Move on up

Electrification and internal migration

Angelika J. Budjan *

October 19, 2021

(Preliminary version - please do not cite.)

Abstract

This study uses the large scale roll-out of electric transmission infrastructure in Nigeria from 2009 to 2015 to quantify the effect on electrification on internal migration. I address endogenous location of electricity infrastructure by estimating effects on peripheral households not directly targeted by the policy in combination with instrumenting for the actual grid path by a hypothetical least cost grid. Results show an increase in individual migration propensity by 6 percent and a reduction of household size by 0.8 individuals, mainly driven by young adults and older teenagers. Theoretically, this result can be explained by rising household incomes with a coinciding lack of employment generation for this sub-population. Results from a gravity model of migration show a reduction in movement costs and a rise in migration to rural, electrified destinations following the electricity supply shock.

Keywords: Electrification, migration, agricultural production

JEL classification: H54, J60, L94

*University of Heidelberg and University of Göttingen, email: angelika@budjan.de

Acknowledgements: I am grateful for helpful comments and suggestions from Gerda Asmus, Anca Balietti, Jonah Busch, Axel Dreher, Sabrina Eisenbarth, Meredith Fowlie, Andreas Fuchs, Dana Kassem, Tobias Korn, Sarah Langlotz, Johannes Matzat, Edward Miguel, Jörg Peters, Dina Pomeranz, Dimitri Szerman, Catherine Wolfram and seminar participants at Heidelberg University, the Alliance Summer School in Science and Policy in Paris, the Sustainability and Development Conference in Michigan, and the Indisciplinary PhD Workshop on Sustainable Development at Columbia University, the EAERE Annual Conference at Manchester University, the Beyond Basic Questions Online Seminar, the German Development Economics Conference, the European Public Choice Society Conference, and the European Economic Association Conference.

1 Introduction

This paper analyses the effect of a local electricity supply shock on internal migration. Investments in rural infrastructure are an important instrument to foster development without relying on urban centers as sole engines of growth. Yet, little is know about the effect of efficiency gains from infrastructure investments on population dynamics. While local growth effects might reduce out-migration incentives (for instance, as documented for United States Lewis and Severnini, 2020), a rise in incomes in a developing country context could also enable out-migration by overcoming credit constrains (Mckenzie and Rapoport, 2007, Bryan et al., 2014, Angelucci, 2015, Bazzi, 2017, Clemens, 2020). This is relevant given that rural infrastructure investments are seen as an alternative to rapid urbanization which, in the case of Sub-Sahara Africa, is often an unplanned and uncoordinated process resulting in congestion, low connectivity and environmental pressures (World Bank, 2016).

The context of this paper is Nigeria in the years 2009 to 2016, where conditions are favorable to expect large productivity gains from electrification. Access to modern electricity has high priority on the global agenda with nearly 1 billion people lacking it (IEA, 2019), but the academic literature finds mixed results regarding its development effects (see Bayer et al., 2020, Lee et al., 2020, for surveys of the literature). Large scale investments in transmission infrastructure and generating capacity along household connections are thought to produce the largest effects (Lee et al., 2020). In addition, complementary factors such as pre-existing industries and market access are assumed to be crucial for its effectiveness. For instance, (Fetter and Faraz, 2020) find a positive electrification effect only in regions that experience a simultaneous shock in demand for local commodities. In the case of Nigeria, the investments in electricity infrastructure were both large scale and in response to wide gap between existing supply and demand.

Understanding the effects of electrification on population dynamics is particularly important in lights of the large gap in productivity and standard of living between urban and rural areas across the developing world (Gollin et al., 2014, Young, 2013). Many scholars see this gap as evidence that moving workers out of the agricultural (rural) sector into the more productive (urban) sector can create large productivity gains (Gollin et al., 2014, Bryan and Morten, 2018). In addition, a high degree of unequal distribution of economic activity across space is associated with low levels of development (Alesina et al., 2016, Lessmann, 2014). One solution is to reduce barriers to migration, as has been the focus of a growing body of research (Allen and Arkolakis, 2014, Bryan and Morten, 2018, Baum-Snow et al., 2020, Bryan et al., 2014, Angelucci, 2015, Lagakos et al., 2018, Bryan et al., 2021, Bah et al., 2020). However, migration might not be desirable for everyone, [1]

[1] In an early research article, Sjaastad (1962) pointed out that migration comes with non-monetary costs, including the disutility from leaving *"familiar surroundings, family, and friends"*. In a similar vein, Blanchard and Kirchberger (2020) muse that *"movement from rural to urban areas may involve*

and can lead to unintended outcomes both at the sending communities (e.g. Baum-Snow et al., 2020) and the receiving urban centers (Henderson, 2002). Thus, investing instead in rural infrastructure as means of fostering country-wide development and of closing the rural-urban gap is a common strategy across the developing world. Whether these investments also slow down internal migration has political significance.

To analyze this question, I rely on data from Nigeria's General Household Survey which offers a rich geo-coded household panel that tracks households and individuals over time. For identification, I use a first-difference estimation conditional on state-wave fixed effects and a number of geographic controls. Endogenous allocation of the transmission infrastructure is addressed in two ways. First, I exploit the fact that transmission lines are large scale connections between two local substations that transport high voltages across long distances.[2] At the local substation, electricity is fed into the local distribution grid, which makes them both an important determinant of the grid locations, but also a highly endogenous variable.[3] However, households located between two of these substations were not the ultimate target of the intervention. Yet, these households benefited greatly from the grid expansion, since distribution lines often follow the path of transmission lines to save costs. This approach builds on Faber (e.g. 2014), who estimates the effect of road construction in China on peripheral cities.

Second, I construct a hypothetical least cost path as instrumental variable for the actual grid path. Given that the path of each transmission line is mainly dictated by the location of the respective substations,it is still possible that policy-makers use the little wiggle room they have to favors certain location – be it for winning voters or for favoring the villages with the highest economic potential. The least cost approach overcomes this concern by isolating supply side factors of infrastructure provision based on the costs of its construction given the characteristics of the terrain. This approach draws heavily on (Faber, 2014), while variations of the least cost approaches find increasing applications in economics (e.g. Banerjee et al., 2020, Kassem, 2020).

Results from first-difference and instrumental variable regression show that the electricity supply shock reduced household size by between 0.3 and 0.8 household members. This decrease is particularly driven by older teenagers aged 13 to 18, while household heads show no increase in migration propensity. At individual level, migration propensity increased by 6 percent. Moreover, I find a significant increase in work related

loss of social connection or information insurance, or the loss of claims to land and other resources in rural areas. There may be barriers for rural people - particularly those who are older - in learning new kinds of work or new way of life." These psychological costs are difficult to quantify and if sufficiently large could explain lower levels of observable migration than expected by theoretical models – without implying resource misallocation.

[2]Transmission lines constructed during the sample period measure on average around 100km

[3]Notable attempts to exploit exogenous variation in substation location exist (Lipscomb et al., 2013), but they are sensitive to model assumptions and more credible for historical grid construction, then for the expansion on an existing grid.

migration by 30 percent for male adults, and of 12 percent for minors. These results seem linked to a combination of increased access to credit with limited job creation for the youth. While, household income proxied in logarithmic food consumption, increased by 23 percent, total working hours and employment outside subsistence agriculture increased only for the household heads, but not of other subgroups.

I complement this analysis with results from a gravity model of migration. These show that also at municipality level migration flows increase after grid construction. What is more, the effect of movements costs on migration goes down to a approximately a third, in line with the existence of barriers to migration in the form of credit constrains. In addition, I find that migrants from municipalities that received new grid are more likely to migrate to rural destinations that also just received a new electricity grid.

This paper is the first rigorous empirical analysis of the impact of electrification on internal migration in a developing country context. Previous studies have either focused on rich countries or applied less empirical rigor. Lewis and Severnini (2020) analyze the effect of the historical expansion of the electricity grid on internal migration in the United States. They find a significant positive effect on population linked to productivity gains in the agricultural sector. Fried and Lagakos (2021) construct a multi-sector spatial model that predicts a reduction of out-migration in electrified villages due to productivity gains. They offer empirical results from difference-in-difference estimation on Ethiopian villages in line with this prediction. However, a simple difference-in-difference estimation is likely to suffer from selection bias as outlined above.

In addition, my paper differs from previous studies by considering credit constrains in the theoretical predictions. While previous theoretical models focus on productivity effects (Lewis and Severnini, 2020, Fried and Lagakos, 2021), the existence of credit constrains might imply sub-optimal migration levels *ex ante* which are adjusted when incomes rise. This can ultimately lead to a net increase in out-migration. This theoretical prediction draws on literature about the income-migration relationship which has mainly focused on the effect of cash transfers (Bryan et al., 2014, Angelucci, 2015, Molina Millán et al., 2020). These studies typically find a positive effect of alleviating credit constrains, particularly for poor households (see Adhikari and Gentilini, 2018, for a survey of this literature). While these studies are useful to understand the isolated role of credit constrains, they do not tell us much about increasing the opportunity costs of migration by raising incomes at home. However, given the current policy debates on ways to slow down rapid urbanization the question of opportunity costs is highly salient. Bazzi (2017) explores the effect of income shocks from variations in rainfall pattern in Indonesia and finds a positive effect on labor migration. While this study is closely linked to this paper, short-lived income increases from rainfall shocks do not change incomes at home for more than one period and will therefore affect the opportunity cost of migration to a limited degree.

Finally, my study contributes to the wide literature on the effects of electrification. Most studies have focused on the effect of electrification on income, employment, health or education (Dinkelman, 2011, Grogan and Sadanand, 2013, van de Walle et al., 2013, Burlig and Preonas, 2016, Lenz et al., 2017, Lee et al., 2020). My results suggest that employment benefits from electrification do not occur homogenously across sub-populations, particularly in an environment of high underemployment. This might explain why some studies tend to find small to no employment effects (Burlig and Preonas, 2016, Lenz et al., 2017, Lee et al., 2020) while others find large effects (Dinkelman, 2011). In addition, investments in the migration of younger household members might not always be accounted for correctly in the assessment of household welfare and the lack thereof might obscure positive effects from electrification.

The remainder of this paper is organized as follows. Section 2 described the context of the study; Section 3 discusses the data sources of the study; Section 4 describes the empirical strategy both at the household-level and at the grid-cell level; Section 5 presents the main results; Section 6 reports robustness tests; Section 7 reports results from the gravity model and finally Section 8 concludes.

2 The context

Nigeria's labor market is characterized by a lack of adequate earning opportunities. In 2011, the World Bank estimated that 53 million Nigerians between the ages of 15 and 64 were working, but half of them in low-productivity agriculture (World Bank, 2016). Despite a moderate level of unemployment, household earnings are often not sufficient to meet basic needs such that a third of the population continues to live below the poverty line. The low earnings are caused by a general lack in labor demand in the formal wage sector. Most work is informal and either self-employed or for a family-owned business. High population growth and rising inequality across regions add additional stress on the labor market.

However, during the study period of this paper sectoral transformation was already on its way. Spurred by macroeconomic growth, in the year 2007 to 2011 the share employment in the agricultural fell from 58 to 50 percent, with new jobs emerging in the private and public wage sector(World Bank, 2016). Wage employment in agriculture is low with only 1 in 20 workers being a wage laborer in 2011. In addition, the World Bank report finds that youth faces barriers to entering the labor market after completing education potentially due to mismatch between skills acquired at school and skills need at potential jobs.

This lack of adequate work, particularly for the youth, is one of the driving forces of internal migration. Using a migration census, (Mberu, 2005) show that on average 58.3 percent of Nigeria's rural-born population are migrants, meaning they reside in a different location than they were born. Of these, 37 percent are rural-urban migrants and 63 percent are rural-rural migrants, illustrating that rural-rural migration constitutes the main share of permanent migration. Amrevurayire and Ojeh (2016) find that in the Ughelli South Local Government Area of Nigeria migration is highest for the age cohorts 15-25 and 26-35 years and decreases in age. Moreover, the authors identify unemployment, a search for education and a lack of basic infrastructure as main reasons for migration. In addition, Dillon et al. (2011) find that agricultural households use the migration of male household members to respond to negative income shocks. These findings suggests that an improvement of earning possibilities and income diversification in remote areas should slow down migration.

While migration might be an optimal strategy for the individual household, outcomes for the sending communities are not always positive. A study in the Niger Delta region shows that rural out-migration leads to sizable labor shortages in the agricultural sector which results in incomplete harvest and foregone revenue (Ofuoku et al., 2017). This mirrors the findings of (Baum-Snow et al., 2020) in China that an increase in migration can have detrimental effects of the economies of origin locations. Thus, migration is not only a result but also a driver of the increasing rural-urban gap.

Infrastructure development is an important component of Nigeria's rural development efforts. In particular the electricity sector holds a crucial position given that increases in power generation capacity have been slow over the last three decades and have not kept track with economic and population growth (Gatugel et al., 2015). Nigeria's electricity consumption was in 2015 one of the lowest in the world with only 156 kWh per capita (World Bank, 2017). Particularly rural areas are under-supplied. The low level of electricity supply hampers productivity across sectors. What is more, it is estimated that the connected population more than half the time faces power problems (Sadiq et al., 2015). Many businesses rely on private electricity generators for production when grid electricity is unavailable or unreliable, raising their costs of production (Pestana et al., 2014).

To address these issues, the Electric Power Sector Reform Act from 2005 demanded the privatization of the entire power sector to create incentives for investments in generation and transmission infrastructure. Among other changes the state-owned Power Holding Company had been unbudled into multiple entities. Since then electric transmission has been managed by the Transmission Company of Nigeria (TCN) (NERC, 2019), which immediately started to undertake efforts to improve grid supply.

Regional efforts to strengthen the coordination of the energy sector in the ECOWAS region already lead in 2007 to the construction of a new transmission line in the South-West at Ikeja West substation and Sakete in Benin. In addition, in 2009 the World Bank committed a credit worth approximately 200 million US dollars for the power sector for the funding period 2009–2014. The proposed project consists in the extension of the generation capacity, the expansion and rehabilitation of the transmission infrastructure and best-practice investments in distribution infrastructure (World Bank, 2009). Out of this 180 million US dollars are solely dedicated to the enhancement of the transmission and distribution grid. As a consequence, a number of major transmission lines were constructed between 2009 and 2015 in context of the World Bank funded Nigeria Electricity and Gas Improvement Project (NEGIP). These investments went along with major investments in generating capacity.

Rural electrification holds a high priority for Nigeria, reflected in the creation of the Rural Electrification Agency in 2005 which lists "driving economic development" as one of its policy objectives according to its website (Rural Electrification Agency, 2021). The website elaborates that this goal consists in "empowering[ing] local industries to play a larger role in the supply chain from materials, manufacturing, construction and operation of the assets" – illustrating that the spatial redistribution of economic activity is an intended consequence of Nigeria's rural electrification efforts. While slowing down migration is not a declared objective, population dynamics are not likely to remain unaffected.

3 Data

To order to analyze how the expansion of the electricity grid affects productivity and migration, I rely on Nigeria's General Household Survey which was collected by the Nigerian National Bureau of Statistics in partnership with the World Bank Living Standards Measurement Study. While the General Household Survey was initiated in 2006, since 2010 it has been collected in a panel structure, following the same approximately 5,000 households over time, and forms a representative sample of the Nigerian population. This study uses 3 waves from the years 2009/2010, 2012/2013, and 2015/2016.[4] It provides detailed information on household consumption, income generating activities, agricultural plot owned by the household and information on each individual household member together with geographic coordinates.

Data on grid expansion and substation location comes from the Energy Database published by the Rural Electrification Agency of Nigeria (Rural Electrification Agency, 2020). This database offers data on various indicators related to energy supply including the exact location and electric tension of substations and main transmission lines as well as the year of construction of the latter.[5] A number of long distance transmission lines

[4] A 4th wave was collected in 2018/2019, but the high degree of attrition from the original panel makes the data useless for the purpose of this study.

[5] For quality assurance, internet research was carried out to verify the construction year of each transmission line. Based on this the following adjustments were made: The transmission line between Dutse Substation and Azare Substation in Jigwara was originally coded as existing in 2000. An alternative source from the World Bank did not report this transmission line. Additional sources reported

were constructed between 2009 and 2015 (Figure 1). They typically measure more than 100 km in length and have a voltage of 132 kV or 329 kV.

My main definition of an electricity supply shock assumes all household affected that were within a 15 km distance of a new transmission line. According to this definition 139 household experienced an electricity supply shock during the observation period (69 between wave 1 and 2 and 70 between wave 2 and 3). These were located across 10 of the 37 states of Nigeria.[6]. These households can be interpreted as the treatment group. Households located in the same states that did not experience an electricity supply shock constitute the control group against which the treatment effect is estimated.

A balancing test on wave 1 observations shows that treatment and control households do not differ significantly across most baseline characteristics (Table 1). On average only 38 percent of the control group households were electrified in 2010. They were on average located 17.59 km away from the closest transmission grid line and 40.30 km from the closest substation. Treatment households were slightly less likely to be already electrified, they were located slightly closer to any existing grid line and slightly more distant from any substation - but none of these differences reaches statistical significance. Importantly, also other geographic characteristics are balanced between both groups. Neither the distance to any major road or the state capitals differ significantly, nor population density or the percentage of cropland and urban land. Given that the identification strategy relies heavily on geography, balancing of geographic characteristics makes it unlikely that difference in time trends across geographic locations are biasing the results.

However, the test shows a few statistically significant differences in household characteristics, particularly in building materials of the accommodation. Treatment households were more likely to have an iron roof and concrete walls and less likely to have a grass roof and unburnt or burnt brick walls. The differences are small and seem uncorrelated with wealth, since agricultural wages and production values do not differ significantly between groups. In addition, the test shows a 90 percent significance difference in the use a diesel generator for lighting and the number of elderly household members. Overall, the significant differences between treatment and control households appear small enough to be driven by chance. A test of joint significance yields are very large p-value of 0.998. Nevertheless, control for potential bias I test all main regressions

---

the construction year of Azare Substation to be 2010. Therefore, the construction year of this line was coded as 2010. The same World Bank map reported the between line Dutse and Kumbotso as existing in 2008, while the Energy Database reported the year of construction as 2010. In combination the wrong year from the neighboring line between Dutse and Azare, it seems that dates of these two lines were accidentally swapped when coded. Therefore, the construction year of the line between Dutse and Kumbotso was re-coded as 2008. The transmission line between Ihovbor and Okada was coded as existing in 2000 and changed to 2018, because the substation construction was found to be only finalized in 2018. The extension of the Odugunya substation was coded as 2010 and changed to 2018, because the additional substation was only created in this year.

[6]These 10 states are Abia, Akwa Ibom, Bauchi, Benue, Ebonyi, Enugu, Imo, Jigara, Kano, and Nasarawa

against the inclusion of all significant difference of the balancing test (appendix Tables B-1 - B-3). In order to avoid problems of multicollinearity these covariates are omitted from the main specifications.

4 Empirical Strategy

4.0.1 Difference-in-difference estimation

I begin the analysis with a difference- in-difference estimation at household level. Changes in the main outcome variables are explained by changes in the proximity to a new transmission line conditional on the distance to the closest substation, other geographic control variables and state-wave fixed effects. Algebraically, it takes the following form:

$$\Delta Y_{ijt} = \alpha \Delta D_{ijt} + \beta' X_{ij} + \gamma_{jt} + \epsilon_{ijt} \tag{1}$$

where Y_{ijt} is a vector of outcome variables at household i in enumeration area j at time t. My independent variable D_{ijt} is a dummy variable that takes a value of 1 if the household was located within a 15km distance of any newly constructed transmission line.[7] Alternatively, I run regressions using a continuous measure of the negative logarithmic distance to the closest new transmission line. The negative sign ensures ease in interpretation of the results as done in similar studies (Lewis and Severnini, 2020). X_{ij} are household specific time-constant geographic control variables which are outlined below. Most importantly, these include the distance to the closest electric substation. γ_{jt} are state-wave fixed effects. The error term ϵ_{ijt} is clustered at enumeration area (which is in most cases equal to the village) to correct for correlated errors due to the sampling structure of the data. Households within a 10 km distance from any substation were excluded from the sample to control for the fact that these might have been directly targeted by the policy. In addition, 7 households were excluded from the dataset which migrated as a whole during the observation period, in order to satisfy the exclusion restriction. [8]

All regressions control for the distance to the closest electric substation. This accounts for fact that substations were directly targeted by the policy, as nodes where electricity

[7]The 15km buffer was selected based on first stage regressions that tested the correlation between distance to the transmission grid and household electrification.

[8]Household migration is very rare in this dataset. Overall in the survey there were 45 household that moved during sample period, but only 7 households located in the treatment states. Due to the fixed effects structure of the main estimation strategy, this number is to low to analyze household migration quantitatively. When analyzing individual level migration, cases where the whole household migrated where excluded, because the identification strategy relies on geographic factors remaining constant. Given the exclusion of household migration, estimates from individual level migration therefore constitute a lower bound for total migration.

was fed into the low voltage distribution grid. The locations are strategically chosen in areas of high electricity demand and therefore highly endogenous. In addition, I present results before and after controlling for a number of additional geographic variables. In particular, these include distance to the respective state capital,[9] distance to the closest major road in 2009, population density within a 40 km buffer, percentage of cropland and percentage of urban land within a 40 km buffer. Details on data sources and metrics of the control variables can be found in appendix table A-1.

These are included to address concerns of non-parallel trends based. Since geographic variables trend to be correlated with each other, there is a risk of non-parallel trends based on geographic location when exploiting geographic variation in the main explanatory variable. For instance, Bensch et al. (2020) find that the instrumental variable for electrification in the influential seminal work of (Dinkelman, 2011) also predicts road access which could drive the results. The geographic controls of this paper reflect that locations might trend differently depending on their market access, political importance, urbanization rate and sectoral composition. However, the risk from non-parallel trends across geographies seems limited since the balance test (table 1) yields only small, statistically insignificant differences between treatment and control households.

At individual level the regression takes a very similar form of:

$$\Delta Y_{cijt} = \alpha \Delta D_{ijt} + \beta' X_{ij} + \beta gender_{cij} + \gamma_{jt} + \epsilon_{cijt} \tag{2}$$

where Y_{cijt} are outcomes at individual level, D_{ijt} is the respective measure of proximity to a newly constructed grid line, X_{ij} are geographic controls remain at household level. At individual level, I only control for $gender_{cij}$ since employment and migration behavior is expected to differ greatly between genders. In addition, I run regression separately based on gender and age group or relationship to the household head. The relationship to the household head is relevant for the main outcomes. Every household member inhibits a different role based on social norms and is expected to contribute to a different degree to the household income.

4.0.2 Instrumental variable estimation

Estimating equations (1) and (2) by OLS risks bias, if the path of the transmission lines was not assigned at random but followed economic and political considerations. To address this concern, I implement the least cost path approach introduced in (Faber, 2014). This approach isolates supply-side factors of infrastructure provision. In particular, I determine for every new transmission line which path it should have followed

[9] In some states the state capital is not the most populated city. Due to the multicollinearity of both variables I do not include both the distance to the state capital and the distance to the largest city in the same regressions. However, results do not depend on which of both measures is used.

in order to connect the terminal substations most cost-effectively. The construction cost are based on characteristics of the terrain that needs to be crossed. Following Faber (2014), I employ gridded land-cover data together with elevation data to measure land gradient. High construction costs are assigned to pixels with a high slope and to pixels that classify as urban areas, waterbodies or wetland.[10] Next, the algorithm selects a path to connect the terminal substations that results in the lowest construction costs. A detailed description of this method can be found in Appendix A.

Figure 2 shows a visual illustration of this approach for an example in the states Jigawa and Bauchi. The new grid line that was actually constructed is approximately concave, while the hypothetical least cost grid is in this case simply a straight line. the difference between both paths suggests that actual grid construction was biased on favor households in the North.

The least cost grid is then used to instrument for the treatment variable D_{ijt} of equations (1) and (2). The two-stage least squares version of equation (1) takes the following form:

$$\Delta Y_{ijt} = \alpha \Delta \hat{D_{ijt}} + \beta' X_{ij} + \gamma_{jt} + \epsilon_{ijt} \tag{3}$$

First stage equations:

$$\Delta D_{ijt} = \alpha \Delta L_{ijt} + \beta' X_{ij} + \gamma_{jt} + \epsilon_{ijt} \tag{4}$$

where $\hat{D_{ijt}}$ are the fitted values of proximity to the actual grid and L_{ijt} indicates proximity to the hypothetical least cost grid.

First stage results are reported in table **??** panel B. Columns (3) and (4) report results for a continuous measure of proximity to any new transmission line, while columns (7) and (8) report results for a dummy variables that turn 1 if the new grid (and the hypothetical least cost grid respectively) was within 15 km proximity. All specification yield very similar estimates of 0.855 to 0.866 points that are highly statistically significant. Kleibergen-Paap F-statistics indicate that the continuous measure of grid proximity leads to higher statistical power, but also the binary measures result in large F-statistics of 55.66 (66.89 respectively). Besides being a strong instrument, the exclusion restriction appears to be satisfied. Proximity to the least cost grid only affects outcomes via proximity to the actual new grid. It would be violated if proximity to the least cost grid correlated with other factors such as sectoral composition that lead to different trends in treatment locations. This seems less likely given the main regressions already control for a number of geographic covariates. In addition, I conduct a series of robustness test to test the

[10] This simple algorithm adopted from Faber (2014) finds its original motivation in the transport engineering literature (Jha and Schonfeld, 2001, Jong and Schonfeld, 2003)

validity of the exclusion restriction.

First, I include baseline covariates that showed significant differences between treatment and control households in the balance test of table 1. Second, test the main results against a specification with household fixed effects that comes out with even weaker assumptions than the first-difference regression. Third, I run a placebo test on future grid lines. If grid locations trended differently from non-grid locations, proximity to future grid lines (instrumented by their hypothetical least cost paths) should lead to similar effects as actual grid. Finally, I test against variation in proximity to road infrastructure.

5 Main Results

5.1 Electricity

Since grid expansion is only a crude measure of electricity access, I first test whether the new transmission lines resulted in increased local electrification. Low reliability of the electricity network and high costs often undermine demand Lee et al. (2019). Grid expansion affects electrification both at local farms, businesses, and private households. However, the data allows only to test for household electrification. While not the only channel, this offers suggestive evidence of changes in local electricity use.

Table 2 reports results of grid expansion on household electrification status for the continuous and the binary treatment measure using both the observed grid and the hypothetical least cost grid. All specifications show a positive correlation between the treatment variable and household electrification. This effect is larger for regression that rely on the hypothetical least cost grid, implying that actual grid location might have favored economically prosperous regions that were already better supplied with electricity. The binary treatment measure shows much larger effects than the continuous measured implying that electrification benefits from new transmission lines fall to 0 after a certain distance threshold. The binary indicators form therefore my preferred treatment measure.

Households located within a 15 km buffer of a new transmission line increase household electrification by 18-54 percent. In all specifications, control variables show no significant effects on changes in household electrification. This suggests that time trends in household electrification are mainly driven by grid construction, i.e. supply side effects. Demand side factors such as urbanization seem to only matter only as long as they lead to new grid construction. In addition, I analyze the effect of the new transmission lines on household fuel choices (appendix Table A-2). Results from the preferred specification show that grid expansion led to a 26 percent increase of electricity, a 31 percent reduction of kerosene use and a 17 percent of battery use as main lighting fuel. The finding underlines that grid expansion created an economically relevant shift in local electricity supply.

5.2 Migration

Table 3 reports the effect of the electricity supply shock on household composition. The table shows OLS and 2SLS results with and without geographic control variables. Across specifications the electricity supply shock reduced the number of household members by between 0.33 to 0.78 persons. This effect is large given that the average household consisted in approximately 6 persons. The effect seems to be partly driven by children. The number of older teenagers aged 13 to 18 went down by 0.14 to 0.335 individuals. Since teenagers in Nigeria often enter the workforce at age 15, this could be both education or work related migration. In either case, given the high unemployment rates among Nigerian youth migration of this age group is probably linked to a household investment in the young member of the household that was previously impossible due to credit constrains. At the same time, the result implies that the productivity shock did not raise employment potential for young people so much that staying and pursuing wage employment would on average be preferred over migration. In addition, the number of young children below the age of 5 decreases by 0.17 to 0.26 individuals.

Turning to results at the individual level, grid construction increased individual migration propensity by 5.6 percent in the preferred specification (Table 4). The effect is smaller and statistically insignificant when using the actual grid path suggesting some bias in the way the actual grid path was selected. At individual level, I distinguish between role of the household member within the household, such as household head, spouse etc. Assuming decision making at the household level, this creates more homogeneous subgroups than grouping by age and/or gender. When analyzing these groups separately, an interesting pattern emerges. Across all specifications migration of the household head is not affected by the productivity shock. This was expected given that migration of whole households was rare and household head migration would typically imply migration of the whole household. The subgroup that mainly showed an increase in migration propensity is the group of children of the household head. Their likelihood to migrate increased by between 10 percent in the preferred specification. This finding is in line with results from Table 3. However, in this table children of the household head are not defined by age, so this group includes also young adults. The oldest 25 percent of this group are aged 17 to 37. This provides additional evidence for households investing newly gained resources in the migration of young household members. For spouses of the household head migration propensity decreased by 5.8 percent in the preferred specification. An interpretation of this result will be discussed below.

To get a clear picture of the migration surge, it is crucial to understand the motives behind out-migration. Given that the rise in out-migration is mainly driven by older teenagers and young adults, work is not the only possible motive. In addition, migration could be linked to a pursuit of education. In both cases, however, the expected returns to migration, must have exceeded expected returns from staying. Appendix Table A-3 reports results on an analysis of migration reasons among the sample of migrants. Since this analysis is performed on the sample of migrants, it does not achieve very high statistical power, but seems nevertheless informative. For this analysis, the sample is grouped by gender and age to keep the sub-samples as large as possible. For both adult male migrants and under-aged migrants there is an increase in work related migration after the productivity shock. This confirms the theoretical expectation that earning potential at destination is one of the main pull factors of migration. For adult men the migration motive "for work" increases by 30 percent relative to other reasons and is significant at the 5 percent level. For under-aged household members this motive increases by 12 percent, again significant at 5 percent. This suggest that work related migration is at least partly responsible for the increase in migration among older teenagers that is visible in Table 3.

In addition, migration of children seemed to be driven by the reason "to join family". This category increased for under-aged migrants by 33 percent (significant at 5 percent) and is consequently much larger than the increase in the migration motive "for work". Without additional details the answer "to join family" is difficult to interpret. Possibly the rise in earnings potential of the adults of the household increased the opportunity cost of child care to such a degree that relatives were charged with this task. This explanation would be in line with the decrease in young children below 5 observed in Table **??** which is most likely not work or education related.

Finally, the results from appendix Table A-3 offer some insight in the reduction of spousal migration. Among female adults the migration reason "divorce/separation" went down by 13 percent and the effect is significant at the 5 percent level. This is in line with the general notion that divorces rise with economic pressures.

5.3 The employment channel

Next, I analyze impact on employment and productivity as channels of the migration effect. Table 6 reports individual level employment effects. While on average across all household members there is no significant effect on employment, there is a significant effect on the employment of the household head. The study distinguishes here between non-farm and farm work to reflect the fact that not unemployment is the major challenge for Nigeria's labor market, but underemployment and deadlocked employment in subsistence agriculture. The variable farm work comprises all cases of work on a family owned farm. Non-farm work comprises all types of wage work or self employment, including wage employment in agriculture. For the household head, non-farm work increased by between 7.5 and 12.1 percent, while farm work remained unaffected by the productivity shock. In addition, working hours of the household head increased by between 4.9 and 10 hours. For their spouses the likelihood of employment and the total working hours seem largely unchanged. This inelastic response of spousal employment is probably linked to traditional expectations about gender roles. Finally, the likelihood that children of the household head are working in non-farm employment decreased by 2.6 to 4.6 percent. This findings confirms the expectation that employment opportunities did not emerge in par for all subgroups. Older teenagers and young adults did not seem to benefit from the increase in labor demand experienced by household heads. While the productivity shock increased access to credit for the household, it did not increase the opportunity costs of migration of this subgroup to a relevant degree. In addition, the negative effect suggests that previously undesirable employment of children was now stopped.

To understand where new employment was generate, I analyze sector of employment in appendix Table A-4. It provides suggestive evidence of sectoral transformation. At baseline 25 percent of the sample population worked in agriculture as their primary sector of employment, 5.9 percent worked in retail and manufacturing and personal services employed 2 percent respectively. When using the full sample (column (2)) the results for most sectors are close to zero. Employment in agriculture diminished by 10 percent, but fails to reach statistical significance. Average employment in retail increased by 3.6 percent and employment in transport by 1 percent, though these estimates reach only 90 percent significance.

Analyzing the sub-groups of household members reveals some nuance. In particular, there were positive employment effects for the personal services sector and the retail sector. The electricity supply shock increased employment of the household head in retail by 11 percent (at 10 percent significance) and by 20 percent for their spouses (at 1 percent significance). In addition, employment of the household head in personal services rose by 5.6 percent (at 99 percent significance) and employment in transport rose by 3 percent (at 10 percent significance). Agricultural employment of the household head fell on average by 9 percent, but does not reach statistical significance. For their spouses, employment in agriculture fell by a similar magnitude (11 percent), again without reaching statistical significance. Moreover, spousal employment in retail increased by a highly significant 21 percent. Most other sector seem unaffected for spouses. Since the fall in agricultural employment of spouses is smaller than the rise of their employment in retail, it appears that spouses partly moved out of from unemployment or under-employment into employment in the retail sector. For children of the household head, we can also observe a statistically insignificant reduction in agricultural employment of 7.6 percent, while the other sector seem unaffected. In addition, there is an 9 percent reduction of grandchildren's employment in the personal services sector and a negative coefficient on agricultural employment.

To understand the employment effect, I next analyze the effect on productivity. Table

5 presents results on productivity of the agricultural sector for which data was readily available in the GHS panel. For agricultural production, there is a significant increase in inputs, in the form of costs of agricultural laborers and plots per households. Labor costs rose by approximately 85 percent in the preferred specification, while the number of wage laborers remained constant, suggesting an increase in agricultural wages. The number of agricultural plots per household increased by 0.76 units. This implies an efficiency gain in agricultural production, since the ratio of workers per lot decreased. What is more, it suggests that the rise in wages was no pure price effect. Surprisingly, the value of agricultural production did not increase to a statistically significant degree. This seems partly driven by poor data quality because the measure shows very large standard errors. In addition, it could mean that rising labor demand in other sectors created a labor shortage in agriculture, leaving harvest incomplete as observed by Ofuoku et al. (2017) as a consequence of migration in the Niger-Delta Region of Nigeria. Finally, I proxy household income by logarithmic food consumption per capita which increased by between 8 and 27 percent. It is therefore evident that the productivity shock had a positive impact on household's earnings and in turn extended their credit line. Overall, these results suggest that productivity gains in the agricultural sector might have freed time - particularly of the household head - to follow other income generating activities.

6 Robustness

6.1 Additional baseline controls

In order to address the concern of non-parallel trends between treatment and control households, I test the main results of Tables 2, 3, and 4 against the inclusion of additional baseline controls. The balance test discussed in section refS:4 shows only marginal differences in most baseline controls between both groups, therefore a violation of the parallel trends assumption is not likely. Statistically significant differences appear for building materials of roof and walls, main lighting fuel and number of elderly household members. In appendix Tables B-1 – B-3, I present replications of the main results while controlling for these baseline covariates. Results of the effect on household composition, individual migration and agricultural production do not change substantially after the inclusion of additional baseline control variables.

6.2 Individual-level fixed effects

Next, I test the main results against an alternative specification using unit fixed effects instead of first-differences. Controlling for the impact of time-constant geographic covariates is algebraically more simple in the first-difference approach. In addition, the first-difference estimator is less sensitive to the strict exogeneity assumption in short panels (Wooldridge, 2010). Thus, comparing the main results to the fixed effects result provide some indication about the presence of bias. The fixed effects equivalent of equation (1) reads:

$$Y_{ijt} = \alpha D_{ijt} + \beta' X_{ij} \times wave_t + \mu_i + \gamma_{jt} + \epsilon_{ijt} \tag{5}$$

where μ_i indicates household fixed effects and the time-constant geographic control variables X_{ij} are interacted with the respective wave $wave_t$ to produce a similar specification to the first-difference estimation.

Results of this exercise are presented in appendix tables B-4, B-5, and B-6. Neither the point estimates nor the standard errors differ greatly between the fixed effects and the first-difference specification. F-statistics of the instrumental variable approach are however smaller by approximately factor 0.5. For household composition, results are qualitatively the same, but the effect for older children aged 13 - 18 loses statistical significance, caused by a slightly larger standard error and a slightly smaller coefficient. At individual level, the results confirm a positive effects on migration propensity in average and on the children of the household head, as well as negative effect on migration propensity of their spouses. For agricultural production the results confirm a positive effect on household food consumption. For the other outcomes, however, results differ to a relevant degree. While the effect on labor costs is positive, it is smaller than the first-difference estimate (0.262 compared to 0.856) and not statistically significant. The same applies to the number of plots (0.083 compared to 0.766). Given that the fixed effects approach is more sensitive, the first-difference results for these two indicators should come closer to the true effect while apparently not being unbiased.

6.3 Future grid lines

Finally, I the conditional exogeneity of grid locations to the main outcomes, by regressing them on future grid lines. If grid lines are conditionally exogenous to the main outcomes, future grid lines should not have an effect.

Data on future grid lines comes from the same data set as actual grid lines (Rural Electrification Agency, 2020). Future grid lines were planed for the years 2018, 2020 and 2025. I code grid lines planned for 2018 and 2020 as occurring between first and the second, and grid lines planned for 2025 as occurring between the second and the third wave of the household sample. Then I replicate Tables 3 – 5 using future grid lines instead of actual grid lines. Results are presented in appendix Tables B-7, B-8, and B-9. Future grid lines have no statistically significant effect on household composition (appendix Table B-7). The coefficient on household size is negative – as the effect of actual grid lines – but very small at a reduction of 0.18 persons (compared to 0.78 persons in the main results). For the group of older teenagers between the ages of 13 to 18 years the point estimate is even smaller at a decrease by 0.01 persons. This makes it unlikely that the negative effect of the main estimates on migration is purely driven by unobserved characteristics of the grid locations.

At individual level, future grid lines show no effect on average nor the subgroups of the spouses, the children and the grandchildren of the household head. The results show a small and weakly significant effect on migration of the household head. The latter seems negligible given its small size and the lack of significant effects on the other subgroups.

Results on agricultural productivity show no effect on most indicators, except for household food consumption. The latter is however substantially smaller than the coefficient of the main results (Table B-9).

6.4 Road construction

While the use of the least cost path instrument addresses demand side factors of electrification, it cannot solve the issues that cost assessments for the construction of other types of infrastructure would favor the same location. This would bias the results if other types of infrastructure were constructed during the treatment locations during the same time period.

This concern can be addressed by directly controlling for the construction of alternative types of infrastructure. Road infrastructure is the most obvious suspect for an omitted variable bias, since the costs of construction are determined by very similar features. To test for a potential bias from road construction, I run regressions on main outcomes controlling for all primary and secondary roads constructed during my sample period. During the time period of my study, the government implemented a large federal road maintenance program that resulted in a number of restored primary and secondary roads. Data on the date of constructions stems from publicly available materials by the Nigerian Federal Road Maintenance Management Agency (FERMA). I combine information on newly constructed or restored roads with their current geographic locations based on OpenStreetMaps (OpenStreetMap, 2018). In then define a binary road treatment variable as being within 15 km of a newly constructed road – similarly to the definition of the grid treatment variable.

Next, I replicate tables 3, 4 and 5 while controlling for the road treatment variable. Results are presented in appendix tables B-10, B-11 and B-12. to allow for easy comparison with the main results, Panel A in each table shows a replication of the respective main results table while controlling for road construction. Panel B of each table shows the coefficient of the road construction variable from the same regression as Panel A. Across all three tables, point estimates of grid treatment hardly differ from their original results in original specifications of tables 3, 4, and 5. This shows that road construction and grid construction did not happen in tandem during the observation period. While in other contexts clustering in infrastructure policies has been observed, this does not seem to had been the case in Nigeria. A possible explanation might be that the electricity grid and the road network are managed by separate ministries and in each case different donors were involved.

The regression results show no effect of road construction on household composition. At individual level, the coefficient for the average household member even shows a negative significant effect of −4 percent. This is a finding in itself since there is limited evidence on the effect of road construction of migration dynamics. Baum-Snow et al. (2017) and Baum-Snow et al. (2020) build the exception but find a positive effect on migration. This invites further research into the mediating factors that exclaim the diverging results in the case of Nigeria.

For agricultural production, the effect of roads shows a negative and highly significant coefficient on the agricultural production value. Moreover, household food consumption decreases slightly by 10 percent (at 10 percent significance). This shows that new roads affect the main outcome variables in a completely different way, making an omitted variable bias unlikely.

6.5 Media use

It is possible that media access caused omitted variable bias. Media access could have increased because related infrastructure was constructed during the same time or because access to electricity made device ownership more attractive. Previous studies have shown that access to mobile phones increases seasonal migration and remittances by reducing information frictions (Aker et al., 2011, Batista and Narciso, 2018). In contrast, access to private television has been linked to reduced internal migration (Farré and Fasani, 2013). I, therefore, regress ownership of media devices on grid expansion. Results show a statistically significant increase in TV ownership of 17 to 18 percent (appendix Table B-13). This is probably due to the rise in income rather than an expansion of the television network. However, it can not be discarded that by wider use of television information friction were reduced. Since this should lead household to correct their expected returns from migration downwards, it should not be a concern for the quality of my main results. In addition, the estimate on internet usage shows a statistically significant negative effect. In the first wave, only 4 households owned an internet connection. In the previous waves internet ownership increased in both the control and the treatment group but remained low. Therefore, the negative point estimate is unlikely to be causal. Importantly, I do not find a statistically significant change in mobile phone ownership. Therefore, the increase in migration is unlikely to be caused by improved connectivity between locations.

6.6 Mobile network coverage

Finally, I test for omitted variable bias from mobile phone infrastructure. As outline above infrastructure investments often happen in tandem. While mobile phone ownership shows no increase, it is still possible that improvements in mobile phone signal drive the effect. To test this, I use data on the 3G mobile phone network from the Collins Bartholomew - Mobile Coverage Explorer (Collins Bartholomew, 2021). This data offer annual shapefiles for the area covered by the mobile networks. The observation period saw the introduction of the 3G network in Nigeria which bear high potential in reducing information frictions (Aker et al., 2011, Batista and Narciso, 2018). I replicate Tables 4 – 6 controlling for a dummy variable indicating whether a household location was within reach of the 3G mobile network. Results are reported in appendix Tables B-14 - B-16. In each table, Panel A reports the respective point estimates for the grid dummy and Panel B reports the corresponding point estimates for the 3G mobile network dummy from the same regressions. The introduction of the 3G dummy control variables affects the size of the main results only slightly. Point estimates for the mobile network dummy show different pattern than the grid dummy making it unlikely that mobile phone access drives the main results. The only indicator of the main results that just loses statistical significance is the number of children aged 13-18, while it's effect size and standard error remain very close the main results. Since the remaining outcomes remain statistically significant including the migration propensity of the children of the household head, this does not affec the main conclusions.

7 Gravity model

This section uses a gravity model of migration to analyze how the electricity supply shock affects dyadic migration pattern. Following the convention in the literature (Bryan and Morten, 2018, Blanchard and Kirchberger, 2020), I construct a directional dyadic mobility measure as following:

$$m_{odt} = \frac{i_{odt}}{i_{ot}} \times 100 \tag{6}$$

where i_{od} is the number of individuals that where reported to have moved from origin district o to destination district d at time t and i_o number of individual that where reported to reside in origin district o at time t. On average, the mobility measure m_{odt} is 0.01 percent, since 99.77 percent of the dyadic flows are 0 (1,000,507 out of 1,002,850 observations). For non-zero flows the average is 4.03 percent. Aggregated over all potential destinations, 6.35 percent of a municipalities population moves to another municipality in every wave. To interpret the mobility measure correctly, a few features of its construction need to be considered. First, the measure captures only migration that happened since the last survey wave, i.e. within the last 3 years. Other studies often focus on lifetime migration and find substantially larger numbers. For instance, an older estimate for Nigeria by Mberu (2005) assumes that 58.3 percent of the rural-born population are living as migrants in 1993. This number includes migrants that eventually return to their home location. As Lucas (2021) note, a substantial share of African migrants returns within 5 years of leaving their origin destination. The well-cited paper by Bryan and Morten (2018) finds a somewhat smaller number for Indonesia where on average of 35.8 percent of the population migrate during their lifetime. Second, it does not include within district migration. While this number can be expected to be sizable, the measure is by construction ignorant to this type of migration. Particularly, for moves from rural or peri-urban areas to the closest urban centers are not captured in the measure. Third, the measure is ignorant about the permanence of a migration move. The questionnaire simply asks respondents whether or not a household member currently resides with the household. Thus, some share of seasonal migration will be contained in the measure.

Following Bryan and Morten (2018) I run regressions of the form:

$$log(m_{odt}) = \gamma_o + \gamma_d + \gamma_t + \beta Grid_{ot} + \delta log(dist_{od}) + \lambda X_{dt} + \epsilon_{odt} \tag{7}$$

where γ_o are origin fixed effects, γ_{dt} are destination fixed effects, γ_t are year fixed effects, $Grid_{ot}$ is dummy variable that indicates new grid construction at origin district, $dist_{od}$ is distance between origin and destination district, X_{dt} is a vector of destination characteristics in year t and ϵ_{odt} is the error term.

The destination characteristics include the percentage of land area of the destination district covered in cropland and the percentage covered in urban land as a proxy for urban/rural characteristics of the location. Data on cropland and urban land comes from the Climate Change Initiative Land Cover Maps dataset (CCI-LC) by the European Space Agency (European Space Agency, 2019). The data provides annual global land cover information for 22 different land cover categories defined by the UN Land Cover Classification System at a spatial resolution of $300m \times 300m$. I define every pixel as agricultural area that is classified as ”cropland, rainfed” or ”cropland, irrigated or post-flooding” in the CCI-LC dataset, while ”urban” constitutes an existing class in the dataset. Due to the fixed effects structure of the regressions, the estimates refer to changes in rural (or urban) area respectively. In addition, I include a dummy variable for grid construction at destination d in year t. Origin and destination municipalities are coded as receiving a grid in year t if one of the new transmission lines intersects with the boundaries of the administrative area. A balancing test between municipalities that received new grid and those who do not finds no significant difference with respect to road density, population, cropland and urban area (appendix Table A-5).

Table 7 represents average effects from all origin-destination pairs, while Table 8 presents a sample split between origins with new grid and those without. Column (1) of Table 7 shows the average effect of the distance variable in my sample. On average, a 1 percent reduction in migration costs in the form of distance between two municipalities results in an approximately 0.01 percent increase in migrants. This relatively small effect is partly driven by the fact that dyadic migration flows are on average only 0.01 percent of the origin location. The effect amounts to approximately 2.5 times the standard deviation of the dependent variable. For this reason, the effect is sizable. In addition, it could indicate that migration costs are not the main barrier to migration in Nigeria. The results on grid construction at the origin municipality are qualitatively in line with previous results from the household panel. After grid construction, the out-going migration flow increases significantly by 0.001-0.003 percent. The small effect size amounts to between 0.4 to 1 standard deviation of the dependent variable and seems therefore relevant. Percentage of cropland and percentage of urban area of the destination municipality have large effects of migration flows. An increase of cropland by 1 percent reduces out-going migration by 0.23 percent, an increase of urban land by 1 percent increases out-going migration by 0.21 percent. Grid construction at destination has on average no significant effect on migration flows.

Next, I perform a sample split to analyze how destinations that have received new grid differ from those that did not. Results are reported in Table 8. Column (1) shows results for origins that did not receive new grid in time t, column (2) shows results for origins that did. Column (3) reports the difference between both coefficients. Across specifications the effect of distance is 3 to 4 times smaller for origins that have received a new. This can be interpreted in two ways: first, migrants from origins that have received new grid migrate over larger distance, and second, the effect of migration costs on migration flows seems to fall. This provides additional salience of credit constrains. Given that the productivity shock increased household incomes by 23 percent, this would imply that raising incomes by 1 percent would increase migration by 13 to 17 percent on average. What is more, migrants from origins that have new grid are 3 times more likely to go to destinations with expanding agricultural area and 3 times less likely to go to destinations with expanding urban area compared to those from origins that did not receive new grid. This finding indicates that the productivity shock changes preferences over destination characteristics. While the average Nigerian migrant showed strong preferences for migration to urban areas, new migrants from origins that received a productivity shock seem to prefer destinations that share more characteristics with their origin. These characteristics are not only linked to the agro-climatic conditions or sectoral composition, but also to the production technology. It is therefore in line with the hypothesis that the productivity shock increased task specific human capital such that returns to migration increased in destinations that possess the same sectors and technologies.

8 Concluding remarks

This paper provides first evidence how investments in electricity infrastructure affect internal migration. Using the expansion of the electric transmission grid in Nigeria in the years 2009 to 2015, I show that the intervention had significant positive effect on out-migration. This seems driven by a increased access to credit with simultaneous lack of employment generation for the youth. While household food consumption increased substantially, the economic boom did not seem to benefit everyone equally. Young adults and older teenagers that suffer particularly from underemployment in Nigeria did neither show an increase in employment nor in working hours. Apparently, the rise in labor demand only affected older, more experienced workers. Instead, we observe a rise in out-migration by this subgroup. The results suggests that this migration spike is mainly labor migration. We also observe that the effect of movement costs on migration decreased by factor 3 for these migrants. This suggests large efficiency gains from easing credit constrains.

Overall, the findings suggest that closing the rural-urban gap with infrastructure investments is extremely difficult. Despite large income gains of the intervention, for a large subgroup of the population employment creation was not sufficient. While raising productivity through public investments is an important tool to harmonize economic activity across space, in the short term youth unemployment might best be tackled by easing credit constrains to enable migration. Policy-makers should, therefore, combine rural infrastructure investments with migration oriented cash transfers to address the rural-urban gap effectively.

The findings of this paper are, however, limited to the short term. While in the short term, employment opportunities might be limited, demand for young, less experienced workers might rise in the long term. It is also not clear whether the observed youth migration is permanent. Since personal costs of living away from ones origin seem to be high, it is possible that young migrants return to their origin locations after collecting more work experience and/or education. In the long term, population dynamics might therefore reverse. This is, however, only possible if economic growth at origin continues, highlighting again the importance of structural investment. Further research is warranted to understand these long-term effects.

Finally, the paper sheds new light on how the electricity shocks affects the ordinal preferences for destination. Following an electricity supply shock migrants are more likely to migrate to rural destinations that also received new grid infrastructure. This finding suggests that the intervention changed not only the household budget, but also the relative expected returns from migration to each destination. This could be linked , for instance, to human capital effects in the form of learning-by-doing or task-specific human capital that is tied to characteristics of the location. Additional research is, therefore, needed to understand how infrastructure investments affect to ordinal preferences for migration destinations, particularly as a tool to channel migration flows consciously.

References

Adhikari, S. and Gentilini, U. (2018), 'Should I Stay or Should I Go: Do Cash Transfers Affect Migration?', *World Bank Policy Research Paper* **8525**(July).

Aker, J. C., Clemens, M. A. and Ksoll, C. (2011), 'Mobiles and mobilit: The Effect of Mobile Phones on Migration in Niger', *Proceedings of the German Development Economics Conference, Berlin 2011, No.2. ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg.* .

Alesina, A., Michaelopoulos, S. and Papaioannou, E. (2016), 'Ethnic Inequality', *Journal of Political Economy* **124**(2).

Allen, T. and Arkolakis, C. (2014), 'Trade and Topography of the Spatial Economy', *Quarterly Journal of Economics* (2002), 1085–1139.

Amrevurayire, E. O. and Ojeh, V. N. (2016), 'Consequences of Rural-Urban Migration on the Source Region of Ughievwen Clan Delta State Nigeria', *European Journal of Geography* **7**(3), 42–57.

Angelucci, M. (2015), 'Migration and financial constraints: Evidence from mexico', *Review of Economics and Statistics* **97**(1), 224–228.

Bah, T., Batista, C., Gubert, F. and McKenzie, D. (2020), 'Information and Alternatives to Irregular Migration', *AEA RCT Registry* .

Banerjee, A., Duflo, E. and Qian, N. (2020), 'On the road: Access to transportation infrastructure and economic growth in China', *Journal of Development Economics* **145**(February), 102442.
URL: *https://doi.org/10.1016/j.jdeveco.2020.102442*

Batista, C. and Narciso, G. (2018), 'Migrant remittances and information flows: Evidence from a field experiment', *World Bank Economic Review* **32**(1), 203–219.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A. and Zhang, Q. (2017), 'Roads, Railroads, and Decentralization of Chinese Cities', *Review of Economics and Statistics* **99**(3), 435–448.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q. and Brandt, L. (2020), 'Does investment in national highways help or hurt hinterland city growth?', *Journal of Urban Economics* **115**(September 2017).

Bayer, P., Kennedy, R., Yang, J. and Urpelainen, J. (2020), 'The need for impact evaluation in electricity access research', *Energy Policy* **137**(March 2019), 111099.
URL: *https://doi.org/10.1016/j.enpol.2019.111099*

Bazzi, S. (2017), 'Wealth heterogeneity and the income elasticity of migration', *American Economic Journal: Applied Economics* **9**(2), 219–255.

Bensch, G., Gotz, G. and Peters, J. (2020), 'Effects of rural electrification on employment:

A comment on Dinkelman (2011)', *Ruhr Economic Papers, No. 840* .

Blanchard, P. and Kirchberger, M. (2020), 'Perputual Motion:Human Mobility and Spatial Frictions in Three African Countries', *CSAE Working Paper* **44**(0).
URL: *https://sites.google.com/site/mkirchberger.*

Bryan, G., Chowdhury, S. and Mobarak, A. M. (2014), 'Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh', *Econometrica* **82**(5), 1671–1748.

Bryan, G. and Morten, M. (2018), 'The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia', *Journal of Political Economy* .

Bryan, G., Shyamal, C., Mobarak, A. M., Morten, M. and Smits, J. (2021), 'Encouragement and Distortionary Effects of Conditional Cash Transfers', *IZA Discussion Paper* (14326).

Burlig, F. and Preonas, L. (2016), 'Out of the Darkness and Into the Light? Development Effects of Rural Electrification in India', (April), 1–38.

Clemens, M. A. (2020), 'The Emigration Life Cycle : How Development Shapes Emigration from Poor Countries The Emigration Life Cycle : How Development Shapes Emigration from Poor Countries', (13614).

Collins Bartholomew (2021), 'Collins bartholomew - mobile coverage explorer 2011 - 2018 [dataset]'. Available at: https://www.collinsbartholomew.com/mobile-coverage-maps/mobile-coverage-explorer/ (accessed March 8, 202).

Dillon, A., Mueller, V. and Salau, S. (2011), 'Migratory responses to agricultural risk in northern Nigeria', *American Journal of Agricultural Economics* **93**(4), 1048–1061.

Dinkelman, T. (2011), 'The Effects of Rural Electrification on Employment : New Evidence from South Africa', *The American Economic Review* **101**(7), 3078–3108.

European Space Agency (2019), 'Climate change initiative - land cover map version 2.0.7 [dataset]'. Available at: http://maps.elie.ucl.ac.be/CCI/viewer/download.php/ (accessed August 23, 2019).

Faber, B. (2014), 'Trade integration, market size, and industrialization: Evidence from China's national trunk highway system', *Review of Economic Studies* **81**(3), 1046–1070.

Farr, T. and Kobrick, M. (2000), 'Shuttle Radar Topography Mission produces a wealth of data', *American Geophysical Union Eos* **81**, 583–585.

Farré, L. and Fasani, F. (2013), 'Media exposure and internal migration - Evidence from Indonesia', *Journal of Development Economics* **102**, 48–61.
URL: *http://dx.doi.org/10.1016/j.jdeveco.2012.11.001*

Fetter, T. R. and Faraz, U. (2020), 'Fracking, farmers, and rural electrification in India', *Ruhr Economic Papers* **864**.
URL: *http://dx.doi.org/10.4419/96973001*

Fried, S. and Lagakos, D. (2021), 'Rural electrification, migration and structural

transformation: Evidence from Ethiopia', *Regional Science and Urban Economics* (December), 103625.
URL: *https://doi.org/10.1016/j.regsciurbeco.2020.103625*

GADM (2015), 'Nigeria shapfile version 2.8 [dataset]'.
URL: *https://biogeo.ucdavis.edu/data/gadm3.6/shp/gadm36˙NGA˙shp.zip*

Gatugel, Z., Abbasoglu, S., Tekbiyik, N. and Fahrioglu, M. (2015), 'Transforming the Nigerian power sector for sustainable development', *Energy Policy* **87**, 429–437.
URL: *http://dx.doi.org/10.1016/j.enpol.2015.09.004*

GeoNames (2020), 'GeoNames Gazetteer extract files Nigeria [dataset]'.
URL: *http://download.geonames.org/export/dump/NG.zip*

Gollin, D., Lagakos, D. and Waugh, M. E. (2014), 'The agricultural productivity gap', *Quarterly Journal of Economics* **129**(2), 939–993.

Grogan, L. and Sadanand, A. (2013), 'Rural Electrification and Employment in Poor Countries: Evidence from Nicaragua', *World Development* **43**, 252–265.
URL: *http://dx.doi.org/10.1016/j.worlddev.2012.09.002*

Henderson, V. (2002), 'Urbanization in Developing Countries', *The World Bank Research Observer* **35**(1), 45–45.

IEA (2019), 'World energy outlook 2019'. OECD Publ.

Jha, M., M. C. and Schonfeld, P. (2001), 'Using GIS, genetic algorithms, and visualization in highway development', *Computer-Aided Civil and Infrastructure Engineering* (16), 399–414.

Jong, J. and Schonfeld, P. (2003), 'Transportation Research Part B: Methodological', *An evolutionary model for simultaneously optimizing three-dimensional highway alignments* (37), 107–128.

Kassem, D. (2020), 'Does Electrification Cause Industrial Development? Grid Expansion and Firm Turnover in Indonesia', *Unpublished manuscript* (143), 102412.

Lagakos, D., Mushfiq Mobarak, A. and Waugh, M. E. (2018), 'The Welfare Effects of Encouraging Rural-Urban Migration', *NBER Working Paper* .

Lee, K., Miguel, E. and Wolfram, C. (2019), 'Does Electrification Supercharge Economic Development?', *Working Paper* .

Lee, K., Miguel, E. and Wolfram, C. (2020), 'Does Household Electrification Supercharge Economic Development?', *The Journal of Economic Perspectives* **34**(1), 122–144.
URL: *https://www.jstor.org/stable/10.2307/26873532*

Lenz, L., Munyehirwe, A., Peters, J. and Sievert, M. (2017), 'Does Large-Scale Infrastructure Investment Alleviate Poverty ? Impacts of Rwanda ' s Electricity Access Roll-Out Program', *World Development* **89**, 88–110.

Lessmann, C. (2014), 'Spatial inequality and development - Is there an inverted-U relationship?', *Journal of Development Economics* **106**, 35–51.

URL: *http://dx.doi.org/10.1016/j.jdeveco.2013.08.011*

Lewis, J. and Severnini, E. (2020), 'Short- and long-run impacts of rural electrification: Evidence from the historical rollout of the U.S. power grid?', *Journal of Development Economics* (143), 102412.

Lipscomb, M., Mobarak, A. M. and Barham, T. (2013), 'Development Effects of Electrification : Evidence from the Topographic Placement of Hydropower Plants in Brazil', *American Economic Journal: Applied Economics* **5**(2).

Lucas, R. E. (2021), *Crossing the Rural-urban Divide: A Compilation of Prior and Fresh Evidence from Developing Countries*, Oxford University Press.

Mberu, B. U. (2005), 'Who moves and who stays? Rural out-migration in Nigeria', *Journal of Population Research* **22**(2), 141–161.

Mckenzie, D. and Rapoport, H. (2007), 'Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico', *Journal of Development Economics* **84**(1), 1–24.

Molina Millán, T., Macours, K., Maluccio, J. A. and Tejerina, L. (2020), 'Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program', *Journal of Development Economics* **143**(September 2018).

NERC (2019), 'Transmission [website]'.
URL: *https://nerc.gov.ng/index.php/home/nesi/404-transmission*

Nigeria National Bureau of Statistics (2017), 'General Household Survey, Panel (GHS-Panel) 2015-2016. 2011-12. 2010'.
URL: *Dataset downloaded from www.microdata.worldbank.org on [30.11.2017]*

Ofuoku, A., Okpara, O. and Obakanurhe, O. (2017), 'The Impact of Rural-Urban Migration on Poultry Production in the Niger Delta Region, Nigeria', *International Journal of Agricultural Science, Research and Technology in Extension and Education Systems* **6**(1), 13–20.

OpenStreetMap (2018), 'Nigeria Settlement Data'.

OpenStreetMap (2020), 'OpenStreetMap Nigeria [dataset]'. Download via Geofabrik GmbH: `http://download.geofabrik.de/africa/nigeria-latest-free.shp.zip`.

Pestana, C., Ibiowie, A. and Managi, S. (2014), 'Nigeria ' s power sector : Analysis of productivity', *Economic Analysis and Policy* **44**(1), 65–73.
URL: *http://dx.doi.org/10.1016/j.eap.2014.02.003*

Roback, J. (1982), 'Wages, Rents, and the Quality of Life', *Journal of Political Economy* **90**(6), 1257–1278.

Rural Electrification Agency (2020), 'Energy database [dataset]'. Available at: `https://database.rea.gov.ng/` (accessed January 17, 2020).

Rural Electrification Agency (2021), 'REA policy objectives [website]'. Last retrieved September 27, 2021.

URL: *https://rea.gov.ng/rea-policy-objectives/*

Sadiq, A., Dada, J. O. and Khalil, I. (2015), 'Current status and future prospects of renewable energy in Nigeria', *Renewable and Sustainable Energy Reviews* **48**, 336–346.
URL: *http://dx.doi.org/10.1016/j.rser.2015.03.098*

Sjaastad, L. A. (1962), 'The Costs and Returns of Human Migration', *Journal of Political Economy* **70**(5), 80–92.
URL: *https://sci-hub.tw/10.1002/iroh.19620470121*

van de Walle, D., Ravallion, M., Mendiratta, V. and Koolwal, G. (2013), Long-Term Impacts of Household Electrification in Rural India.
URL: *http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-6527*

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.

World Bank (2009), 'Project Appraisal Document: Nigeria - Electricity and Gas Improvement Project (English)'.
URL: *http://documents.worldbank.org/curated/en/796771468288655228/pdf/479450PAD0IDA1101*

World Bank (2016), 'More, and More Productive, Jobs for Nigeria: A Profile of Work and Workers', (Washington, D.C.).
URL: *http://documents.worldbank.org/curated/en/650371467987906739/More-and-more-productive-jobs-for-Nigeria-a-profile-of-work-and-workers*

World Bank (2017), 'World Development Indicators'.

WorldPop and Center for International Earth Science Information Network (CIESIN), C. U. (2018), 'Global high resolution population denominators project [dataset]'. Funded by The Bill and Melinda Gates Foundation (OPP1134076). Available at: `https://dx.doi.org/10.5258/SOTON/WP00674` (accessed November 11, 2020).

Young, A. (2013), 'Inequality, the Urban-Rural Gap, and Migration', *Quarterly Journal of Economics* pp. 1727–1785.

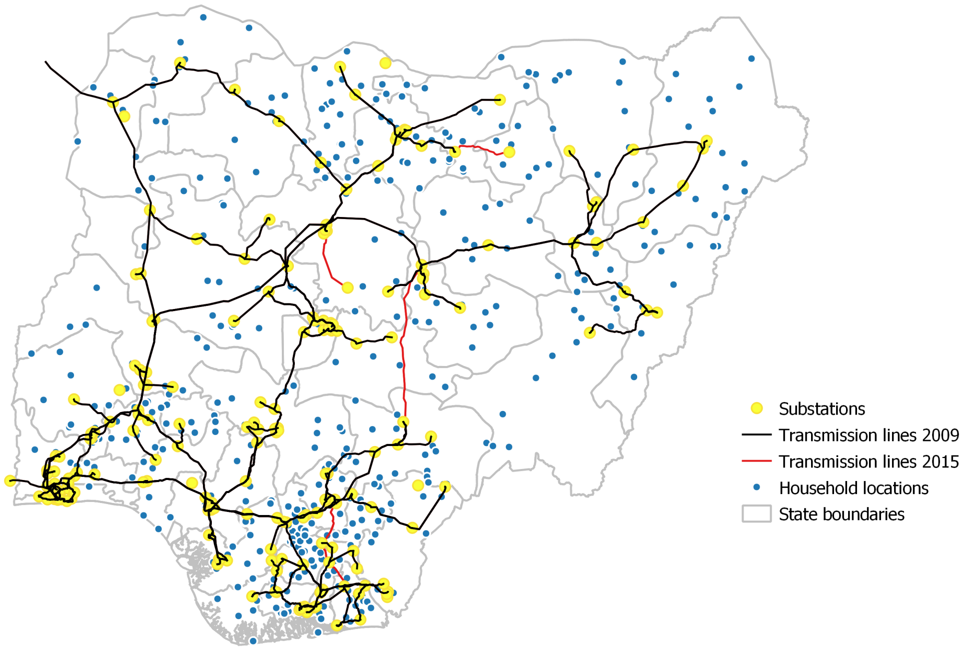

Figure 1 – Location of households, transmission lines and substations

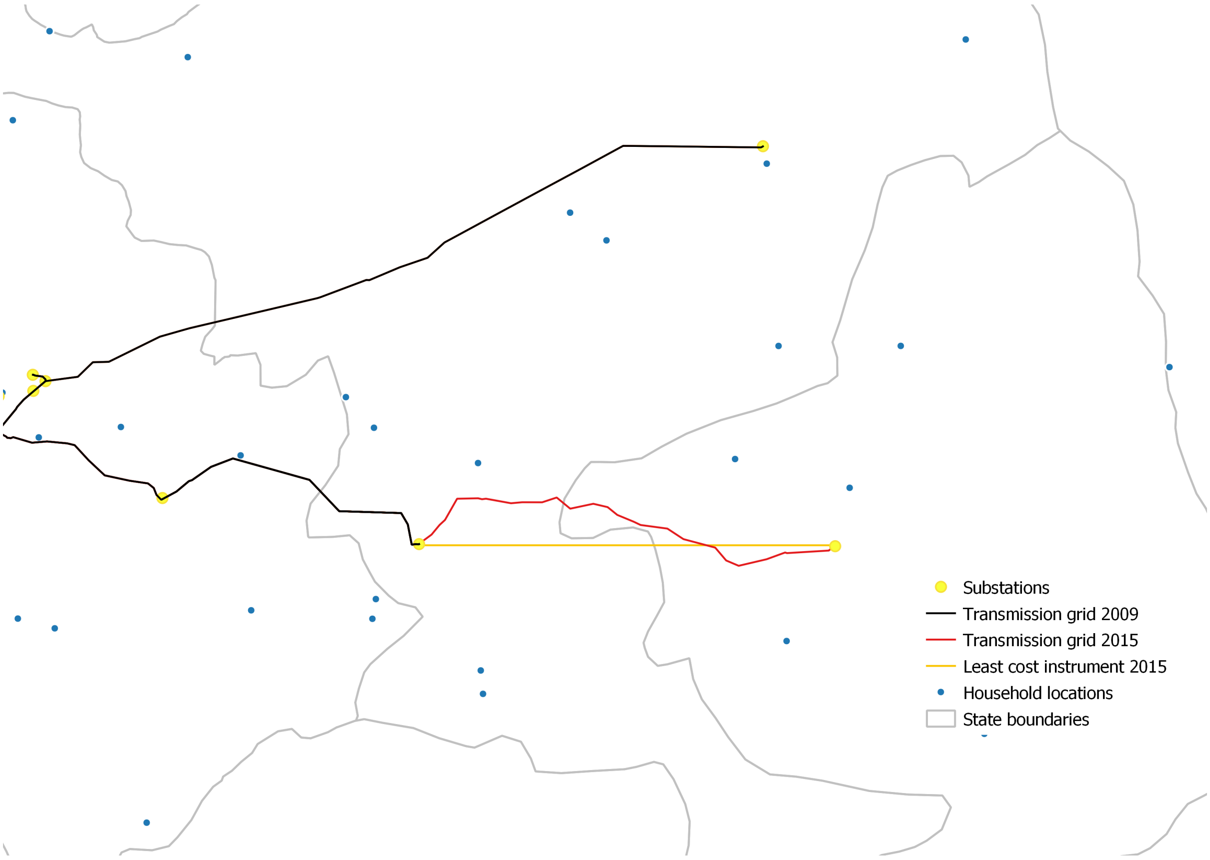

Figure 2 – Example of actual and least cost grid

Table 1 – Balancing between treatment and control households in 2009

		Control	SD	treatment	SE	N
Electrified		0.380	0.485	-0.061	0.066	3,615
(Log) grid distance		9.775	1.317	0.182	0.249	3,632
(Log) substation distance		10.604	0.640	-0.050	0.146	3,632
(Log) distance to capital		11.003	0.776	-0.499	0.311	3,632
(Log) road distance		4.007	1.496	-0.287	0.272	3,632
% of cropland		0.426	0.317	0.011	0.027	3,632
Population density		2.109	3.129	-0.866	0.652	3,632
% of urban land		0.003	0.016	-0.000	0.001	3,632
Rural		0.824	0.381	-0.050	0.089	3,632
House value		975,496.668	2,743,567.399	310,147.835	245,754.801	2,867
Roof	Grass	0.209	0.407	-0.085**	0.041	3,611
	Iron sheets	0.702	0.457	0.103**	0.042	3,611
	Clay tiles	0.015	0.123	0.003	0.009	3,611
	concrete	0.010	0.099	0.005	0.014	3,611
	Plastic sheeting	0.009	0.093	-0.001	0.001	3,611
	Asbestos sheet	0.020	0.140	-0.002	0.002	3,611
	Other	0.035	0.184	-0.024	0.017	3,611
Walls	Grass	0.077	0.267	-0.010	0.024	3,603
	Mud	0.445	0.497	-0.059	0.064	3,603
	Compacted earth	0.035	0.184	-0.012	0.016	3,603
	Mud bricks (unfired)	0.065	0.247	-0.049**	0.021	3,603
	Burnt bricks	0.014	0.118	-0.015*	0.008	3,603
	Concrete	0.347	0.476	0.146**	0.070	3,603
	Wood	0.012	0.110	-0.001	0.001	3,603
	Iron sheets	0.004	0.061	-0.001	0.001	3,603
Lighting fuel	Collected firewood	0.092	0.289	-0.018	0.022	3,607
	Purchased firewood	0.031	0.174	0.001	0.017	3,607
	Kerosene	0.390	0.488	0.062	0.054	3,607
	Electricity	0.203	0.402	-0.029	0.039	3,607
	Generator	0.032	0.176	-0.014*	0.008	3,607
	Battery	0.210	0.407	0.010	0.047	3,607
	Other	0.042	0.202	-0.012	0.020	3,607
(Log) agri production value		10.366	3.145	0.162	0.351	2,767
(Log) daily wage, men		3.945	3.517	-0.242	0.226	1,821
(Log) daily wage, women		2.151	3.133	-0.368	0.234	1,542
# of paid workers (men)		2.393	5.567	-0.094	0.351	2,715
# of paid workers (women)		1.449	6.593	-0.391	0.275	2,715
# of plots		1.560	1.397	-0.075	0.261	3,632
HH size		5.900	3.016	0.088	0.335	3,632
# of elderly		0.075	0.311	0.044*	0.026	3,632
# of children		3.236	2.479	0.024	0.298	3,632

Test of joint significance	F-stat	p-value
	0.43	0.998

Note: Balancing is tested using a regression with state fixed effects and standard errors clustered at village-level as in the main regressions. (*** p<0.01, ** p<0.05, * p<0.1)

Table 2 – The effect of new transmission lines on household electrification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	- (Log) grid distance				Dummy grid			
	OLS		2SLS		OLS		2SLS	
Panel A: Main regression								
INDICATOR	0.057*	0.055	0.108***	0.104***	0.180**	0.177**	0.539***	0.536***
	(0.034)	(0.033)	(0.039)	(0.040)	(0.085)	(0.084)	(0.139)	(0.153)
Substations	-0.029	-0.037	-0.033	-0.034	-0.028	-0.036	-0.034	-0.031
	(0.029)	(0.032)	(0.033)	(0.033)	(0.028)	(0.030)	(0.036)	(0.036)
Capital		0.012		0.002		0.012		-0.013
		(0.018)		(0.020)		(0.017)		(0.022)
Road distance		0.003		0.004		0.002		0.000
		(0.006)		(0.006)		(0.006)		(0.008)
% Cropland		0.071		0.077		0.065		0.066
		(0.081)		(0.082)		(0.080)		(0.084)
Population density		0.005		0.006		0.006		0.009
		(0.006)		(0.006)		(0.006)		(0.006)
% Urban		1.571		1.213		1.452		0.385
		(6.115)		(6.190)		(5.985)		(6.100)
Observations	2,289	2,289	2,289	2,289	2,289	2,289	2,289	2,289
R-squared	0.029	0.032			0.033	0.036		
Panel B: First stage results								
Least cost IV			0.8633***	0.8551***			0.866***	0.863***
			(0.0512)	(0.0477)			(0.106)	(0.116)
Substations			0.2059*	0.1739*			0.037	0.035
			(0.1130)	(0.0988)			(0.038)	(0.036)
Capital				0.0667**				0.020
				(0.0299)				(0.021)
Road distance				-0.0092				0.008
				(0.0115)				(0.007)
% Cropland				-0.1836				-0.020
				(0.1150)				(0.041)
Population density				-0.0197*				-0.010**
				(0.0110)				(0.005)
% Urban				-0.4542				-1.269
				(7.1669)				(2.421)
F-stat			284.3	321.3			66.89	55.66

Note: All regression control for wave-state fixed effects. Dependent variable is a binary measure for household electrification. Each column reports results for a different measure of grid construction: columns (1) and (2) report results on negative log distance to actual grid location, columns (3) and (4) on 2SLS estimates for negative log distance instrumented through the hypothetical least cost grid path, columns (5) and (6) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid, and columns (7) and (8) on the grid dummy instrumented by a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid. For all control variables distances are measured as negative logarithmic distance. Substations is the negative least cost distance to any substation in 2015, capital is the one to the state capital, road distance is the negative logarithmic distance to any primary or secondary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table 3 – The effect of new transmission lines on household composition

	(1)	(2)	(3)	(4)	(5)
	Baseline	OLS		2SLS	
	mean	no controls	controls	no controls	controls
# of household members	5.963	-0.330**	-0.350**	-0.691***	-0.778***
		(0.140)	(0.150)	(0.196)	(0.226)
# of elderly	0.071	-0.061*	-0.060*	-0.038	-0.045
		(0.033)	(0.034)	(0.032)	(0.036)
# of children (total)	3.259	-0.300***	-0.329***	-0.506***	-0.576***
		(0.102)	(0.099)	(0.148)	(0.142)
# of children (age 0-5)	1.176	-0.207**	-0.222**	-0.174*	-0.259**
		(0.093)	(0.096)	(0.096)	(0.099)
# of children (age 6-12)	1.301	0.064	0.053	-0.020	0.003
		(0.071)	(0.070)	(0.102)	(0.102)
# of children (age 13-18)	0.802	-0.137	-0.137	-0.335**	-0.335**
		(0.089)	(0.089)	(0.133)	(0.135)
Observations		2,259	2,259	2,259	2,259
F-stat				66.430	55.295

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean for the year 2009. columns (2) and (3) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, and columns (4) and (5) on the grid dummy instrumented by a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. For all control variables distances are measured as negative logarithmic distance. Control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital, road distance is the negative logarithmic distance to any primary or secondary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table 4 – The effect of new transmission lines on migration (individual level)

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
All HH members	0.019	0.015	0.016	0.054*	0.056*	15,993	100.701
		(0.016)	(0.017)	(0.027)	(0.028)		
HH head	0.003	0.005	0.006*	0.011	0.012*	2,716	68.260
		(0.003)	(0.003)	(0.007)	(0.007)		
HH spouse	0.035	-0.018	-0.019	-0.057**	-0.058***	2,536	96.434
		(0.017)	(0.016)	(0.024)	(0.020)		
HH child	0.091	0.030	0.030	0.099**	0.102**	9,338	102.018
		(0.023)	(0.024)	(0.041)	(0.043)		
HH grandchild	0.159	0.103	0.194**	0.023	0.210*	564	164.612
		(0.089)	(0.086)	(0.074)	(0.110)		
Other	0.180	0.058	0.045	0.130	0.163	828	43.684
		(0.081)	(0.083)	(0.233)	(0.238)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Control variables of columns (3) and (5) include distance to the state capital, distance to the closest secondary or primary road in 2009, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table 5 – The effect of new transmission lines on agricultural production

		(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
Output	(Log) production value	10.121	0.203	0.101	0.532	0.545	1,876	36.634
			(0.518)	(0.505)	(0.913)	(0.910)		
Factor inputs	(Log) labor costs	1.436	0.048	0.015	0.939**	0.856**	1,885	40.376
			(0.460)	(0.464)	(0.387)	(0.364)		
	(Log) # of paid workers	0.600	-0.092	-0.092	0.066	-0.002	1,885	40.376
			(0.094)	(0.088)	(0.149)	(0.150)		
	# of plots	1.784	0.166	0.212	0.680*	0.766**	2,323	55.700
			(0.258)	(0.263)	(0.338)	(0.357)		
Profit	(Log) food consumption	4.011	0.081**	0.082**	0.257***	0.271***	2,250	55.241
			(0.039)	(0.038)	(0.074)	(0.085)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean for the year 2009. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. For all control variables distances are measured as negative logarithmic distance. Control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital, road distance is the negative logarithmic distance to any primary or secondary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table 6 – The effect of new transmission lines on employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Non-farm work		Farm work		Working hours		Obs	F-stat
	OLS	2SLS	OLS	2SLS	OLS	2SLS		
All	0.002	-0.011	-0.040	-0.012	-0.488	1.329	12,808	146.481
	(0.019)	(0.019)	(0.045)	(0.071)	(1.670)	(1.822)		
HH head	0.075**	0.121**	0.001	-0.008	4.870**	10.072***	2,696	68.115
	(0.031)	(0.048)	(0.069)	(0.076)	(2.215)	(2.868)		
HH spouse	0.000	0.025	0.010	0.084	-3.379	5.541	2,387	92.864
	(0.065)	(0.066)	(0.065)	(0.141)	(5.032)	(5.393)		
HH child	-0.026*	-0.046**	-0.052	-0.012	-0.476	0.356	6,808	197.905
	(0.014)	(0.019)	(0.069)	(0.101)	(1.816)	(2.945)		
Other	0.084	-0.012	-0.055	0.204	1.116	-1.081	594	1,469.557
	(0.060)	(0.133)	(0.097)	(0.252)	(3.234)	(4.109)		

Note: All regression control for wave-state fixed effects and distance to the closest substation, the negative logarithmic distance to the state capital, negative logarithmic distance to secondary or primary roads in 2009, % Cropland, % urban landcover and population density measured within a 40km buffer. Columns (1), (3) and (5) show regression results on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid, and columns (2), (4) and (6) show regression results on the grid dummy instrumented by a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid. Column (7) reports number of observations and column (7) display Kleibergen-Papp-F-statistics of column (6) results. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table 7 – Gravity model estimates - effect on (log) migrants (pooled)

Dependent variable = $log(m_{odt})$	(1)	(2)	(3)	(4)	(5)
$Grid_{ot}$		0.001**	0.003**	0.003**	0.001**
		(0.001)	(0.001)	(0.001)	(0.001)
$log(dist_{od})$	-0.007***	-0.007***	-0.014***	-0.014***	-0.007***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
$\%\ Cropland_{dt}$			-0.232***		
			(0.033)		
$\%\ Urban_{dt}$				0.211***	
				(0.023)	
$Grid_{dt}$					-0.002
					(0.001)
Destination FE			x	x	x
Origin FE		x	x	x	x
Wave FE			x	x	x
Destination-Wave FE	x	x			
Origin-Wave FE	x				
Observations	1,001,556	1,001,556	498,493	498,493	1,001,556

Note: The dependent variable of all regressions is $log(m_{odt})$, i.e. the annual logarithmic percentage of migrants from origin o to destination d relative to total inhabitants of o. $Grid_{ot}$ is a dummy variable indicating construction of new transmission lines within the boundaries of district o, $log(dist_{od})$ is the logarithmic distance between the centroids of origin district to destination district, $\%Cropland_{dt}$ is the percentage of destination district's land area covered in cropland, $\%Urban_{dt}$ is the percentage of destination district's land area covered in urban area, and $Grid_{dt}$ is a dummy variable indicating construction of new transmission grid within the boundaries of the destination district. Standard errors are two-way clustered at year and municipality level and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table 8 – Gravity model estimates - effect on (log) migrants (sample split)

	(1) $Grid_{ot} = 0$	(2) $Grid_{ot} = 1$	(3) Difference (2) - (1)
Panel A: Heterogenous effect of cropland			
$Log(dist_{od})$	-0.0173***	-0.0054***	0.0120***
	(0.0007)	(0.0007)	(0.0010)
$\%\ Cropland_{dt}$	-0.2794***	-0.0938**	0.1856***
	(0.0419)	(0.0418)	(0.0592)
Panel B: Heterogenous effect of urban land			
$Log(dist_{od})$	-0.0173***	-0.0053***	0.0120***
	(0.0007)	(0.0007)	(0.0010)
$\%\ Urban_{dt}$	0.2578***	0.0824***	-0.1754***
	(0.0293)	(0.0248)	(0.0384)
Panel B: Heterogenous effect of urban land			
$Log(dist_{od})$	-0.0094***	-0.0025***	0.0068***
	(0.0004)	(0.0003)	(0.0005)
$Grid_{dt}$	-0.0031**	0.0033*	0.0064***
	(0.0013)	(0.0019)	(0.0023)
Observations	749,232	252,324	1,001,556

Note: The dependent variable of all regression is $log(m_{odt})$, i.e. the annual logarithmic percentage of migrants from origin o to destination d relative to total inhabitants of o. $Grid_{ot}$ is a dummy variable indicating construction of new transmission lines within the boundaries of district o, $log(dist_{od})$ is the logarithmic distance between the centroids of origin district to destination district, $\%Cropland_{dt}$ is the percentage of destination district's land area covered in cropland, $\%Urban_{dt}$ is the percentage of destination district's land area covered in urban area, and $Grid_{dt}$ is a dummy variable indicating construction of new transmission grid within the boundaries of the destination district. Standard errors are two-way clustered at year and municipality level and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

A Appendix A - Additional tables

Table A-1 – List of variables

Variable	Definition	Source
	Dependent variables	
Electrified	1 if household has access to grid electricity	Nigeria National Bureau of Statistics (2017)
# of household members	number of all household members	Nigeria National Bureau of Statistics (2017)
# of elderly	number of all household members above age 65	Nigeria National Bureau of Statistics (2017)
# number of children	number of all household members below age 19	Nigeria National Bureau of Statistics (2017)
Migration all HH members	1 if HH members left household since last wave	Nigeria National Bureau of Statistics (2017)
Migration HH head	1 if HH head left household since last wave	Nigeria National Bureau of Statistics (2017)
Migration HH spouse	1 if spouse of HH head left household since last wave	Nigeria National Bureau of Statistics (2017)
Migration HH child	1 if child of HH head left household since last wave	Nigeria National Bureau of Statistics (2017)
Migration HH grandchild	1 if grandchild of HH head left household since last wave	Nigeria National Bureau of Statistics (2017)
(Log) production value	logarithmic value of all harvest produced by the household (self-reported)	Nigeria National Bureau of Statistics (2017)
(Log) labor costs	aggregate cost of agricultural workers hired by the household for this season (self-reported)	Nigeria National Bureau of Statistics (2017)
# of plots	number of plots owned by the household	Nigeria National Bureau of Statistics (2017)
(log) food consumption	logarithmic good consumption per HH member (in the past 7 days)	Nigeria National Bureau of Statistics (2017)
Non-farm work	1 if HH member worked self-employed or outside the home for wage (in the past 7 days)	Nigeria National Bureau of Statistics (2017)
Farm-work	1 if HH member worked on a family farm (in the past 7 days)	Nigeria National Bureau of Statistics (2017)
Working hours	total working hours in primary and secondary employment (in the past 7 days)	Nigeria National Bureau of Statistics (2017)
$log(m_{odt})$	logarithmic fraction of migrants from municipality o that moved to municipality d in wave t divided by total number of residents of o in wave t	Nigeria National Bureau of Statistics (2017)

Independent variables		
- (Log) grid distance	negative logarithmic distance to the closest newly constructed transmission line	Nigeria National Bureau of Statistics (2017), Rural Electrification Agency (2020)
- (Log) grid distance instrument	negative logarithmic distance to the least cost path of the closest newly constructed transmission line	Nigeria National Bureau of Statistics (2017), Rural Electrification Agency (2020)
Dummy grid	1 if household was within 15 km distance of any newly constructed transmission line	Nigeria National Bureau of Statistics (2017), Rural Electrification Agency (2020)
Dummy grid instrument	1 if household was within 15 km of the least cost path of any newly constructed transmission line	Nigeria National Bureau of Statistics (2017), Rural Electrification Agency (2020)
$Grid_{ot}$	1 if new transmission lines were constructed within the boundaries of municipality o in year t	GADM (2015), Rural Electrification Agency (2020)
Control variables: baseline model		
Substations	negative logarithmic distance to the closest substation	Nigeria National Bureau of Statistics (2017), Rural Electrification Agency (2020)
Capital	negative logarithmic distance to the state capital	GeoNames (2020)
Road distance	negative logarithmic distance to any primary or secondary 2009 road	Own construction based on OpenStreetMap (2020)
% Cropland	percentage of area covered in cropland within a 40 km buffer	European Space Agency (2019)
Population density	Population density with a 40 km buffer	WorldPop and Center for International Earth Science Information Network (CIESIN)
% Urban	percentage of area covered in urban land within a 40 km buffer	European Space Agency (2019)
3G mobile network	dummy variable indicating that the location is within reach of the 3G mobile network	Collins Bartholomew (2021)
Control variables: gravity model		
$log(dist_{od})$	logarithmic distance of municipality centroids	ADM2 boundaries from GADM (2015)
$\%Cropland_{dt}$	percentage of area covered in cropland within ADM2 boundaries	European Space Agency (2019), GADM (2015)

$\%Urban_{dt}$	percentage of area covered in urban land within ADM2 boundaries	European Space Agency (2019), GADM (2015)
$Grid_{dt}$	1 if new transmission lines were constructed within the boundaries of municipality d in year t	GADM (2015), Rural Electrification Agency (2020)

Table A-2 – Effect on main lighting fuel

	(1)	(2)	(3)	(4)	(5)
	Baseline mean	-(Log) grid distance		Dummy grid	
		OLS	2SLS	OLS	2SLS
Collected firewood	0.088	0.002	0.005	-0.006	0.029
		(0.016)	(0.019)	(0.027)	(0.026)
Purchased firewood	0.038	0.006	0.015	0.008	0.007
		(0.009)	(0.011)	(0.015)	(0.013)
Kerosene	0.463	-0.051	-0.072*	-0.134	-0.310**
		(0.032)	(0.039)	(0.103)	(0.112)
Electricity	0.156	0.008	0.031	0.065	0.255***
		(0.028)	(0.028)	(0.078)	(0.083)
Generator	0.021	0.026	0.035	0.069	0.151
		(0.019)	(0.029)	(0.057)	(0.112)
Battery	0.202	-0.010	-0.029	-0.051	-0.171**
		(0.024)	(0.029)	(0.044)	(0.064)
Other	0.031	0.019**	0.014	0.049	0.039
		(0.009)	(0.012)	(0.031)	(0.035)
Observations		2,308	2,308	2,308	2,308
F-stat			331.727		55.568

Note:All regression control for wave-state fixed effects. Dependent variable is a binary measure of each fuel type. Each column reports results for a different measure of grid construction: column (2) reports results on negative log distance to actual grid location, column (3) on negative log grid distance instrumented by negative log distance to the hypothetical least cost grid, column (4) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid, and column (5) on the grid dummy instrumented by a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid. For all control variables distances are measured as negative logarithmic distance. Substations is the negative least cost distance to any substation in 2015, capital is the one to the state capital, road distance is the negative logarithmic distance to any primary or secondary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table A-3 – The effect of new transmission lines on migration reasons

	(1) Mean	(2) All	(3) Male adults	(4) Female adults	(5) Children
Divorce/separation	0.0460	-0.0636**	-0.0016	-0.1321**	0.0188
		(0.0294)	(0.0053)	(0.0562)	(0.0168)
Studies/education	0.1057	-0.0194	-0.0991	0.0491	0.0655
		(0.0733)	(0.0979)	(0.0793)	(0.2134)
For work	0.1802	0.1088	0.3005**	-0.0482	0.1181**
		(0.1111)	(0.1203)	(0.1335)	(0.0529)
To find better land	0.0487	0.0152	-0.0286	0.0480	-0.0514
		(0.0517)	(0.0878)	(0.0466)	(0.0530)
Health reasons	0.0060	-0.0209	-0.0080	-0.0256	-0.0475
		(0.0139)	(0.0098)	(0.0215)	(0.0329)
Security reasons	0.0066	-0.0213	0.0001	-0.0133	-0.1068
		(0.0249)	(0.0037)	(0.0169)	(0.0908)
Marriage/cohabitation	0.2618	-0.0074	-0.1107	0.1172	-0.1119
		(0.0600)	(0.0948)	(0.1160)	(0.1035)
To join family	0.1840	-0.0083	-0.0962	-0.0388	0.3322**
		(0.0804)	(0.0758)	(0.0990)	(0.1546)
Moved with family	0.0268	-0.0058	0.0157	-0.0232	-0.0195
		(0.0302)	(0.0509)	(0.0300)	(0.0367)
To set up home	0.0690	-0.0103	-0.0040	-0.0030	0.0103
		(0.0428)	(0.0700)	(0.0363)	(0.0095)
Unable to stay due to conflict	0.0044	0.0044	0.0136	0.0014	0.0014
		(0.0043)	(0.0112)	(0.0017)	(0.0017)
Dispute with other HH member	0.0027	0.0015	0.0059	-0.0035	-0.0035
		(0.0034)	(0.0049)	(0.0046)	(0.0046)
Other	0.0581	0.0273	0.0124	0.0720	-0.2079
		(0.0706)	(0.1099)	(0.0619)	(0.1555)
Missing values	0.0083	0.0049	-0.0027	0.0188	-0.0009
		(0.0046)	(0.0059)	(0.0164)	(0.0012)
Observations		26,486	6,367	7,595	12,524
F-stat		113.1559	64.2769	59.6364	265.5221

Note:All regression control for wave-state fixed effects and geographic controls. Geographic controls are the distance to any 2015 substation, the distance to the state capital, distance to the closest secondary or primary road in 2009, % of cropland, % of urban landcover and population density. The latter three are measured within a 40 km buffer in the year 2000. All distances are measured in negative logarithmic meters. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table A-4 – The effect of new transmission lines on main employment sector

	(1) Baseline	(2) All	(3) HH head	(4) HH spouse	(5) HH child	(6) HH grandchild
Agriculture	0.2511	-0.1043	-0.0911	-0.1130	-0.0762	-0.1102
		(0.0664)	(0.1313)	(0.1175)	(0.0538)	(0.0817)
Mining	0.0001	0.0005	0.0028	0.0000	0.0001	
		(0.0006)	(0.0036)	(0.0001)	(0.0001)	
Manufacturing	0.0221	0.0018	-0.0069	0.0362	0.0024	-0.0022
		(0.0089)	(0.0179)	(0.0348)	(0.0080)	(0.0045)
Technical Activities	0.0024	0.0015	0.0049	0.0040	-0.0009	0.0010
		(0.0019)	(0.0071)	(0.0037)	(0.0014)	(0.0018)
Electricity/Water/Gas/Waste	0.0007	0.0012	0.0024		0.0006	
		(0.0012)	(0.0029)		(0.0008)	
Construction	0.0066	-0.0058	-0.0287		0.0032	
		(0.0047)	(0.0287)		(0.0023)	
Transportation	0.0082	0.0090*	0.0281*	-0.0004	0.0056	
		(0.0045)	(0.0155)	(0.0007)	(0.0051)	
Buying and Selling	0.0595	0.0361*	0.1108*	0.2080***	-0.0148	-0.0090
		(0.0192)	(0.0585)	(0.0650)	(0.0106)	(0.0106)
Finance/Insurance/Real estate	0.0002	-0.0034	-0.0213	0.0026	-0.0001	
		(0.0029)	(0.0160)	(0.0033)	(0.0002)	
Personal Service	0.0224	0.0016	0.0568**	0.0009	-0.0100	-0.0927*
		(0.0078)	(0.0272)	(0.0095)	(0.0117)	(0.0537)
Education	0.0069	-0.0024	-0.0048	-0.0150	0.0008	
		(0.0045)	(0.0154)	(0.0294)	(0.0013)	
Health	0.0035	0.0023	0.0052	0.0068	0.0008	
		(0.0035)	(0.0101)	(0.0112)	(0.0009)	
Public Administration	0.0065	0.0025	0.0271	0.0009	-0.0028	-0.0020
		(0.0045)	(0.0180)	(0.0043)	(0.0040)	(0.0021)
Other	0.0035	0.0036	-0.0059	0.0056	0.0055	
		(0.0055)	(0.0195)	(0.0086)	(0.0044)	
Observations		15,993	2,716	2,536	9,338	564
F-stat		100.7010	68.2600	96.4335	102.0181	164.6119

Note: All regression show results on a on a grid dummy instrumented by a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid and control for wave-state fixed effects and distance to the closest substation, the negative logarithmic distance to the state capital, the negative logarithmic distance to the closest primary or secondary road in 2009, % Cropland, % urban landcover and population density measured within a 40km buffer. Columns (1) reports mean at baseline. Column (2) reports results for all individuals in the sample, column (3) uses the sample of household heads, column (4) uses the sample of household spouses, column (5) uses the sample of children of the household head, and column (6) uses the sample of grandchildren of the household head. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table A-5 – Balancing between control und treatment municipalities

	Control	SD	treatment	SE	N
Primary road density	0.273	0.558	-0.001	0.162	316
(Log) population	11.985	0.453	0.162	0.132	322
% of cropland	0.364	0.306	0.035	0.086	322
% of urban land	0.049	0.170	0.045	0.049	322

Note: Balancing is tested using a regression with state fixed effects and standard errors clustered at village-level as in the main regressions. (*** $p<0.01$, ** $p<0.05$, * $p<0.1$)

A Appendix B - Robustness tables

Table B-1 – Effect of new transmission lines on migration controlled for baseline differences

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
# of household members	5.963	-0.308**	-0.327**	-0.711***	-0.799***	2,310	58.444
		(0.147)	(0.159)	(0.201)	(0.230)		
# of elderly	0.071	-0.048	-0.048	-0.030	-0.036	2,310	58.444
		(0.031)	(0.032)	(0.033)	(0.036)		
# of children (total)	3.259	-0.279**	-0.310***	-0.512***	-0.586***	2,310	58.444
		(0.105)	(0.103)	(0.167)	(0.159)		
# of children (age 0-5)	1.176	-0.201**	-0.218**	-0.160	-0.247**	2,247	58.169
		(0.093)	(0.096)	(0.098)	(0.095)		
# of children (age 6-12)	1.301	0.087	0.076	-0.002	0.016	2,247	58.169
		(0.074)	(0.073)	(0.107)	(0.106)		
# of children (age 13-18)	0.802	-0.141	-0.144	-0.362**	-0.358**	2,247	58.169
		(0.089)	(0.089)	(0.137)	(0.140)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. In addition, all regression control for baseline household control variables that were found to differ significantly between treatment and control households in table 1. These include household roof material, household wall material, main lighting fuel and number of elderly household members. Column (1) shows the sample mean for the year 2009. columns (2) and (3) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, and columns (4) and (5) on the grid dummy instrumented by a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. For all control variables distances are measured as negative logarithmic distance. Geographic control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital, the negative logarithmic distance to the closest secondary or primary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-2 – The effect of new transmission lines on migration (individual level) controlled for baseline covariates

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
All HH members	0.019	0.015	0.015	0.060**	0.062**	15,729	100.476
		(0.017)	(0.018)	(0.027)	(0.029)		
HH head	0.003	0.006**	0.007**	0.010*	0.012**	2,682	69.652
		(0.003)	(0.003)	(0.005)	(0.005)		
HH spouse	0.035	-0.020	-0.019	-0.049*	-0.052**	2,495	97.169
		(0.018)	(0.017)	(0.025)	(0.021)		
HH child	0.091	0.031	0.031	0.109**	0.112**	9,180	101.136
		(0.024)	(0.026)	(0.043)	(0.044)		
HH grandchild	0.159	0.122	0.230**	0.060	0.286**	551	177.046
		(0.096)	(0.096)	(0.089)	(0.128)		
Other	0.180	0.060	0.047	0.088	0.128	810	38.440
		(0.080)	(0.082)	(0.239)	(0.241)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. In addition, all regression control for baseline household control variables that were found to differ significantly between treatment and control households in table 1. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Geographic control variables of columns (3) and (5) include distance to the state capital, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-3 – Effect of new transmission lines on agricultural production controlled for baseline differences

		(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
Output	(Log) production value	10.121	0.194	0.095	0.555	0.558	1,868	39.552
			(0.551)	(0.536)	(0.939)	(0.940)		
Factor inputs	(Log) labor costs	1.436	0.161	0.146	1.117***	1.027**	1,877	43.519
			(0.448)	(0.448)	(0.411)	(0.385)		
	(Log) # of paid workers	0.600	-0.095	-0.093	0.050	-0.015	1,877	43.519
			(0.098)	(0.094)	(0.175)	(0.175)		
	# of plots	1.784	0.173	0.219	0.685*	0.778**	2,310	58.444
			(0.261)	(0.263)	(0.358)	(0.373)		
Profit	(Log) food consumption	4.011	0.086**	0.086**	0.269***	0.282***	2,239	58.431
			(0.042)	(0.041)	(0.074)	(0.085)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. In addition, all regression control for baseline household control variables that were found to differ significantly between treatment and control households in table 1. These include household roof material, household wall material, main lighting fuel and number of elderly household members. Column (1) shows the sample mean for the year 2009. columns (2) and (3) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, and columns (4) and (5) on the grid dummy instrumented by a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. For all control variables distances are measured as negative logarithmic distance. Geographic control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital, negative logarithmic distance to the closest secondary or primary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-4 – Fixed effects regression on household composition

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
# of household members	5.963	-0.276*	-0.313*	-0.756***	-0.963***	3,524	22.125
		(0.158)	(0.181)	(0.254)	(0.294)		
# of elderly	0.071	-0.029	-0.037	-0.069*	-0.096**	3,524	22.125
		(0.032)	(0.034)	(0.037)	(0.043)		
# of children (total)	3.259	-0.207*	-0.247*	-0.411***	-0.555***	3,524	22.125
		(0.107)	(0.127)	(0.158)	(0.190)		
# of children (age 0-5)	1.176	-0.207**	-0.240**	-0.175**	-0.281***	3,459	21.732
		(0.096)	(0.103)	(0.077)	(0.097)		
# of children (age 6-12)	1.301	0.070	0.038	0.002	-0.050	3,459	21.732
		(0.065)	(0.069)	(0.063)	(0.085)		
# of children (age 13-18)	0.802	-0.086	-0.068	-0.243*	-0.232	3,459	21.732
		(0.084)	(0.097)	(0.127)	(0.163)		

Note: All regression control for household fixed effects. In addition, all regressions control for wave-state fixed effects and the interaction of wave and distance to the closest substation in order to provide a similar specification to the first-difference estimates of table 3. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if future grid construction was planned within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the future grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost path for the future grid lines was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. In order to control for different trends by time-constant location characteristics, control variables of columns (3) and (5) include the interactions of wave and distance to the state capital, wave and distance to the closest 2009 road, wave and % cropland, wave and % urban land, and wave and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-5 – Fixed effects regression on migration (individual level)

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
All HH members	0.019	0.014	0.013	0.058*	0.059*	26,486	113.030
		(0.014)	(0.015)	(0.031)	(0.032)		
HH head	0.003	0.001	0.001	0.001	0.002	4,173	62.096
		(0.001)	(0.001)	(0.001)	(0.002)		
HH spouse	0.035	-0.015	-0.016	-0.052***	-0.054***	4,104	112.771
		(0.015)	(0.015)	(0.018)	(0.018)		
HH child	0.091	0.035*	0.034	0.113**	0.114**	15,661	124.926
		(0.021)	(0.022)	(0.046)	(0.048)		
HH grandchild	0.159	-0.018	0.002	-0.036	0.004	1,083	98.915
		(0.072)	(0.076)	(0.103)	(0.121)		
Other	0.180	0.006	-0.001	0.060	0.054	1,454	57.224
		(0.053)	(0.054)	(0.179)	(0.179)		

Note: All regression control for individual level fixed effects. In addition, all regressions control for wave-state fixed effects and the interaction of wave and distance to the closest substation in order to provide a similar specification to the first-difference estimates of table 4. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. In order to control for different trends by time-constant location characteristics, control variables of columns (3) and (5) include the interactions of wave and distance to the state capital, distance to the closest 2009 road, wave and % cropland, wave and % urban land, and wave and population density within a 40km buffer in year 2000. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-6 – Fixed effects regression on agricultural production

		(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
Output	(Log) production value	10.121	-0.453	-0.241	-0.901	-0.685	3,029	19.136
			(0.715)	(0.509)	(1.350)	(1.080)		
Factor inputs	(Log) labor costs	1.436	-0.055	-0.065	0.307	0.262	3,037	20.801
			(0.296)	(0.327)	(0.340)	(0.374)		
	(Log) # of paid workers	0.600	0.052	-0.036	-0.014	-0.180	3,037	20.801
			(0.084)	(0.107)	(0.120)	(0.167)		
	# of plots	1.784	0.048	-0.088	0.315	0.083	3,524	22.125
			(0.240)	(0.193)	(0.330)	(0.266)		
Profit	(Log) food consumption	4.011	0.063*	0.052	0.202***	0.225**	3,451	22.421
			(0.037)	(0.041)	(0.070)	(0.090)		

Note: All regression control for household fixed effects. In addition, all regressions control for wave-state fixed effects and the interaction of wave and distance to the closest substation in order to provide a similar specification to the first-difference estimates of table 3. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if future grid construction was planned within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the future grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost path for the future grid lines was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. In order to control for different trends by time-constant location characteristics, control variables of columns (3) and (5) include the interactions of wave and distance to the state capital, distance to the closest 2009 road, wave and % cropland, wave and % urban land, and wave and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-7 – Placebo test of future transmission lines on migration

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS	(5) 2SLS	(6) Obs	(7) F-stat
# of household members	5.963	-0.144	-0.122	-0.206	-0.181	2,323	113.196
		(0.107)	(0.102)	(0.145)	(0.142)		
# of elderly	0.071	-0.011	-0.008	-0.016	-0.014	2,323	113.196
		(0.021)	(0.020)	(0.026)	(0.025)		
# of children (total)	3.259	-0.063	-0.039	-0.099	-0.065	2,323	113.196
		(0.095)	(0.090)	(0.114)	(0.109)		
# of children (age 0-5)	1.176	-0.116*	-0.090	-0.139	-0.113	2,259	106.210
		(0.068)	(0.062)	(0.082)	(0.076)		
# of children (age 6-12)	1.301	0.100	0.105	0.083	0.100	2,259	106.210
		(0.080)	(0.078)	(0.101)	(0.101)		
# of children (age 13-18)	0.802	-0.046	-0.054	-0.006	-0.014	2,259	106.210
		(0.077)	(0.079)	(0.094)	(0.097)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if future grid construction was planned within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the future grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost path for the future grid lines was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Control variables of columns (3) and (5) include distance to the state capital, distance to the closest primary or secondary road in 2009, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-8 – Placebo test of future transmission lines on migration (individual level)

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
All HH members	0.019	0.004	0.004	0.011	0.005	15,993	110.885
		(0.015)	(0.015)	(0.021)	(0.021)		
HH head	0.003	0.014**	0.013**	0.011*	0.010*	2,716	192.178
		(0.006)	(0.006)	(0.006)	(0.006)		
HH spouse	0.035	0.000	0.001	-0.007	-0.016	2,536	72.799
		(0.014)	(0.014)	(0.021)	(0.019)		
HH child	0.091	-0.002	-0.002	0.017	0.009	9,338	100.710
		(0.022)	(0.021)	(0.030)	(0.030)		
HH grandchild	0.159	-0.037	-0.001	-0.011	0.014	564	81.082
		(0.086)	(0.084)	(0.118)	(0.108)		
Other	0.180	0.128*	0.140*	0.125	0.126	828	52.410
		(0.074)	(0.076)	(0.109)	(0.104)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if future grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost path for future grid lines was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Control variables of columns (3) and (5) include distance to the state capital, distance to the closest secondary or primary 2009 road, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-9 – Placebo test of future transmission lines on agricultural production

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
(Log) production value	10.121	-0.534*	-0.532	-0.360	-0.376	1,876	72.340
		(0.322)	(0.327)	(0.438)	(0.442)		
(Log) labor costs	1.436	0.160	0.062	0.216	0.067	1,885	74.849
		(0.374)	(0.386)	(0.455)	(0.489)		
(Log) # of paid workers	0.600	0.026	-0.003	-0.006	-0.032	1,885	74.849
		(0.096)	(0.102)	(0.138)	(0.140)		
# of plots	1.784	-0.191	-0.186	-0.012	0.001	2,323	113.196
		(0.196)	(0.201)	(0.156)	(0.166)		
(Log) food consumption	4.011	0.067***	0.063**	0.112***	0.106***	2,250	117.100
		(0.026)	(0.025)	(0.036)	(0.036)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean for the year 2009. columns (2) and (3) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls without and with geographic controls. For all control variables distances are measured as negative logarithmic distance. Control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital, negative logarithmic distance to the closest secondary or primary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-10 – The effect of new transmission lines on migration controlling for new roads

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls
Panel A: New Grid Construction					
# of household members	5.963	-0.330**	-0.349**	-0.693***	-0.778***
		(0.140)	(0.150)	(0.197)	(0.227)
# of elderly	0.071	-0.061*	-0.061*	-0.038	-0.044
		(0.033)	(0.035)	(0.032)	(0.037)
# of children (total)	3.259	-0.300***	-0.325***	-0.506***	-0.582***
		(0.102)	(0.100)	(0.148)	(0.141)
# of children (age 0-5)	1.176	-0.206**	-0.223**	-0.174*	-0.255**
		(0.093)	(0.097)	(0.096)	(0.102)
# of children (age 6-12)	1.301	0.063	0.056	-0.020	-0.006
		(0.071)	(0.071)	(0.102)	(0.102)
# of children (age 13-18)	0.802	-0.136	-0.136	-0.335	-0.337
		(0.089)	(0.090)	(0.133)	(0.135)
Panel B: New Road Construction					
# of household members	5.963	0.148	0.146	0.144	0.127
		(0.163)	(0.173)	(0.162)	(0.171)
# of elderly	0.071	0.027	0.029	0.027	0.030
		(0.031)	(0.032)	(0.031)	(0.031)
# of children (total)	3.259	0.023	0.015	0.021	0.003
		(0.162)	(0.168)	(0.160)	(0.165)
# of children (age 0-5)	1.176	0.030	0.044	0.030	0.043
		(0.129)	(0.130)	(0.127)	(0.129)
# of children (age 6-12)	1.301	-0.083	-0.102	-0.084	-0.105
		(0.090)	(0.090)	(0.089)	(0.089)
# of children (age 13-18)	0.802	0.077	0.075	0.075	0.066
		(0.181)	(0.181)	(0.178)	(0.179)
Observations		2,259	2,259	2,259	2,259
F-stat				66.461	55.961

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean for the year 2009. columns (2) and (3) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, and columns (4) and (5) on a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. For all control variables distances are measured as negative logarithmic distance. Control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-11 – The effect of new transmission lines on migration (individual level)

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
			Panel A: New Grid Construction				
All HH members	0.019	0.015	0.014	0.053*	0.056*	15,993	100.312
		(0.016)	(0.017)	(0.027)	(0.028)		
HH head	0.003	0.005	0.005	0.011	0.011	2,716	68.572
		(0.003)	(0.003)	(0.007)	(0.007)		
HH spouse	0.035	-0.019	-0.019	-0.057**	-0.059***	2,536	94.928
		(0.017)	(0.016)	(0.024)	(0.021)		
HH child	0.091	0.029	0.029	0.098**	0.102**	9,338	101.025
		(0.023)	(0.024)	(0.042)	(0.043)		
HH grandchild	0.159	0.100	0.184**	0.020	0.198*	564	168.629
		(0.089)	(0.085)	(0.074)	(0.106)		
Other	0.180	0.058	0.045	0.130	0.164	828	44.469
		(0.082)	(0.083)	(0.234)	(0.239)		
			Panel B: New Road Construction				
All HH members	0.019	-0.045**	-0.044**	-0.043**	-0.041**	15,993	100.312
		(0.020)	(0.020)	(0.019)	(0.019)		
HH head	0.003	-0.024	-0.024	-0.024	-0.024	2,716	68.572
		(0.020)	(0.020)	(0.019)	(0.019)		
HH spouse	0.035	-0.035	-0.017	-0.038	-0.019	2,536	94.928
		(0.032)	(0.014)	(0.032)	(0.014)		
HH child	0.091	-0.044*	-0.044*	-0.040	-0.039	9,338	101.025
		(0.025)	(0.027)	(0.025)	(0.026)		
HH grandchild	0.159	-0.194	-0.150	-0.196	-0.150	564	168.629
		(0.120)	(0.125)	(0.118)	(0.120)		
Other	0.180	0.062	0.063	0.063	0.065	828	44.469
		(0.243)	(0.255)	(0.241)	(0.254)		

Note: Panel A and Panel B show results from the same regressions, whereas Panel A reports point estimates for the dummy indicating new transmission grid construction, Panel B reports point estimates for the dummy indicating new road construction. All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Control variables of columns (3) and (5) include distance to the state capital, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-12 – The effect of new transmission lines on agricultural production

	(1) Baseline mean	(2) Dummy grid no controls	(3) Dummy grid controls	(4) Dummy least cost grid no controls	(5) Dummy least cost grid controls	(6) obs	(7) F-stat
			Panel A: New Grid Construction				
(Log) production value	10.121	0.206	0.108	0.544	0.562	1,876	37.105
		(0.519)	(0.512)	(0.911)	(0.916)		
(Log) labor costs	1.436	0.047	0.003	0.935**	0.866**	1,885	41.024
		(0.458)	(0.465)	(0.385)	(0.358)		
(Log) # of paid workers	0.600	-0.092	-0.093	0.065	-0.003	1,885	41.024
		(0.094)	(0.087)	(0.149)	(0.151)		
# of plots	1.784	0.165	0.208	0.682*	0.769**	2,323	56.279
		(0.258)	(0.263)	(0.339)	(0.356)		
(Log) food consumption	4.011	0.081**	0.080**	0.258***	0.273***	2,250	55.914
		(0.039)	(0.038)	(0.074)	(0.085)		
			Panel B: New Road Construction				
(Log) production value	10.121	-0.773**	-0.971**	-0.776*	-0.966**	1,876	37.105
		(0.384)	(0.427)	(0.382)	(0.424)		
(Log) labor costs	1.436	0.281	0.232	0.274	0.241	1,885	41.024
		(0.455)	(0.490)	(0.453)	(0.488)		
(Log) # of paid workers	0.600	0.048	0.084	0.047	0.085	1,885	41.024
		(0.305)	(0.317)	(0.302)	(0.314)		
# of plots	1.784	-0.260	-0.223	-0.254	-0.198	2,323	56.279
		(0.220)	(0.224)	(0.218)	(0.222)		
(Log) food consumption	4.011	-0.114*	-0.117*	-0.112*	-0.108*	2,250	55.914
		(0.064)	(0.063)	(0.064)	(0.063)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. The table replicates table 6 while controlling for primary and secondary road construction. Panel A reports results on the grid dummy for comparison with table 6, Panel B reports results on the road dummy from the same regression as Panel A. Column (1) shows the sample mean for the year 2009. Columns (2) and (3) use regressions on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, and columns (4) and (5) on a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. New road construction is also defined as a binary variable that turns 1 if a new primary or secondary road was constructed within 15km distance. For all control variables distances are measured as negative logarithmic distance. Control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-13 – The effect of new transmission lines on media device ownership

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) IV no controls	(5) IV controls	(6) Obs	(7) F-stat
Radio	0.553	-0.101	-0.092	-0.166	-0.161	2,300	55.424
		(0.067)	(0.065)	(0.114)	(0.105)		
TV set	0.205	0.031	0.038	0.191**	0.210**	2,300	55.424
		(0.051)	(0.053)	(0.077)	(0.089)		
Computer	0.009	0.010	0.010	-0.002	-0.002	2,300	55.424
		(0.010)	(0.010)	(0.016)	(0.017)		
Internet	0.000	-0.028	-0.027	-0.085*	-0.084*	2,312	55.693
		(0.024)	(0.024)	(0.044)	(0.045)		
Mobil	0.275	-0.021	-0.019	0.042	0.037	12,019	115.238
		(0.024)	(0.023)	(0.031)	(0.031)		

Note: All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean for the year 2009. columns (2) and (3) on a dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, and columns (4) and (5) on a dummy variable turning 1 if the hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. For all control variables distances are measured as negative logarithmic distance. Control variables include the negative logarithmic distance to any substation in 2015, the negative logarithmic distance to the state capital, the negative logarithmic distance to any primary or secondary road in 2009. % Cropland, % urban landcover and population density are measured within a 40km buffer. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-14 – The effect of new transmission lines on migration controlling for 3G mobile network

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls
Panel A: New Grid Construction					
# of household members	5.963	-0.324**	-0.347**	-0.666***	-0.755***
		(0.143)	(0.152)	(0.216)	(0.243)
# of elderly	0.071	-0.061*	-0.062*	-0.042	-0.047
		(0.033)	(0.035)	(0.032)	(0.037)
# of children (total)	3.259	-0.292***	-0.321***	-0.469***	-0.550***
		(0.101)	(0.100)	(0.157)	(0.149)
# of children (age 0-5)	1.176	-0.212**	-0.226**	-0.193**	-0.272**
		(0.092)	(0.096)	(0.090)	(0.099)
# of children (age 6-12)	1.301	0.074	0.063	0.019	0.027
		(0.070)	(0.071)	(0.101)	(0.105)
# of children (age 13-18)	0.802	-0.132	-0.134	-0.318	-0.321
		(0.088)	(0.089)	(0.136)	(0.139)
Panel B: New 3G Mobile Network Coverage					
# of household members	5.963	-0.208	-0.206	-0.195	-0.199
		(0.176)	(0.177)	(0.182)	(0.183)
# of elderly	0.071	0.027	0.029	0.027	0.028
		(0.021)	(0.021)	(0.021)	(0.021)
# of children (total)	3.259	-0.292**	-0.293**	-0.285**	-0.289**
		(0.119)	(0.122)	(0.121)	(0.123)
# of children (age 0-5)	1.176	0.139**	0.142**	0.138**	0.143**
		(0.065)	(0.065)	(0.065)	(0.065)
# of children (age 6-12)	1.301	-0.282***	-0.280***	-0.279**	-0.279***
		(0.101)	(0.100)	(0.102)	(0.100)
# of children (age 13-18)	0.802	-0.130	-0.136	-0.121	-0.131
		(0.112)	(0.113)	(0.111)	(0.111)
Observations		2,259	2,259	2,259	2,259
F-stat				62.647	53.359

Note: Panel A and Panel B show results from the same regressions, whereas Panel A reports point estimates for the dummy indicating new transmission grid construction, Panel B reports point estimates for the dummy indicating 3G mobile network coverage. All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Control variables of columns (3) and (5) include distance to the state capital, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-15 – The effect of new transmission lines on migration (individual level) controlling for 3G mobile network

	(1) Baseline mean	(2) OLS no controls	(3) OLS controls	(4) 2SLS no controls	(5) 2SLS controls	(6) Obs	(7) F-stat
Panel A: New Grid Construction							
All HH members	0.019	0.016	0.016	0.058**	0.058**	15,993	96.735
		(0.017)	(0.017)	(0.026)	(0.028)		
HH head	0.003	0.005	0.005	0.012*	0.012*	2,716	66.511
		(0.003)	(0.003)	(0.007)	(0.007)		
HH spouse	0.035	-0.014	-0.016	-0.044**	-0.048**	2,536	90.614
		(0.016)	(0.016)	(0.020)	(0.019)		
HH child	0.091	0.030	0.029	0.100**	0.101**	9,338	96.849
		(0.023)	(0.025)	(0.043)	(0.044)		
HH grandchild	0.159	0.118	0.191**	0.049	0.210**	564	168.798
		(0.088)	(0.084)	(0.075)	(0.103)		
Other	0.180	0.057	0.041	0.142	0.175	828	43.174
		(0.083)	(0.084)	(0.232)	(0.236)		
Panel B: New 3G Mobile Network Coverage							
All HH members	0.019	-0.013	-0.006	-0.016	-0.008	15,993	96.735
		(0.019)	(0.019)	(0.020)	(0.020)		
HH head	0.003	-0.001	-0.000	-0.001	-0.000	2,716	66.511
		(0.010)	(0.011)	(0.010)	(0.011)		
HH spouse	0.035	-0.052***	-0.047**	-0.050**	-0.045**	2,536	90.614
		(0.019)	(0.019)	(0.018)	(0.019)		
HH child	0.091	0.004	0.012	-0.002	0.007	9,338	96.849
		(0.030)	(0.031)	(0.031)	(0.033)		
HH grandchild	0.159	-0.119	-0.058	-0.114	-0.059	564	168.798
		(0.090)	(0.095)	(0.087)	(0.091)		
Other	0.180	-0.072	-0.080	-0.071	-0.076	828	43.174
		(0.082)	(0.081)	(0.081)	(0.081)		

Note: Panel A and Panel B show results from the same regressions, whereas Panel A reports point estimates for the dummy indicating new transmission grid construction, Panel B reports point estimates for the dummy indicating 3G mobile network coverage. All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Control variables of columns (3) and (5) include distance to the state capital, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

Table B-16 – The effect of new transmission lines on agricultural production controlling for 3G mobile network

	(1) Baseline mean	(2) Dummy grid no controls	(3) Dummy grid controls	(4) Dummy least cost grid no controls	(5) Dummy least cost grid controls	(6) obs	(7) F-stat
Panel A: New Grid Construction							
(Log) production value	10.121	0.185	0.098	0.508	0.522	1,876.000	36.837
		(0.520)	(0.509)	(0.911)	(0.912)		
(Log) labor costs	1.436	0.046	0.001	0.939**	0.871**	1,885.000	40.508
		(0.461)	(0.468)	(0.388)	(0.362)		
(Log) # of paid workers	0.600	-0.092	-0.095	0.065	-0.002	1,885.000	40.508
		(0.094)	(0.087)	(0.149)	(0.151)		
# of plots	1.784	0.177	0.216	0.734**	0.811**	2,323.000	53.876
		(0.256)	(0.263)	(0.299)	(0.334)		
(Log) food consumption	4.011	0.078**	0.080**	0.245***	0.263***	2,250.000	53.314
		(0.037)	(0.037)	(0.073)	(0.085)		
Panel B: New 3G Mobile Network Coverage							
(Log) production value	10.121	-0.401	-0.322	-0.378	-0.293	1,876.000	36.837
		(0.411)	(0.443)	(0.408)	(0.434)		
(Log) labor costs	1.436	-0.039	-0.041	0.005	-0.001	1,885.000	40.508
		(0.074)	(0.093)	(0.109)	(0.119)		
(Log) # of paid workers	0.600	-0.023	-0.048	-0.015	-0.044	1,885.000	40.508
		(0.157)	(0.152)	(0.154)	(0.150)		
# of plots	1.784	-0.397***	-0.385***	-0.419**	-0.395**	2,323.000	53.876
		(0.138)	(0.135)	(0.150)	(0.142)		
(Log) food consumption	4.011	0.091*	0.090*	0.083	0.085	2,250.000	53.314
		(0.051)	(0.052)	(0.048)	(0.051)		

Note:Panel A and Panel B show results from the same regressions, whereas Panel A reports point estimates for the dummy indicating new transmission grid construction, Panel B reports point estimates for the dummy indicating 3G mobile network coverage. All regression control for wave-state fixed effects and distance to the closest substation. Column (1) shows the sample mean. Columns (2) and (3) present regression results using dummy variable turning 1 if actual grid construction was within 15km distance and closer than any existing grid without and with geographic controls, columns (4) and (5) present results using the grid dummy instrumented by a dummy variable turning 1 if hypothetical least cost grid was within 15km distance and closer than any existing grid without and with geographic controls. Column (6) displays number of observations of column (5) results and column (7) display Kleibergen-Papp-F-statistics of column (5) results. Control variables of columns (3) and (5) include distance to the state capital, % cropland, % urban land and population density within a 40km buffer in year 2000. All distances are always measured as negative logarithmic distance. Standard errors are clustered at the level of the survey enumeration area and stated in parentheses below point estimates. *** 1%, ** 5%, and * 10 % significance levels.

B Appendix C - Least cost approach

My approach draws heavily on (Faber, 2014). To construct grid that assigns construction costs to each pixel I use data on elevation from 90-meter Shuttle Radar Topography Mission (SRTM) Global Digital Elevation Model (Farr and Kobrick, 2000) and gridded data on landcover categories from the Climate Change Initiative Land Cover Maps (CCI-LC) by the European Space Agency (European Space Agency, 2019). After converting elevation data into a gridded raster of terrain slope measured in degrees, I aggregate both datasets to a resolution of 900m x 900m for computational feasibility. Then, I generate a conductance raster that assigns construction costs to each grid cell, based on the following equation:

$$c_i = 1 + slope_i + 25 \times wetland_i + 25 \times urban_i + 25 \times water_i \tag{B-1}$$

whre c_i represents the construction costs at grid cell i, $slope_i$ is the average land gradient at i ranging from 0 to approximately 32 degrees. The terms $wetland_i$, $urban_i$ and $water_i$ are dummy variables that mark the respectively land cover categories that make infrastructure construction extremely costly. Finally, I determine the least cost path for each pair of substations based on this conductance raster.

C Appendix D - Theoretical model

In this section we describe an extension for the static migration model from Bryan and Morten (2018) that distinguishes between urban and rural locations drawing on two-sector Rosen-Roback models (Roback, 1982). While the baseline model predicts a increase in population following a productivity shock, I show how the presence of credit constrains can alter this outcome. 7.

C.1 Baseline model

The economy consists of N locations. Each location is the origin o of a number of L_o workers and can belong either to the rural sector, $r_o = 1$, or the urban sector, $r_o = 0$. Overall, there are N^r rural locations and N^u urban locations, with $N^r + N^u = N$. Following Bryan and Morten (2018), skill is location specific. For every possible destination location d workers i draw a skill level s_{id} from a Fréchet distribution (and respectively s_{io} for the origin location), such that

$$F(s_1, \ldots, s_N) = exp\left(-\left\{\sum_{d=1}^{N} s_d^{-[\tilde{\theta}/(1-\rho)]}\right\}^{1-\rho}\right) \tag{B-2}$$

where larger values of $\tilde{\theta}$ are associated with more evenly distributed skill across locations and larger values of ρ are associated with a higher correlation of skill across locations. For simplicity, I will use $\theta = \tilde{\theta}/(1-\rho)$ for the remaining of the text. Following again the original model by Bryan and Morten (2018), innate skills are multiplied with the schooling quality at origin to form human capital.

$$h_{ido} = s_{id}q_o \tag{B-3}$$

Wage of worker i from origin o working and living at destination d then determined by:

$$wage_{ido} = w_d h_{ido} = w_d s_{id} q_o \tag{B-4}$$

where w_d can be thought of as the wage per effective unit of labor in destination d or the productivity of location d. Bryan and Morten (2018) also include an error term in the wage equation to account for any factor that causes workers from origin o to increase their labor demand at a certain destination d. Since this is exactly the type of variation in labor demand that I am interested in, I omit the inclusion of the error term here. The term w_d is determined by the price level at destination d, p_d, and a technology term A_d, such that:

$$w_d = p_d A_d \tag{B-5}$$

The indirect utility function of a worker i staying in her origin location depends on the amenities α_o this location offers and the consumption level determined by received wage, such that:

$$U_{ioo} = \alpha_o w_o h_{ido} \tag{B-6}$$

Again the original model of Bryan and Morten (2018) includes an error term that describes random variation in amenities at d that depends on origin o. As I am only interested in qualitative predictions and for the sake of simplicity this error term is again omitted. Moving to another location is costly and must be compensated by a higher income. So indirect utility for a worker i from origin o living and working in destination d becomes:

$$U_{ido} = \alpha_d w_d h_{ido}(1 - \tau_{do}) \tag{B-7}$$

where $\tau \ni [0, 1]$ is defined as the movement costs. The proportion of persons i from origin o that decide to migrate to destination d is given by:

$$m_{od} = \frac{\tilde{w}_{do}}{\tilde{w}_{do} + \tilde{w}_o} \tag{B-8}$$

with $\tilde{w}_{do} = \alpha_d w_d h_{ido}(1 - \tau_{do})$ and $\tilde{w}_o = \alpha_o w_o$. Here $\tilde{w}_{do}$ measures here the attractiveness of location d for someone from o. This is the main sorting equation. In contrast, to Bryan and Morten (2018) my sorting equation includes human capital. This is essential, since one of my model extensions focuses how a technology induced change in human capital affects migration. In my empirical analysis of section 5 we only observe total out-migration from origin o. This can be written as:

$$M_o = \sum_{d=1}^{N} m_{od} = \sum_{d=1}^{N} \frac{\tilde{w}_{do}}{\tilde{w}_{do} + \tilde{w}_o} \tag{B-9}$$

C.2 Productivity shock

Denote e as a measure of local electricity access. We assume technology of production A_o is a positive function of local electricity access, such that $A_o(e)' = t$, with $t \ni [0, 1]$. This implies that an increases in e increase local productivity:

$$\frac{\partial \tilde{w}_{oo}}{\partial e} = \alpha_o p_o t h_{ioo} \geq 0 \tag{B-10}$$

when productivity rises, wages at origin o increase and in turn the attractiveness of destination o for workers from all locations increases. As a net effect, we expect to see falling migration flows from origin o to all other destinations d:

$$\frac{\partial M_o}{\partial e} = -\alpha_o p_o th_{ioo} \sum_{d=1}^{N} \frac{\tilde{w}_{do}}{(\tilde{w}_{do} + \tilde{w}_{oo})^2} \leq 0 \tag{B-11}$$

This simple prediction is in line with previous works. Lewis and Severnini (2020) use a Rosen-Roback style model to show that productivity boosts from rural electrification should lead to an increase in population locally. Bryan and Morten (2018) assume an exogenous technology term, but include a normally distributed error term to allow for any unmeasured characteristics that increase productivity which in turn should increase migration to this location.

C.3 Credit constrains

Credit constrains are a common market failure in developing countries. In the context of migration, scholars have shown that a lack of credit is an important barrier to optimal migration, since it prevents households from being able to pay for movement costs upfront even if the expected return from migration is positive (Bryan et al., 2014). We remain in a static setting and model credit constraints such that movement costs cannot exceed wage at origin, i.e. $\tau_{do} \leq wage_{ioo}$. One can think of the restriction as a simplification of a dynamic setting, where movement costs have to be paid by earnings and saving of the previous period. Our dyadic mobility measure then becomes:

$$m_{od} = \begin{cases} \frac{\tilde{w}_{do}}{\tilde{w}_{do} + \tilde{w}_{oo}}, & \text{if } \tau_{do} \leq wage_{ioo} \\ 0, & \text{otherwise} \end{cases} \tag{B-12}$$

This restrictions reduces the aggregate measure of out-migration from origin o to:

$$M_o = \sum_{d=1}^{N} \frac{\tilde{w}_{do}}{\tilde{w}_{do} + \tilde{w}_{oo}} \times 1_{\tau_{do} \leq wage_{ioo}} \tag{B-13}$$

where $1_{\tau_{do} \leq wage_{ioo}}$ is an indicator function turning 1 for. This means when credit constrains are present the migration flows from origin o are smaller than at the optimal. Now, if origin o receives an increase in electricity access, productivity and wages at o increase. This reduces the set of origin-destination pairs for which $\tau_{do} > wage_{ioo}$. At the same time, rising wages increase $\tilde{w}_{oo}$, i.e. the attractiveness of location o. Both effects have opposing implications for aggregate out-migration M_o. The net effect depends on the number of origin-destination pairs for which $\tau_{do} \geq wage_{ioo}$ is satisfied before and after the productivity shock. Assume without electricity at origin o there are m destinations where wage at origin is smaller than movement costs, with electricity there are only n destinations for which this is the case, with $n < m < N$. Given these assumptions, an increase in electricity access at origin o increases the out-migration if:

$$\left| -\alpha_o p_o th_{ioo} \sum_{d=1}^{N-m} \frac{\tilde{w}_{do}}{(\tilde{w}_{do} + \tilde{w}_{oo})^2} \right| < \sum_{d=N-m}^{N-m+n} \frac{\tilde{w}_{do}}{\tilde{w}_{do} + \tilde{w}_{oo}} \tag{B-14}$$

Note that the productivity shock only affects the attractiveness of the origin $\tilde{w}_{oo}$ and the set of destinations that workers from o can afford to migrate to. It does not affect the relative attractiveness of potential destinations $\tilde{w}_{do}$. This means for any pair of destinations $d_1, d_2 \ni N$ if destination d_1 is preferred over destination d_2 before technology shock at origin o, this is still the case after the technology shock:

$$U(\tilde{w}_{d_1o}|e = 0) > U(\tilde{w}_{d_2o}|e = 0) \iff U(\tilde{w}_{d_1o}|e = 1) > U(\tilde{w}_{d_2o}|e = 1) \tag{B-15}$$