - (18)) remain large and positive, although, since the sample shrinks with the exclusion of some communities, the estimate is no longer statistically significant at conventional levels. The AR confidence interval extends from [0; 0,2].

3 Main results with corrections for spatial correlation in unobservables

To check that the main results (both coefficient estimates and statistical significance) are robust to spatial correlation in the error term, I re-estimate all regression results using the approach of Conley (1999). Results appear in Web Appendix Tables 3 to 6. In this approach, standard errors are generated using a weighted estimator, where the weights are the product of two weight functions, or kernels (one with an East-West orientation and the other with a North-South orientation). Each kernel declines linearly and is zero beyond a cutoff number. The cutoff number I choose here is 0.7 degrees (roughly 70 kilometers) and results are robust to cutoffs from 0.6 degrees to 1 degree (60 kms to 100kms).

Two points are apparent from these tables. For the chosen cutoff values, the coefficient estimates remain stable. And, standard errors are not uniformly larger when corrected for spatial correlation: sometimes they are larger and sometimes smaller than standard errors clustered at the sub-district level. The reason for this is that clustering standard errors at the sub-district level already takes account of most of the spatial correlation in errors.

Overall, the tables show that OLS and IV estimates and inference related to these estimates is robust to this alternative form of computing standard errors.

4 Testing for differences between male and female employment effects

Web Appendix 3 Table 7, I test for differences in the effect of electrification on male and female employment. I implement the test by differencing male and female outcome variables within community and then estimating the same set of OLS and IV regressions on this new variable. This test respects the correlated structure of errors across male and female regressions. Results indicate we cannot reject that new Eskom Projects had the same impact on male and female employment growth.

5 Characteristics of outmigrants compared to incumbents across high and low electrification areas, LFS 2002

The September 2002 Labor Force Survey contains a special module on migrants attached to households, from which information on outmigrants can be derived. In Web Appendix 3 Table 8, I show the fraction of people who are outmigrants from rural KZN magisterial districts as measured in these data. The table also presents mean employment rates and mean years of education for outmigrants and incumbents.

The table show differences in these summary statistics across communities with high and low rates of electrification (column 3) and indicates whether these differences are statistically significant (column 4). High electrification areas are defined as any magisterial districts in which more than 40% of the households have electric lighting: 66 percent of individuals live in such areas. The remaining 34% of individuals live in areas where less than 25 percent of households have electric lighting (there is no density in between 25 and 40 percent coverage).

Note that:

- Outmigration rates from rural KZN are high, and significantly higher in areas with low rates of electrification
- Outmigrants have higher average education than those who remain behind

- Employment rates of either group are about the same across high and low electricity districts
- Employment rates are significantly (significant at the 1% level) higher among incumbents compared to outmigrants in both low and high electrification rate districts.

References

Banerjee, Abhijit and Rohini Somanathan, "The political economy of public goods: Some evidence from India," *Journal of Development Economics*, 2007, *82*, 287314.

Conley, Timothy, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, September 1999, *92* (1), 1–45.

Web Appendix 3 Table 1: First stage OLS regressions

	Outcome is Eskom Project indicator			
	(1)	(2)	(3)	(4)
Gradient*10	-0.082**	-0.074***	-0.083**	-0.087***
	(0.003)	(0.003)	(0.003)	(0.003)
Political competition index	4.221***	1.321		
	(0.153)	(0.171)		
Other baseline controls?	Y	Y		
District Fixed Effects?	N	Y		
Sample	Full sample		Sample excluding communities with main roads	
N	1,781	1,781	1,792	1,792
R2	0.10	0.18	0.10	0.18
Mean of outcome variable	0.20	0.20	0.20	0.20
F-statistic on instrument	6.39	7.32	6.135	10.43
Probability>F	0.01	0.01	0.079	0.179

Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level. Ten district fixed-effects included in columns (2) and (4), all other controls included in each regression. Land gradient in degrees. Political competition is a measure of political heterogeneity: 1-the sum of (vote share)^2 where the sum is over all parties and elections data are from 2000 municipal elections.

Web Appendix 3 Table 2: Effects of electrification on employment: Additional controls and different subsamples

	Outcome is															
	Δ_t female employment rate				Δ_t male employment rate				Δ_t female employment rate				Δ_t male employment rate			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Eskom Project	-0.001	-0.001	0.084	0.089	-0.009	-0.010*	0.012	0.012	0.000	0.001	0.059	0.069	-0.014**	-0.009	0.056	0.019
	(0.005)	(0.005)	(0.055)	(0.057)	(0.006)	(0.006)	(0.066)	(0.069)	(0.005)	(0.005)	(0.050)	(0.045)	(0.006)	(0.006)	(0.071)	(0.060)
Political competition index	0.034*	0.038*	0.020	0.021	0.029	0.035	0.025	0.031								
	(0.021)	(0.020)	(0.030)	(0.030)	(0.026)	(0.027)	(0.030)	(0.031)								
Other baseline controls?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District Fixed effects?	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Sample	Full sample								Sample excludes communities with main roads							
N	1,781	1,781	1,781	1,781	1,781	1,781	1,781	1,781	1,792	1,792	1,792	1,792	1,792	1,792	1,792	1,792
Standard 95% C.I.	[-0.01;0.01]	[-0.01;0.01]	[-0.06;0.11]	[-0.02;0.19]	[-0.02;0]	[-0.02;0]	[-0.12;0.14]	[-0.12;0.15]	[-0.01;0.01]	[-0.01;0.01]	[-0.04;0.16]	[-0.02;0.16]	[-0.03;0]	[-0.02;0]	[-0.08;0.2]	[-0.1;0.14]
AR C.I.				[0.05;0.35]				[-0.15;0.2]				[0;0.20]				[-0.09;0.15]

Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level.

Web Appendix 3 Table 3: First stage assignment to Eskom Project: OLS results with standard errors corrected for spatial correlation

	Outcome is Eskom Project = 1			
	(1)	(2)	(3)	(4)
Gradient*10	-0.083**	-0.075**	-0.078***	-0.077***
	(0.040)	(0.034)	(0.027)	(0.027)
	[0.054]	*[0.042]*	*[0.031]*	*[0.031]*
District FE	N	N	Y	Y
Sample	All	All	All	All
Mean of Y variable	0.201	0.201	0.201	0.201
N	1,816	1,816	1,816	1,816

Table shows coefficients from OLS regression of Eskom project indicator on gradient, all other control variables and district fixed effects. Robust standard errors, clustered at the sub-place level are shown below each coefficient in (parentheses). Standard errors adjusted for spatial correlation using Conley's spatial weighting matrices and a cutoff of 0.7 degrees (about 70 kilometers) are also presented in square [brackets]. Significant at p<0.01***, p<0.05** or p<0.1* level.

Web Appendix 3 Table 4: Effects of electricity projects on household energy sources and other services, standard errors corrected for spatial correlation

	OLS			IV	
Outcome is Δ_t	No controls	Controls	Reduced form coefficient on gradient*10	No controls	Controls
	(1)	(2)	(3)	(4)	(5)
(1) Lighting with electricity	0.251***	0.239***	-0.077***	0.577***	0.658***
(Mean=0.8)	(0.032)	(0.031)	(0.017)	(0.188)	(0.144)
	[0.046]	*[0.04]*	*[0.014]*	*[0.228]*	*[0.271]*
(2) Cooking with wood	-0.045***	-0.039***	0.022**	-0.266	-0.275*
(Mean=-0.04)	(0.012)	(0.012)	(0.010)	(0.179)	(0.147)
	[0.012]	*[0.01]*	*[0.009]*	*[0.203]*	*[0.161]*
(3) Cooking with electricity	0.068***	0.056***	-0.019***	0.250**	0.228**
(Mean=0.04)	(0.009)	(0.009)	(0.006)	(0.107)	(0.101)
	[0.01]	*[0.008]*	*[0.005]*	*[0.12]*	*[0.121]*
(4) Water nearby	-0.029	0.005	0.029	-0.483*	-0.372
(Mean=0.01)	(0.029)	(0.024)	(0.018)	(0.249)	(0.248)
	[0.028]	*[0.023]*	*[0.02]*	*[0.271]*	*[0.225]*
(5) Flush toilet	0.003	0.008	-0.005	0.018	0.067
(Mean=0.03)	(0.006)	(0.005)	(0.005)	(0.069)	(0.068)
	[0.007]	*[0.006]*	*[0.005]*	*[0.075]*	*[0.061]*

Each cell contains the coefficient on Eskom Project indicator from OLS or IV regressions of dependent variable on electrification dummy; all control variables listed in Table 3 are included in columns (2) and (5). Robust standard errors, clustered at the sub-place level are shown below each coefficient in (parentheses). Standard errors adjusted for spatial correlation using Conley's spatial weighting matrices and a cutoff of 0.7 degrees (about 70 kilometers) are presented in square [brackets]. Significant at $p<0.01$***, $p<0.05$** or $p<0.1$* level. Each regression contains N=1,816 except for change in fraction of households using wood; I set 9 observations to missing (rather than to zero) for the 2001 observations.

Web Appendix 3 Table 5: Effects of electrification on female employment with standard errors corrected for spatial correlation

Outcome is Δ_t female employment rate	OLS				Reduced form	IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Eskom Project	-0.004	-0.001	0.000	-0.001		0.025	0.074	0.090*	0.095*
	(0.005)	(0.005)	(0.005)	(0.005)		(0.045)	(0.060)	(0.055)	(0.055)
	[0.005]	*[0.005]*	*[0.005]*	*[0.005]*		*[0.057]*	*[0.07]*	*[0.056]*	*[0.056]*
Gradient*10					-0.007**				
					(0.003)				
					[0.003]				
Other control variables	N	Y	Y	Y	Y	N	Y	Y	Y
District Fixed Effects	N	N	Y	Y	Y	N	N	Y	Y
N	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816

Table shows OLS and IV regression coefficients for the outcome change in employment rate of African females. Robust standard errors, clustered at the sub-place level are shown below each coefficient in (parentheses). Standard errors adjusted for spatial correlation using Conley's spatial weighting matrices and a cutoff of 0.7 degrees (about 70 kilometers) are presented in square [brackets]. Significant at p<0.01***, p<0.05** or p<0.1* level. See Table 3 for notes on additional control variables included.

Web Appendix 3 Table 6: Effects of electrification on male employment with standard errors corrected for spatial correlation

Outcome is Δ_t male employment rate	OLS				Reduced form	IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Eskom Project	-0.017**	-0.015***	-0.009	-0.010*		-0.063	0.069	0.033	0.035
	(0.007)	(0.006)	(0.006)	(0.006)		(0.073)	(0.082)	(0.064)	(0.066)
	[0.009]	*[0.005]*	*[0.006]*	*[0.006]*		*[0.099]*	*[0.111]*	*[0.064]*	*[0.065]*
Gradient*10					-0.003				
					(0.005)				
					[0.005]				
Other control variables	N	Y	Y	Y	Y	N	Y	Y	Y
District FE	N	N	Y	Y	Y	N	N	Y	Y
N	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816	1,816

Table shows OLS and IV regression coefficients for the outcome change in employment rate of African males. Robust standard errors, clustered at the sub-place level are shown below each coefficient in (parentheses). Standard errors adjusted for spatial correlation using Conley's spatial weighting matrices and a cutoff of 0.7 degrees (about 70 kilometers) are presented in square [brackets]. Significant at p<0.01***, p<0.05** or p<0.1* level. See Table 3 for notes on additional control variables included.

Web Appendix 3 Table 7: Effects of electrification on Female employment growth - Male employment growth

	Δ_t female employment rate - Δ_t male employment rate	
	OLS	IV
	(1)	(2)
Eskom Project	0.010**	0.060
	(0.004)	(0.060)
Other baseline controls?	Y	Y
District Fixed effects?	Y	Y
N	1,816	1,816
R2	0.13	0.00
Standard 95% C.I.	[0;0.02]	[-0.06;0.18]
AR C.I.		[-0.09;0.35]

Table shows coefficients from regressions of the differential difference in employment rates for women - men on all control variables and district fixed effects. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level.

Web Appendix 3 Table 8: Characteristics of outmigrants and incumbents by district-level electrification status in rural KZN: LFS 2002 data

	District has high electrificatio n rates	District has low electrificaton rates	Difference	Significant?
	(1)	(2)	(3)	(4)
Fraction of adult who outmigrate	0.39	0.43	-0.04	**
Mean yrs of education of adults remaining	7.18	6.87	0.31	***
Mean yrs of education of outmigrant adults	7.35	6.97	0.38	**
Mean employment rate among adults remaining	0.34	0.37	-0.02	
Mean employment rate among outmigrants	0.20	0.18	0.02	

Table shows descriptive statistics for adult Africans attached to households in rural KZN and enumerated in the September 2002 Labor Force Survey. Means are weighted using survey weights. Outmigrants are individuals identified by the household as belonging to the household, but who are outmigrants. There are 3,201 adults who are not migrants, and 2,146 adults who are outmigants. I define high electrification areas as any magisterial districts in which more than 40% of the households have electric lighting: 66% of individuals live in such areas. The remaining 34% of individuals live in areas where less than 25% of households have electric lighting (there is no density in between 25 and 40% coverage). Significant differences in variable means between high and low electrification areas are presented in columns (3) and (4) (t-tests).

NOT FOR PUBLICATION

Web Appendix 4: Measurement error

1 Measurement error in the Census data: Employment

The Census data undoubtedly measures employment with some error. While the employment questions are broad, the Census does not probe for employment information as the household surveys do. This section discusses the extent of this measurement error by comparing the Census data to individual level household survey data.

In Web Appendix 4 Table 1, I present population totals and employment rates for six different surveys: columns (1) and (4) present household-level data from the 1996 October Household Survey and the 2001 September Labor Force Surveys. These are the closest surveys we have to the relevant Census years and I use the weights in these surveys (constructed using the relevant Census as a benchmark) to create population totals and employment rates. In columns (2) and (4), I use the micro data from the 10% sample of the Census in 1996 and 2001 to create the same statistics using the Census weights; and in columns (3) and (6) I present the statistics taken from the 100% Census community databases in 1996 and 2001. Note that the unit of observation is the individual in columns (1), (2), (4) and (5) and the community in columns (3) and (6). Another important difference is that the individual level data in the Census and household surveys can only be restricted to African adults living in rural KZN while the Census community data can be disaggregated further to include adult Africans living in tribal areas of KZN. Tribal areas refer to the former homelands.

The difference between columns (1) and (2), and between (4) and (5) is largely the result of differences in the Census questions for employment versus the more detailed household survey questions. The difference between columns (2) and (3), and columns (5) and (6) is due to a restriction to tribal areas as well as the use of communities (rather than individuals) as the unit of observation.

The individual Census and household survey data provide population totals that are not substantially different from each other in most cases (the largest difference is in the male population total in 2001). However, employment rates are quite different across household survey and Census data. In every case, employment rates are lower in the 10% Census data compared with the 1996/2001 household survey data. The gaps also seem to be larger in 2001 than in 1996.

In addition to these differences over time in how closely the individual level data correspond, there are differences between the community data and the individual Census data. In every year, for men and for women, the Census community data present lower employment rates: between one half and one third of the employment rate is measured in the community data. A large part of the explanation for this is that the community data are restricted to tribal areas, which are not identical to all of the rural areas in the province (the Census community data does not provide a clean variable to separate rural areas into tribal/non-tribal). Hence individuals who live in rural communities with better average labor market outcomes than in the tribal, rural areas of the province are excluded from the community level data.

Web Appendix 4 Table 2 shows correlations between the individual and community Census data for different years and for men and women at the magisterial district level. The first four columns show that the community census employment data predict only a fraction of the individual census employment data and that the fraction explained for women is higher than it is for men. The final two columns show the correlation between the change in employment rates measured at the individual level and the change in employment rates measured at the community level. Again, more of the change in female employment at the individual level is predicted by the change in female employment in the community data than for men. This suggests that the Census community data may undercount male employment to a larger extent than female employment.

There are a few important points to note from Web Appendix 4 Table 1:

- The Census community data that is restricted to tribal areas under-counts employment, relative to all rural areas.
- The 2001 Census data (both individual and community data) measure lower levels of employment compared to the household survey data. This is probably due to the way the Census asked about employment in 2001: "Did you work for at least 1 hour last week?" compared with the 1996 question, "Did you work for a formal wage/salary, in informal work, or on a farm last week?" The 2001 question may not have been interpreted to include informal sector work or farm work by respondents in 2001, so the main types of employment that are under-counted in 2001 are probably these types of jobs. As long as the prevalence of these jobs is uncorrelated with gradient, then under-counting of employment in the 2001 data should not be problematic for the paper's main research design.
- Changes in employment in the Census community data more strongly predict changes in employment in the individual level data for women compared to men. This suggests that the community level data may be missing more of the employment story for men, than for women, in these areas.

- Even though the Census community data under-counts employment, the strong message from the individual level data is that there are very low levels of employment in these rural areas: under 50% of men are employed and under 30% of women are employed. These employment rates fall even further when we restrict to tribal areas of the province in using the Census community data. The low levels of employment in these areas are not an artifact of mis-measured (i.e. missing) employment.

- Finally, focusing on the occupation distribution for men and women, the individual Census data count fewer men and women employed in agriculture than the household surveys do. Agricultural employment in both of the individual sources is higher than in the Census community data - bearing in mind that the community data count people living only in tribal areas - and yet is still very low, below 10%. Regardless of which data set is considered, only a small fraction of individuals work in agriculture in the rural areas of KZN.

2 Measurement error in the electrification project variable

Since Eskom region boundaries do not line up with Census boundaries, I assign values of T_{jdt} in the following way: for any community that lies even partially inside an Eskom project area, all information from that project is assigned to that community. This means some communities are assigned full electrification status when only a fraction of households in the area are electrified. In addition, non-NEP electrification continued during this period in areas where households were willing to pay for their connections.

Measurement error in the binary project status variable could contribute to the difference between OLS and IV coefficients. OLS will underestimate the effect of electricity on outcomes when there is a negative covariance between δ_j and ΔT_{jdt} (which I have argued is likely) and when ΔT_{jdt} is measured with error. However, the valid IV that is uncorrelated with $\delta_j + \Delta\epsilon_{jt}$ will tend to be correlated with any non-classical measurement error in the binary variable ΔT_{jdt}. In this situation, even if the instrument deals with the omitted variables bias, the measurement error in ΔT_{jdt} could lead to an upwards-biased IV estimator.[1]

To get a sense of how much of the difference in OLS and IV results is due to measurement error, I restrict to samples where I expect ΔT_{jdt} to be measured with less error. The first

[1]This result is conditional on the measurement error in electrification status not being too extreme (Kane et al, 1998). See Bound and Solon (1999) and Kane, Rouse and Staiger (1998) for a discussion of what the IV estimator is consistent for in the presence of non-classical measurement error.

two columns of Web Appendix 4 Table 4 reproduce the main result for females in the full sample while columns (3) to (6) present results for successive sample limitations. To identify communities where projects had greater coverage, I exclude electrified areas with less than a 10 percent change in coverage of electric lighting, and areas where the connection rate between 1996 and 2001 was under 80 percent of households. All communities that did not have an electricity project during the period are included in all columns. Under the first restriction in columns (3) and (4), the OLS coefficient rises substantially and the IV coefficient is the same as the main result at 13 percentage points. The movement in the OLS coefficient suggests that there is some measurement error is present in the electrification variable. Columns (5) and (6) impose the second restriction. Again, the OLS estimate is large and positive and the IV result is now slightly higher than the main result (at 0.155), although neither is statistically significant due to the smaller sample size.

Although effects estimated under the OLS specification for these sub-samples are between 1 and 1.2 percentage points higher than the OLS result for the full sample, they are still well smaller than the IV results. This is evidence that measurement error in the electrification dummy alone is unable to account for the entire gap between OLS and IV estimates.

References

Bound, John and Gary Solon, "Double Trouble: on the value of twins-based estimation of the returns to schooling," *Economics of Education Review*, 1999, *18*, 169–182.

Kane, Thomas, Cecilia Elena Rouse, and Douglas Staiger, "Estimating returns to schooling when schooling is misreported," 1998. NBER Working Paper No. 7235.

Web Appendix 4 Table 1: Comparing measures of employment in the Census and October Household/Labor Force Surveys

	1996			2001		
	OHS 1996	10% Census	100% Census	LFS 2001	10% Census	100% Census
Unit and place of observation	Indiv. RURAL	Indiv. RURAL	Comm. TRIBAL	Indiv. RURAL	Indiv. RURAL	Comm. TRIBAL
Panel A: Women						
Population totals	1,398,856	1,299,475	1,290,869	1,144,854	1,479,848	1,632,826
Total employment/population	0.19	0.13	0.07	0.28	0.13	0.08
Occupational distribution						
Managers, profs, assoc. profs	0.04	0.02	0.02	0.03	0.02	0.02
Clerks	0.01	0.00	0.00	0.01	0.01	0.01
Services	0.06	0.01	0.00	0.04	0.01	0.00
Agriculture[1]	0.01	0.01	0.00	0.03	0.01	0.00
Crafters	0.04	0.01	0.00	0.07	0.01	0.00
Machine Operators	0.07	0.00	0.00	0.09	0.00	0.00
Elementary Occupations[2]	0.13	0.07	0.03	0.11	0.06	0.04
Missing occupations data	0.04	0.02	0.01	0.00	0.01	0.00
Panel B: Men						
Population totals	1,036,785	993,888	1,079,777	777,350	1,181,795	1,323,726
Total employment/population	0.39	0.25	0.15	0.38	0.22	0.12
Occupational distribution						
Managers, profs, assoc. profs	0.04	0.01	0.02	0.03	0.02	0.02
Clerks	0.01	0.00	0.02	0.03	0.01	0.01
Services	0.04	0.02	0.01	0.07	0.02	0.00
Agriculture[1]	0.07	0.02	0.00	0.09	0.02	0.00
Crafters	0.11	0.04	0.03	0.10	0.03	0.02
Machine Operators	0.02	0.04	0.02	0.01	0.04	0.02
Elementary Occupations[2]	0.09	0.07	0.03	0.16	0.07	0.04
Missing occupations data	0.01	0.05	0.03	0.00	0.02	0.00

Table shows population totals and means from the October Household Survey (OHS) microdata, the Labor Force Survey (LFS) microdata, the 10% Census microdata and 100% Census community aggregate data. The sample is restricted to rural Africans living in KZN, aged 15-59 inclusive. Means and totals from the OHS/LFS/10% Census data are weighted using population weights provided in each survey. Agriculture [1] includes skilled and subsistence agriculture. Elementary occupations [2] include domestic workers. Census Community data are weighted by the number of people in each community.

Web Appendix 4 Table 2: Correlation between community Census and individual Census employment data

	Female employment		Male employment			
	Census Individual 2001	Census Individual 1996	Census Individual 2001	Census Individual 1996	Δ Female empl. Individual	Δ Male empl. Individual
	(1)	(2)	(3)	(4)	(5)	(6)
Female employment, Community, 2001	0.627***					
	(0.08)					
Female employment, Community, 1996		0.508***				
		(0.08)				
Male employment, Community, 2001			0.284***			
			(0.05)			
Male employment, Community, 1996				0.340***		
				(0.04)		
Δ Female employment, Community					0.384***	
					(0.09)	
Δ Male employment, Community						0.140***
						(0.05)
N	42	42	42	42	42	41
R^2	0.55	0.44	0.39	0.33	0.17	0.01

Table shows coefficients (standard errors) from OLS regressions of employment rates measured in the individual Census data on employment rates measured in the aggregate Census data, where data from each Census has been aggregated up to the magisterial district level. Robust standard errors in parentheses. Significance at *** p<0.01, ** p<0.05, * p<0.1. Δ variables refer to change in the employment rate between 1996 and 2001.

Web Appendix 4 Table 3: Comparing the difference in employment measurement error gaps by gradient (OLS regressions)

	Females		Males		Females	Males
	Community - Individual data, 2001	Community - Individual data, 1996	Community - Individual data, 1996	Community - Individual data, 2001	[2001 underestimate - 1996 underestimate]	[2001 underestimate - 1996 underestimate]
	(1)	(2)	(3)	(4)	(5)	(6)
Gradient aggregated to MD	-0.00311**	-0.00388**	-0.00631***	-0.004	0.001	-0.002
	(0.001)	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)
N	41	41	41	41	41	41
R^2	0.11	0.08	0.08	0.04	0.01	0.01

*** p<0.01, ** p<0.05, * p<0.1, Robust standard errors in parentheses. Each coefficient is from a regression of the outcome variable on land gradient. The first four columns use the (community - individual data) difference in employment rates wthin a year as the outcome variable. The final two columns use the difference in the difference in employement rates across time as the outcome variable. The unit of observation is the magisterial district.

Web Appendix 4 Table 4: Contribution of measurement error in electrification project status to female employment result

Outcome is Δ_t in female employment	Full sample		Restricted to areas with > 10% change in electricity coverage		Restricted to areas with over 80% coverage by 2001	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)
Eskom Project	-0.001	0.095*	0.009	0.095	0.011	0.082
	(0.005)	(0.055)	(0.007)	(0.060)	(0.009)	(0.087)
N	1,816	1,816	1,461	1,461	1,273	1,273

Table shows the Eskom project coefficient (s.e.) from an OLS or IV regression of the change in female employment on all controls as described in Table 3, for different samples: the full sample in columns (1) and (2), the sample restricted to areas with a large change in electric lighting in columns (3) and (4), and the sample restricted to areas with the highest levels of electric lighting use by 2001 in the last two columns. Robust standard errors clustered at sub-district level. Significant at p<0.01***, p<0.05** or p<0.1* level.