Pre-analysis plan
"The Impacts of Electrification on Secondary Education: Evidence from Kenya"

Kenneth Lee* Edward Miguel † Robert Pickmans ‡
Catherine Wolfram §

2021-06-01

1 Summary

This document outlines the statistical analysis to be performed on a dataset that merges secondary school electrification data from the Kenya Power and Lighting Company (KPLC) and test score data from the Kenya National Examinations Council. The goal of this study is to estimate the impact of electrification on secondary school outcomes, complementing the household-level analysis of 2019-2020 follow-up data that builds on the Lee et al. (2020) study.[1] This document describes the regression specifications and outcome variable definitions that we intend to follow. We may also carry out additional analyses beyond those that are specified in this document; this document is not intended to be comprehensive or to preclude us from running additional analyses.

*Energy Policy Institute at the University of Chicago
†University of California, Berkeley
‡University of California, Berkeley
§University of California, Berkeley

[1]We intend to use the AEA author randomization tool to randomize author order when writing the analysis. The pre-analysis plan for the household-level analysis, "The Economic and Social Impacts of Electrification: Evidence from Kenya," as well as information on prior pre-analysis plans from previous survey rounds, can be found at https://www.socialscienceregistry.org/trials/350.

2 Introduction

To date, the research on mass electrification has yielded mixed evidence on the medium to long-term impacts on many socio-economic outcomes for households in low- and middle-income countries. While some research studies have estimated positive effects from electrification (see e.g., Dinkelman, 2011; Grogan and Sadanand, 2012), other work such as Burlig and Preonas (2016) and Lee et al. (2020) have found more muted impacts from electrification. As the latter article notes, complementary inputs to electrification may be necessary to generate meaningful impacts.

In this analysis, we plan to examine the impact of electrification on school-level outcomes in Kenya, with particular attention to complementarities to electrification. Kenya's electrification roll-out in the early 21st century is emblematic of electrification initiatives across Sub-Saharan Africa, and also relates to larger infrastructure initiatives to electrify communities made elsewhere (Lipscomb et al., 2013; Kline and Moretti, 2014; Meeks et al., 2021). With regards to impacts on household-level education outcomes, Khandker et al. (2012) find improvements on completed schooling years and study hours for both boys and girls in Bangladesh, and Khandker et al. (2013) report higher school enrollment rates for girls and boys in Vietnam as a result of electrification. The recent literature on solar lanterns is more mixed in its findings (see e.g., Aevarsdottir et al., 2017; Kudo et al., 2019). Koima (2020) is related to this study, examining the impact of electrification on test scores among 8th grade students in Kenya. We examine the impact of electrification on a different sample with distinct data, and also focus on the role of complementary inputs.

We also examine the possibility of heterogeneous treatment effects of electrification: revisiting earlier work by Burlig and Preonas (2016), Fetter and Uzmani (2020) find the impact from India's Rajiv Gandhi Grameen Vidyutikaran Yojana scheme on nonagricultural employment to be higher in regions specializing in the production of guar, an important input in the hydraulic fracturing process whose price increased sharply in the late 2000s. Meeks et al. (2021) find that micro-hydro plants have less of an impact on the number of manufacturing establishments in areas located farther from the historical electricity grid, and also document differential impacts by worker type. As discussed below, we plan to incorporate data on both local road quality and population density to examine the extent of heterogeneous impacts of school-level electrification along these dimensions. This project's ability to bring together household level data and school level data in the same analysis will allow for a richer investigation into heterogeneous impacts.

2.1 Data construction

For the education outcomes, we use 2001-2016 records from the Kenya National Examinations Council (KNEC), which includes the distribution of Kenya Certificate of Secondary Education (KCSE) scores (English, Kiswahili, Math, Computer Studies) at the secondary school level, disaggregated by gender.[2] We construct a consistent panel of schools and merge that data with connections data from Kenya Power and Lighting Company (KPLC) that describes meter connection dates and the customer names associated with the meter. As a starting point, we further restrict the data to schools that became connected to the national grid between 2005 and 2011, a period of rapid expansion of such connections. This allows us to (i) conduct the analysis on the largest possible dataset for which we have the same schools featured throughout, to (ii) examine differential impacts over the four years within the secondary school level, and to (iii) examine an equivalent number of years prior to connection to examine pre-trends for all schools. We will explore alternative specifications and subsets of the data for robustness. The resulting dataset will provide the "main sample" of schools for the analysis described herein. We will plan to match these data with school-level characteristics from a Kenya Ministry of Education survey to perform additional statistical analyses.

3 Analysis

Below we outline the econometric analysis we intend to carry out. We note that this outline offers a plan for the core analysis to be carried out, and does not preclude us from performing additional analysis.

[2]Computer Studies is one subject among several that meets the "technical" subject requirement for students, and which may be particularly relevant to electrification.

3.1 Main specification

The primary regression specification will be as follows:

$$y_{it} = \sum_{l=-4}^{4} \beta_l \mathbf{I}\{t - E_i = l\} + \lambda_i + \gamma_t + \varepsilon_{it} \tag{1}$$

where y_{it} is the education outcome of interest (described below) for school i in year t, E_i is the year in which school i became connected, λ_i are school fixed effects, γ_t are year fixed effects, and ε_{it} is the idiosyncratic error term. The β_l coefficients represent dynamic treatment effects associated with indicators for being l periods relative to receiving a connection. We are primarily interested in the β_l for which l is positive (i.e., post-connection effects), using the negative values of l to examine pre-trends. In all analyses, we will cluster disturbance terms by school.

3.2 Outcomes of interest

The main outcomes of interest are as follows:

1. Test score: school-level KCSE exam scores, where the exam is administered at the end of the four-year secondary school program, and is required for graduation. For this outcome, we weill use school-level subject scores (English, Kiswahili, Math, and Computer Science) as well as an "overall" score;

2. A binary measure of testing performance: proportion of school KCSE scores that meet the "passing" threshold;

3. Number of students taking the KCSE exam

The first two outcomes will help us assess the main query of interest: the impact of grid electrification on students' educational performance. The third outcome will allow us to examine whether the composition of students changes across schools (i.e., if students are able to be placed in schools that they prefer and if students or their parents have a preference for electrified schools). To investigate the latter hypothesis, we will compare the number of students enrolled at schools (as reflected in the number taking the exam) to the number enrolled at nearby schools over time.

3.3 Heterogeneity specification

3.3.1 Population

We examine local population density as it often correlates with local socioeconomic development (Becker et al., 1999). Oak Ridge National Laboratory collects global population data that represents an ambient population (averaged over 24 hours) distribution.[3] The population data is developed through modeling that uses sub-national level census counts for each country, as well as various geospatial datasets, including land cover, roads, slope, urban areas, village locations, and high resolution imagery analysis. Cells are then weighted for the possible occurrence of population during a day, to produce a dataset available at at approximately 1 km (30" X 30") spatial resolution. Using this data, we estimate a regression specification that builds on (1):

$$y_{it} = \sum_{l=-4}^{4} \delta_l \mathbf{I}\{t - E_i = l\} + \sum_{l=-4}^{4} \rho_l \mathbf{I}\{t - E_i = l\} \times P_{it} + \psi P_{it} + \lambda_i + \gamma_t + \varepsilon_{it} \tag{2}$$

where the notation is the same as from (1), and P represents the population located within the cell that the school is featured in. We are particularly interested in the ρ_l terms, as they allow us to assess heterogeneity in impact by local population density across years where $l > 0$. To align with the grid cells that we have from the data on roads (described below), we aim to match the school with the average population of the 11 x 11 km area in which the school is centered.

3.3.2 Roads

Another source of complementary effects to electrification may come from roads. For example, higher quality roads may allow students easier and more regular access to schools, unlocking higher gains from electrification. Roads may also act as a proxy for broader infras-

[3]See https://landscan.ornl.gov/documentation for more information.

tructure that supports electrification. Jedwab and Storeygard (2021) use road locations data from Nelson and Deichmann (2004) as well as data from 64 digitized Michelin road maps to construct a measure of road quality in Sub-Saharan Africa. An intermediary dataset features the length of road by road quality type (highway, other paved roads, or improved roads) at the 0.1 by 0.1 degree grid square (approximately 11 x 11 km). Using this data on road quality from Jedwab and Storeygard (2021), we estimate the following specification using the road data from 2000:

$$
\begin{aligned}
y_{it} = &\sum_{l=-4}^{4} \delta_l \mathbf{I}\{t - E_i = l\} + \sum_{l=-4}^{4} \rho_l \mathbf{I}\{t - E_i = l\} \times H_i + \sum_{l=-4}^{4} \phi_l \mathbf{I}\{t - E_i = l\} \times L_i \\
&+ \lambda_i + \gamma_t + \varepsilon_{it}
\end{aligned} \tag{3}
$$

where the notation is the same as from (1), H represents the length of paved and highway road in the corresponding grid cell, and L represents the length of unpaved/improved road in the corresponding grid cell. (These terms are not included separately in the regression as they are absorbed in the local fixed effect.) The road lengths will be demeaned by road type. We will test the statistical significance of the ρ_l and ϕ_l coefficients, both separately and jointly, to determine the presence of complementarities to education according to road presence and quality. In future analysis, we may also incorporate additional road length and quality data, and explore particular functional forms for travel times and access, to augment the core analysis laid out here.

3.3.3 School-specific inputs

In addition to the above dimensions of heterogeneity, school specific-inputs may also affect the impact of electrification. We will follow the format of the above specifications using data from the Kenya Ministry of Education, which contains information on the number of teachers (to construct a teacher-to-student ratio), type of school (day vs. boarding and public vs. private), and school sponsor.

3.3.4 Gender

We are interested in examining heterogeneity in the impact of electrification on schooling outcomes by gender. The KNEC data contains data on average examinee outcomes by gender, allowing us to explore this dimension of heterogeneity. For gender, we plan to carry out the following estimation, where each observation is at the school-year-gender level (and g denotes the gender of examinees, taking on values of 0 for male observations and 1 for female observations):

$$y_{itg} = \sum_{l=-4}^{4} \delta_l \mathbf{I}\{t - E_i = l\} + \sum_{l=-4}^{4} \rho_l \mathbf{I}\{t - E_i = l\} \times F_{itg} + \psi F_{itg} + \lambda_i + \gamma_t + \varepsilon_{itg} \tag{4}$$

where the notation is the same as from (1) above, and here F_{itg} is an indicator that takes on a value of one if the observation is the average for females in that school-year (and zero for males). We are particularly interested in ρ_l to assess heterogeneity in electrification impacts by gender across years where $l > 0$ (i.e., post-connection).

3.4 Robustness and other issues

Although the proposed analyses described above rely on the two-way fixed effects regression model, recent econometric literature has found the model to be potentially biased in the presence of differential timing of treatment and heterogeneous treatment effects across units and time. We will implement the methods described in Callaway and Sant'Anna (2020) and Goodman-Bacon (2018) (and any other newer methods that are developed by the time of writing) to assess and address these particular concerns. We leave open other ways in which we might explore the robustness of our results.

References

Aevarsdottir, A. M., N. Barton, and T. Bold (2017). The impacts of rural electrification on labor supply, income and health: experimental evidence with solar lamps in tanzania. Working paper.

Becker, G. S., E. L. Glaeser, and K. M. Murphy (1999). Population and economic growth. *American Economic Review 89*, 145–149.

Burlig, F. and L. Preonas (2016). Out of the darkness and into the light? development effects of rural electrification. Technical Report 268.

Callaway, B. and P. H. Sant'Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*. Forthcoming.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from south africa. *American Economic Review 101*(7), 3078–3108.

Fetter, T. R. and F. Uzmani (2020). Fracking, farmers, and rural electrification in india. Working paper.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. National Bureau of Economic Research Working Paper 25018.

Grogan, L. and A. Sadanand (2012). Rural electrification and employment in poor countries. *World Development 43*, 252–265.

Jedwab, R. and A. Storeygard (2021). The average and heterogenous effects of transportation investments: Evidence from sub-saharan africa 1960-2010. Working paper.

Khandker, S. R., D. F. Barnes, and H. A. Samad (2012). The welfare impacts of rural electrification in bangladesh. *The Energy Journal 33*(1), 187–206.

Khandker, S. R., D. F. Barnes, and H. A. Samad (2013). Welfare impacts of rural electrification: A panel data analysis from vietnam. *Economic Development and Cultural Change 61*(3), 659–692.

Kline, P. and E. Moretti (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *Quarterly Journal of Economics 129*, 275–331.

Koima, J. (2020). School electrification and academic outcomes in kenya. Working paper.

Kudo, Y., A. S. Shonchoy, and K. Takahashi (2019). Can solar lanterns improve youth academic performance? experimental evidence from bangladesh. *World Bank Economic Review 33*, 436–460.

Lee, K., E. Miguel, and C. Wolfram (2020). Experimental evidence on the economics of rural electrification. *Journal of Political Economy 128*(4).

Lipscomb, M., A. M. Mobarak, and T. Barham (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in brazil. *American Economic Journal: Applied Economics 5*(2), 200–231.

Meeks, R. C., H. Thompson, and Z. Wang (2021). Electrification to grow manufacturing? evidence from nepal. Working paper.

Nelson, A. and U. Deichmann (2004). African population database documentation. Technical report.