Pivot to Power: Resolving a Local Public Good Problem*

Torsten Figueiredo Walter   Niclas Moneke**

March 30, 2024

Abstract

Economists have long studied the underprovision of public goods. Theoretical mechanisms designed to achieve efficient provision abound. However, empirical evidence how underprovision can be overcome is rare. In this paper, we characterise the provision of last-mile infrastructure as a local public good problem. We study the recent, country-wide public roll-out of the electric grid in Zambia which connected hundreds of rural locations, but not individual agents (the 'last mile'). We show that, conditional on grid arrival, the last mile electric trunk network required to connect agents appears only in some locations, but not others – leading to a bimodal distribution of adoption. We provide causal evidence that whenever agents with pivotal expected benefits of electricity (here: grain mills) are present at baseline, locations see dramatically higher electricity adoption among end-users. Consequently, any beneficial effects of electrification on various measures of development are confined to the few locations with pivotal agents at baseline.

JEL classification: J24, O13, O14, O18, Q41, R1

*We thank Ben Faber, Paco Buera, Gharad Bryan, Taryn Dinkelman, Kevin Donovan, Doug Gollin, Karen Macours, Craig McIntosh, Pritchard Mukuka, Simon Quinn, Lupupa Siluapwa, Daniel Sturm, Anant Sudarshan, Chris Woodruff and audiences at PSE, Oxford, Linkages in Labour Economics Workshop, LSE Environment Week, UEA Washington DC, Y-Rise Kingston, Hitotsubashi, NUS, SSE, EEA Barcelona and STEG Annual Conference for helpful comments. Gregory Cuff, Felix Johnson, Patrick Mathewson, Hassane Meite, Ana-Maria Radu, Mackenzie Tan Phei Huey and Likun Tian provided outstanding research assistance. Financial support from the IGC is gratefully acknowledged.

**Figueiredo Walter: t.f.walter@nyu.edu, https://sites.google.com/site/torstenfigueiredowalter/, Social Science Division, New York University Abu Dhabi. Moneke: niclas.moneke@economics.ox.ac.uk, https://niclasmoneke.com, Department of Economics, University of Oxford.

1 Introduction

Economists have long studied the underprovision of public goods. Free-riding and coordination failure can lead to underprovision in equilibrium.[1] Theoretical mechanisms designed to achieve efficient provision abound. In particular, a large literature discusses the determinants of the degree of underprovision.[2] However, empirical evidence on how the underprovision of local public goods can be overcome is rare.[3]

This paper provides new evidence on how local public good problems of last-mile infrastructure can be resolved. To answer what it takes to overcome the underprovision of local public goods, we proceed in four steps. First, we note that last-mile infrastructure allocations often display public good characteristics. For example, local irrigation, sewerage and low-voltage electricity networks are often both imperfectly excludable and imperfectly rivalrous. Persistent problems in their provision are well-documented.[4] However, the absence of last-mile access greatly undermines the benefits and development effects of the entire infrastructure investment, which remains underutilised.

In our context, the provision of the local electric trunk network required for last-mile connections can be characterised as a local public good problem: once erected in the form of a set of poles, overhead lines and service drops, connections to the local trunk network are imperfectly excludable and non-payment for electricity is common, both in our setting of rural Zambia and elsewhere around the world.[5] Furthermore, usage of electricity consumed from the local network is imperfectly rivalrous since the electric trunk network is designed by expert engineers to meet peak local demand, while supply is global and unaffected by local variation in consumption from individual end-users.[6,7]

Second, we document a new empirical regularity across rural grid-connected locations: household electricity connections are bimodal conditional on prior grid arrival – with locations either experiencing near-complete electrification or none at all. This bi-

[1] Samuelson (1954) first documented underprovision of public (compared to private) goods in equilibrium, which Vickrey (1961) and Groves and Ledyard (1977) formalised more generally. Bergstrom et al. (1986) introduce their private provision. Cornes and Sandler (1996) introduce impure public goods.

[2] cf. Olson (1964), Cornes (1993), Wade (1994), Alesina and La Ferrara (2000), Bardhan (2000), Bardhan and Dayton-Johnson (2002a), Banerjee (2004), Banerjee et al. (2005), Bardhan et al. (2007)

[3] Despite many theoretical contributions being well-motivated by anecdotal evidence, systematic empirical evidence that causally documents underprovision and its resolution remains scarce.

[4] In fact, evidence on the eponymous 'last-mile problem' exists for electric trunk, irrigation and freshwater networks, and is not confined to low-income countries (e.g. fibre broadband, parcel delivery).

[5] Burgess et al. (2020a, 2020b) highlight non-payment in India, whereas Figure A1 and Table A1 provide evidence of widespread non-payment in Zambia, which is more severe in rural than urban areas.

[6] Figure A2 depicts an electric transformer as point of supply in excess of peak load (cf. Figure A3 for load curves). Local electricity demand is low relative to the minimum size of required supply equipment.

[7] Non-rivalry also extends to the electric trunk network's construction: fixed costs amount to multiples of annual income (cf. Figure A4). Resulting cost curves slope downward in adoption (cf. Figure A5).

modality in the adoption of last-mile electricity – where the local electric trunk network is erected in some locations but not others – holds more widely across many low-income countries in Sub-Saharan Africa, potentially affecting half a billion people.[8]

Third, in line with theoretical mechanisms proposed in the literature on how under-provision of local public goods can be overcome, we highlight the importance of pivotal end-users for the provision of the local trunk network: household and end-user adoption of electricity is concentrated in Zambian locations that have agents with disproportionately high expected benefits of electricity present at baseline (in our context: grain mills). We find that electrified locations with grain mills at baseline experience significantly higher causal household adoption of electricity for lighting, weakly higher household adoption for cooking and significantly higher nighttime luminosity (both higher mean and maximum luminosity, extending for two kilometres from the point of grid arrival).

Fourth, we provide causal evidence that any development effects of rural electrification are concentrated in locations where such pivotal agents are present at baseline, and where the local public good problem gets resolved. It is in these locations that we find structural transformation out of agriculture into services (and, weakly, manufacturing), especially by women, higher construction activity, especially within one kilometre of the point of grid arrival, consequently less green vegetation indicative of agricultural activity in the location's centre, but significantly more in the direct surroundings of locations with grain mills at baseline.[9] In addition, we find positive, significant effects on educational attainment and literacy, due to a combination of increased education and in-migration.[10] Finally, event study analyses of the staggered, country-wide rural electrification expansion confirm that dynamic effects take up to five years to materialise.[11]

With respect to mechanisms, we present evidence that the effect of mills in determining electrification's success follows an inverted U-shape in distance between the mill and the point of grid arrival: locations in which a mill is located around one kilometre from the electric transformer ('the grid') generate some of the largest nighttime luminosity responses following grid arrival.[12] That mills located at the outskirts of settlements have

[8]Figure 1 documents bimodal adoption patterns in rural areas across several large economies in Sub-Saharan Africa. If taken at face value across all countries we have data on, approximately 500 million people may live in locations that are suffering from underprovision of the local electric trunk network.

[9]In line with previous findings of female-centred benefits (Dinkelman, 2011; Dinkelman & Ngai, 2022).

[10]cf. Proffen's (2024) findings in Ethiopia. Figueiredo Walter and Moneke (2024) document an additional mechanism, i.e. electrification attracting/retaining more qualified teachers to rural public schools.

[11]Similar to findings by Collinson et al. (2024), and also hinted at by Burlig and Preonas (forthcoming).

[12]Figure A6 depicts the inverted U-shape in distance between mill and transformer. In our context, grid arrival and transformer location are accurately determined from the universe of 14,380 public schools, their location and annual reports. Electrification of public infrastructure, such as schools, follows a legal mandate, whereas no legal mandate for erecting a local electric trunk network exists. Any such local

stronger effects on electrification than mills directly near the transformer (or those further than two kilometres away, i.e. beyond most settlements' spatial extent) is suggestive of the mill being indeed pivotal for a local electric trunk network to emerge. Similarly, the bimodality of adoption across locations can, to a large extent, be explained by locations with (without) pivotal end-users seeing near-complete (near-zero) adoption.[13]

In addition, given the documented disproportionately high expected benefits of grid electricity for mills, their role in resolving the underprovision of the local trunk network appears likely and can be anecdotally confirmed from various sources in the context of rural Zambia.[14] For example, the utility provides eager mill-owners high-powered incentives to decrease their own cost share of the local trunk network (which enables their own connection to the grid) by rewarding them for bringing more end-users to sign up: instead of paying for the entire trunk network themselves, the miller (or another pivotal player like them) is incentivised by only having to pay their 'fair' share of the total construction cost if they attract or convince more end-users from the same settlement to get connected. Anecdotal evidence also indicates that back payments may indeed function as one practical way how pivotal users ensure the local public good gets constructed.[15] Once the local trunk network is constructed, households and other end-users of electricity face dramatically lower connection cost than those in a counterfactual location without a mill coordinating the erection of (at least parts of) the trunk network. Mills may thus act as pivotal players, enabling the construction of the local trunk network and wider electricity adoptions across a location.

Our setting resonates beyond the specific issue of the provision of last-mile infrastructure for electricity access in rural Zambia: low-income countries are affected by various local public good problems around last mile infrastructure. In fact, most of the recent literature on underprovision of local public goods arose from studies of irrigation systems, while sewerage systems and local trunk roads share similar characteristics. In high-income countries, contemporary issues around the expansion of fibre broadband and the delivery of parcels to the doorstep are conceptually equivalent. Regarding rural electrification in particular, the US in the last century provides historical precedent: a lack of electricity adoption motivated the formation of the Rural Electrification Administration (REA) in 1935, tasked with providing (subsisided) funding to overcome

electric trunk network would span from the transformer, i.e. the school, to the remainder of the village.

[13] cf. Figures A7 and A8 for bimodality in rural Zambia, and pivotal users' role in explaining it.

[14] Figure 2 shows that electricity load from grain mills can account for 50-90% of cumulative local load.

[15] A theoretically isomorphic view is that the presence of a mill resolves a local coordination failure, exhibiting a positive externality on its neighbours, cf. Murphy et al. (1989) and Buera et al. (2023).

the underprovision of last mile electrification.[16] REA in the US also encouraged and subsidised the formation of farmer cooperatives to overcome coordination failure and free-riding among farmers with supposedly high expected benefits of electricity.

Beyond providing one mechanism how local public good problems can get resolved, this paper contributes to the wider literature on electrification: electricity is a key technology for modern production and electrification is therefore highly positively correlated with economic development.[17] However, evidence on the causal effect of electrification on development remains mixed:[18] whereas long-term studies highlights its potentially transformative effects,[19] several rigorous microeconomic studies find only limited or virtually zero short-term effects of electrification on a wide variety of economic outcomes.[20]

Thus, whether and how electricity access can translate into economic development remains unclear. We contribute evidence on when (and where) electrification can succeed by studying the determinants of successful rural electrification in a large-scale, country-wide grid expansion program in Zambia over the last two decades.[21]

To track electrification, we develop a novel measure of electric grid arrival with high temporal and spatial coverage derived from administrative records on the universe of primary schools.[22] This new, geo-identified data on rural electrification allows us to obtain a cross-validated, village-level measure of electrification over time.[23] We confirm that hundreds of rural localities gained access to the grid since 2009.[24] We then test for reduced-form causal effects of electrification and can confirm disappointing local average treatment effects across several specifications in a rural electrification context that resembles recent experience in many low income countries.[25] Interestingly, despite widespread grid arrival, we find that adoption in newly-connected localities is virtually

[16]Kitchens and Fishback (2015) evaluate the REA, noting its crucial role in resolving underprovision.

[17]Figure A9 confirms a robust, positive cross-sectional correlation, and also highlights that most sub-Saharan African countries, including Zambia, remain at disproportionately low levels of electricity access.

[18]Lee et al. (2020a) provide an excellent evidence review, including historical expansion programs.

[19]cf. Dinkelman (2011) in South Africa, Rud (2012) in India, Lipscomb et al. (2013) in Brazil, Fried and Lagakos (2017) in Ethiopia and Kassem (2018) in Indonesia for potentially transformative effects.

[20]cf. Lee et al. (2014), Lee et al. (2020b), Burgess et al. (2020b), Burlig and Preonas (forthcoming).

[21]An obvious source of variation across studies is the underlying margin: firm electrification (Rud, 2012; Grimm et al., 2013; Kassem, 2018) may have quite different effects from electrification of households (Bensch et al., 2011; Peters & Vance, 2011; Lee et al., 2015; Lee et al., 2020b), whereas urban electrification is characteristically different from its rural counterpart. Allcott et al. (2016) show that extensive margin expansion is likely constrained by the intensive margin of available electricity supply in connected locations. We study rural electrification and its effect on households and the local economy.

[22]By government mandate, primary schools are initial target points of Zambian rural grid expansion.

[23]cf. Figure A10 and Table A2 for graphical and empirical evidence that our electrification measure correlates highly with alternative, but imperfect measures of electrification across rural Zambia.

[24]cf. Figure 3 for country-wide electrification after inception of the Rural Electrification Master Plan.

[25]Similar rural or universal electrification plans were implemented in Ghana (1990), Ethiopia (2005), Kenya (2009), Rwanda (2017) in SSA; Indonesia, Thailand, Vietnam, Laos, Myanmar in Southeast Asia.

bimodal: either very few or almost all potential end-users adopt electricity. A core result is that only localities with pre-existing pivotal end-users – predominantly grain mills in our context – experience positive, large treatment effects (and wide adoption).

Since the electrification planners in the 2000s lacked accurate data on mill locations, the resulting rural electrification program did not specifically target mills and their surrounding locations. Instead, planners fell back on population growth-derived demand projections to determine suitable target locations. Having obtained two independent sources of data on the exact location of mills (and other potentially pivotal end-users with expected high benefits of grid electricity) across rural Zambia at baseline, we therefore explicitly test if pre-existing mill presence acts as a potential determinant of successful electrification, i.e. widespread local adoption and positive development effects.

The context of the recent Zambian rural electrification program provides a promising empirical setting for at least three reasons: first, it features a large-scale electrification expansion coupled with a natural experiment implied by the roll-out plan. Second, we collate extremely rich administrative data that allows us to generate a novel measure of electrification status and progress across space and over time, derived from the universe of Zambian primary schools. Third, we geo-locate rich outcome data that also allows us to determine locations' pre-existing presence of potentially pivotal end-users.

Regarding the first, the large-scale expansion we study originates from the Zambian Rural Electrification Master Plan (REMP) which set the objective of a 51% rural electrification rate by 2030, starting from approximately 3% rural electrification at baseline in 2009. Extensive international technical assistance on electricity network expansion cumulated in the REMP, a sophisticated expansion plan. It defined a location-specific ranking of electrification priority, and an algorithm to optimally bundle different locations. Bundling locations of vastly different ranks, the plan admits a classic spatial inconsequential units design (Redding & Turner, 2015). We study the subset of locations that are 'accidentally' electrified earlier than their rank would imply since they happen to lie on a path between the existing grid and a highly-ranked location further afield.[26]

Regarding the second, the study of rural electrification in low-income countries often suffers from lack of accurate information on the exact timing and placement of the low-voltage grid as it expands at scale. We present a novel measure to track the status and progress of rural electrification across time and over space, which provides us with

[26] Figure 5 provides a schematic drawing of the concept of bundling RGCs into packages. The figure displays one electrification package, characteristically with an existing substation (orange circle), a highest-ranked RGC as target (red circle), as well as lower-ranked, inconsequential RGCs (blue circles) on the way, that accidentally receive electricity earlier than their own ranking would imply. Figure A14 confirms that this quasi-experimental ideal holds surprisingly well in the actual roll-out plan.

a cross-validated, village-level dataset of electrification in Zambia from 2005 to 2020. Our measure is derived from administrative records collected annually by the Ministry of General Education from the universe of Zambian primary schools, which includes a question on schools' grid access. After tedious geo-identification of more than 14,000 primary schools, we obtain a panel dataset of school-level (and, therefore, village-level) electrification. Legal and technical aspects of rural electrification provide peculiar features that make public schools prime electrification targets. For example, the REMP includes a legal mandate for Zambia's REA to ensure all public infrastructure in electrified locations gets connected.[27] Therefore, school electrification confirms the presence of a transformer in the direct vicinity of the school – and the presence of a transformer in a location equals a necessary condition for any households, firms, streetlights or any other end-user to receive a grid connection. We are not aware of any previous work that made use of this novel geo-identified, cross-validated, location-level measure of grid arrival.

Regarding the third, we obtained and geo-identified unusually rich data on outcomes and pre-existing conditions in Zambia before and during grid expansion since the late 2000s: two full Population & Housing Censi (2000/2010), National Labour Force Surveys (quarterly, 2017-2022) and two Living Conditions Measurement Surveys (2010/2015).[28] To test if the underprovision of the local electric trunk network can be overcome if pivotal end-users with high expected benefits of electricity are present at baseline, we exploit a large, fully geo-identified dataset of 58,000 'points of interest', including mills, churches, banks, administrative buildings and commercial operations, e.g. shops or workshops.[29]

Our analysis employs a triple differences empirical strategy restricted on the sample of inconsequential locations identified in the plan: we compare inconsequential locations electrified by 2020 against inconsequential locations that remain unelectrified, interacting electrification status with pre-existing presence of pivotal end-users, such as mills.[30] We find that the strong bimodality in adoption is confirmed in regression results: any statistically significant positive effects of electrification on adoption, usage and development effects are confined to those locations with a pre-existing mill at baseline. The prominent role of pre-existing mills in determining local electrification effects is confirmed across various specifications, outcome variables and a plethora of robustness checks.

Why does electrification work in some places, but not in others, and what is the

[27] In rural locations, the only instance of public infrastructure is often the public primary school.

[28] Figure 4 provides an overview of the various administrative, household and remotely-sensed datasets.

[29] The 'points of interest' dataset was enumerated as by-product of the 2010 Population and Housing Census, but was not included in the original Census datasets.

[30] Figure 9 confirms that the identifying assumption of parallel ratios appears to hold in event study plots of the interacted difference-in-differences, both for mean/max nighttime luminosity (lower panel).

underlying determinant of bimodal adoption? In this paper, we argue that rural electrification poses a local public good problem in grid-electrified villages due to the high fixed cost of last mile infrastructure, i.e. the local electric trunk network to connect end-users. To provide evidence on the underlying mechanism, we present a series of evidence to confirm that pivotal end-users indeed drive electrification outcomes via resolution of the underprovision of the local low-voltage trunk network. For example, mills that are far, but not too far away, mills in larger settlements and mills in locations without alternative public buildings matter the most. Finally, the utlity's high-powered incentives for interested end-users to rally their neighbours for a costly connection have the most bite for those end-users with the highest expected benefits, i.e. mills in rural locations.

The remainder of this paper is organised as follows: Section 2 characterises last mile electrification as a local public good problem. Section 3 introduces the empirical context of rural electrification in Zambia. Section 4 describes the data. We then present the empirical strategy (Section 5) and reduced-form results (Section 6), distinguishing between locations with and without pivotal end-users present at baseline. Section 7 concludes.

2 Last mile electrification: a local public good problem

2.1 The electric trunk network as local public good problem

In this paper, we characterise the provision of the local electric trunk network as a local public good problem. In general, last mile infrastructure allocations often display public good characteristics. For example, local irrigation, sewerage and low-voltage electricity networks are often both imperfectly excludable and imperfectly rivalrous. Persistent problems in their provision are well-documented: in fact, evidence on the eponymous 'last mile problem' exists for electric trunk, irrigation, freshwater and sewerage networks. The 'last mile problem' is also not confined to low-income countries, with landline phones, mail delivery and rural electrification representing historical examples in the US and elsewhere, whereas today's high-income countries are addressing issues around fibre broadband access, 5G mobile towers, natural gas grids or district heating systems.

In fact, the US Rural Electrification Administration was founded in 1935 not only to advance universal rural electrification, but in part also to overcome underprovision of local electric trunk networks and free-riding (via waiting for a neighbour to connect first) in the US (Rural Electrification Administration [REA], 1960, 1981; Kitchens & Fishback, 2015; Lewis & Severnini, 2020). However, the absence of last mile access greatly undermines the benefits and effects of the entire infrastructure investment required to reach a given location in the first place, which remains underutilised.

Our setting is rural electrification in one of today's low-income countries, Zambia. Rural electrification implies end-user access to the electric grid. However, end-user low-voltage access requires a local low-voltage trunk network that bridges the distance between end-user and point of grid arrival (the 'transformer', a piece of equipment that steps down voltages from medium- to low-voltage). Grid arrival itself requires an extension of the medium-voltage line to the location in question, which we take as given in our context. To fix ideas, Figure A2 provides photographs of the relevant pieces of infrastructure in rural Zambia: in the upper panel, a medium-voltage line on a tall wooden pole passes through a given location. A grey transformer to step down medium- to low-voltage is installed almost directly underneath, from which a local low-voltage trunk network on lower wooden poles is spanned in three directions, leading to end-users. In the lower panel, an end-point of the local low-voltage trunk network on a wooden pole is pictured, from which four individual end-user connections ('service drops') depart, and an electrical earthing. Both the white and brown houses ('end-users') on the left have fuse boxes and electric metres installed on their exterior walls, from which internal wiring connects appliances, e.g. a suspected television with satellite dish in the white house.

In our context, the provision of a local electric trunk network required for last mile end-user connections can be characterised as a local public good problem: once erected in the form of a set of poles, overhead lines and service drops, connections to the local trunk network are imperfectly excludable and non-payment for electricity is common, both in our setting of rural Zambia and elsewhere around the world. Burgess et al. (2020a) and Burgess et al. (2020b) show large non-payment in India, which we can confirm in our sample. Figure A1, based on Living Conditions Measurement Survey 2015 household data, shows that especially in rural Zambia, a large minority of locations (here: enumeration areas) feature end-users who self-report (sic!) not paying for their electricity consumption. Although technical non-payment is certainly a possibility (i.e. meters not yet being read, while connections are live), Table A1 confirms that approximately one quarter of locations in rural Zambia have end-users who do and do not pay co-existing in the same location. We interpret the latter to strongly suggest non-payment and non-excludable consumption of electricity once a local low-voltage trunk network is in place. In fact, visual inspections of selected rural locations in our sample by the authors revealed widespread illegal service drops, and common self-reported non-payment.

Furthermore, usage of electricity consumed from the local network is imperfectly rivalrous since the electric trunk network is designed by expert engineers to meet peak local electricity demand, while the supply of electricity is determined globally by large-scale power generation infrastructure elsewhere, which is unaffected by local variation in electricity consumption from individual end-users in the rural settlements we study.[31] In other words, local electricity demand in our sample is low even relative to the minimum available size of infrastructure equipment (cf. the transformer in Figure A2) required to supply it. Standard load curves (cf. Figure A3) also highlight that most end-users in our sample are small and inconsequential, with the exception of electrified grain mills, which we discuss in detail below. However, even for relatively large (rural) end-users such as grain mills, consuming a marginal hour more electricity is highly unlikely to generate scarce supply (in form of an electric outage) for neighbours, confirming imperfect rivalry.

Beyond electricity consumption, non-rivalry also extends to the local electric trunk network's construction: the fixed costs required to erect the trunk network amount to multiples of annual income (cf. Figure A4), and the resulting average total cost curve per end-user is sloping downward in adoption (cf. Figure A5), as shown elsewhere for rural Kenya. At least initially, for every new end-user the total cost per end-user falls for everyone. Therefore, we treat the provision of the local electric trunk network as a local public good, with a particular focus on its potential underprovision.[32]

2.2 Bimodality in electricity adoption

If the local electric trunk network is understood as a local public good, underprovision due to failure to resolve the local public good problem becomes a natural issue: most end-users can remain unconnected despite the electric grid having arrived. Therefore, locations can find themselves in one of two extreme worlds: either the local electric trunk network gets built and many connect, or it does not and virtually no end-user connects.[33]

We document a new empirical regularity across grid-connected locations in rural Zambia which also holds across many countries across rural Sub-Saharan Africa: locations' share of household electricity connections is bimodal conditional on prior grid arrival – with locations either experiencing near-complete electrification or none at all. This bimodality in the adoption of last-mile electricity – where the local electric trunk network is erected in some locations but not others – holds more widely across dozens

[31] Whereas large end-users, such as a copper mine or an aluminium smelter could introduce frequency variation in the electric grid from sudden consumption changes, no such end-user exists in our sample.

[32] It is important to note that, as Bardhan et al. (2007) show, even if instances exist where the trunk network resembles a club good more than a public good, the theoretical predictions presented below still hold. We understand our context as a case of an imperfectly rivalrous and imperfectly excludable good.

[33] In our context (cf. Section 3.2), public infrastructure in Zambia is mandated to be a target of rural electrification, so at least one end-user will always connect conditional on grid arrival: the public school.

of low-income countries in Sub-Saharan Africa potentially affecting half a billion people. Figure 1 documents bimodal adoption patterns in rural areas across several large economies in Sub-Saharan Africa, which account for approximately 500 million people that live in locations that suffer from underprovision of the local electric trunk network.

The data underlying the suggestive evidence on bimodal end-user connections in locations with grid access come from nationally representative, standardised and comparable household surveys, such as the World Bank's Living Standards Measurement Surveys. In particular, we employ data from the Ethiopian Socioeconomic Survey 2018, the Malian Enquête Harmonisée sur le Conditions de Vie des Ménages 2018, the Nigerian Living Standards Survey 2018 and the Zambian Living Conditions Measurement Survey 2015.[34] We are interested in identifying rural locations (here: enumeration areas) that technically have grid access in the form of a transformer present. Two measures are available: either, the enumerator themselves reports if a given enumeration area has grid access, or we can derive (a lower bound for) grid access from those enumeration areas where at least one respondent (or any other, arbitrary threshold) reports having grid access. We then plot the remaining household's share of electrification for this subset of grid-electrified locations. Especially less densely populated rural areas in physically larger countries appear to show a bimodality pattern (cf. Figure 1), while no bimodal pattern can be detected in urban areas. Our finding confirms previous anecdotal evidence, e.g. from Ethiopia, documenting common 'ghost electrification', i.e. no grid connections beyond the initial transformer, and problems around 'last mile' grid access reported elsewhere.[35]

What are the potential reasons underlying the bimodal distribution of electricity adoption in electrified locations highlighted above? The underlying issue driving the bimodality result appears to be the local electric trunk network's high fixed cost (cf. Figure A4), where the lion's share of cost arises from the 'communal' part of the local electric trunk network, especially poles and overhead lines, leading to a downward sloping cost curve along (most of) the within-location adoption spectrum.[36] Therefore, free-riding and coordination failure abound: every additional connection would spread the high initial fixed cost of the trunk network over more end-users, therefore lowering the cost for every end-user. However, at the same time, the utility's costing approach (explained in

[34]We can also confirm our bimodality results using more widely available Afrobarometer surveys.

[35]It is important to note that our finding is conceptually distinct from 'under grid' end-users (Lee et al., 2014). By definition, those 'under grid' are within 200 metres of an existing trunk network. In contrast, our finding pertains to locations that could erect an electric trunk network, but fail to do so.

[36]As Lee et al. (2020a) show, depending on the estimation strategy for the average total cost per connection, scenarios exist where cost per end-user may increase again beyond 70% location-wide connections. In our context, the second ('full adoption') mode usually appears around 75% adoption, in line with increasing cost per end-user for the final few connections, most likely located at the outskirts.

detail in Section 3.3) can unintentionally reward potential end-users for waiting, free-riding on their neighbours' investment into the trunk network, then connecting at a fraction of the fixed cost later. ZESCO, the utility in charge of electricity distribution in Zambia confirmed in private conversations with the authors 'strategic waiting' behaviour as binding constraint for expanding rural electricity access, which inspired the design of ZESCO incentive structures to try to address this issue, as we discuss below.

2.3 Resolving underprovision: pivotal agents and unequal benefits

Economists have long studied the underprovision of public goods. Free-riding and coordination failure can lead to underprovision in equilibrium. Samuelson (1954) first documented underprovision of public (compared to private) goods in equilibrium, which Vickrey (1961) and Groves and Ledyard (1977) formalised more generally.

Theoretical mechanisms designed to achieve efficient provision of public goods abound. In particular, a large literature discusses the determinants of the degree of underprovision: Olson (1964) spawned an entire field of study on collective action problems, beginning with observations how different group characteristics would ease or hinder the provision of public goods or taking of collective action. If the unequal distribution of pre-existing characteristics, e.g. wealth, will exacerbate or defuse collective action can depend on the functional form of individual contributions (Cornes, 1993), the purity of the public good (Cornes & Sandler, 1996), or the nature of the externality (Bardhan et al., 2007). Empirical examples on the role of individual actors and inequality within or across groups include the study of irrigation networks (Wade, 1994; Bardhan, 2000), social capital Alesina and La Ferrara (2000), dyke maintenance (Cornes, 1993), or place-based infrastructure provision (Banerjee, 2004; Banerjee et al., 2007).

However, systematic empirical evidence on how the underprovision of local public goods can be overcome is rare. This paper provides new evidence on how a local public good problem in the context of last mile infrastructure can be resolved.

The particular local public good problem faced here, the provision of a local electric trunk network, closely resembles a classic example in the literature, the 'public project'. A public project represents the situation in which a community needs to decide whether to build a public project at a cost c. The cost are to be equally divided among members of the community such that $\frac{c}{n}$ represents cost per user. This closely resembles our context where each additional end-user helps bring down cost per end-user for everybody who participates. Let us define utility derived from the public project as v_i, which is allowed to vary across potential end-users i.[37] The decision to construct the public project adheres to the simple rule:

$$\sum_i \hat{v}_i \geq c$$

such that construction will go ahead iff the sum of reported valuations $\hat{v}_i$ exceeds cost c. Free-riding by any given member of the community can be represented as deviations between the sum of actual and the sum of reported valuations: $\sum_i v_i \geq \sum_i \hat{v}_i$.[38]

When will a given public project be built? How can underprovision of the local public good be resolved? Vickrey (1961), Clarke (1971), Groves (1977) and Groves and Ledyard (1977) highlight a theoretical mechanism that could overcome underprovision: the pivotal end-user. In Clarke's (1971) formulation, a mechanism can be designed using transfers such that the one end-user without whose valuation total (actual) valuation of the public project would fall below total cost incentivises the other community members to report their valuations truthfully.[39] Although the mechanism will not achieve budget balance (Groves & Ledyard, 1977), it can achieve provision of the local public good and overcome free-riding. As discussed below, since government and utility are incurring financial losses on rural electrification, the constrained-optimal outcome would be to ensure utilisation of existing, newly-built medium-voltage infrastructure by means of erecting local electric trunk networks. Hence, we are not concerned with the implied lack of budget balance.

If one interprets the Vickrey-Groves-Clark mechanism as requiring end-users with disproportionately high expected benefit of rural electrification, our empirical context features at least one group of end-users that could reasonably be described as 'pivotal': hammer mills, which are by far the biggest expected end-users and beneficiaries of grid electricity in rural Zambia. As Figure 2 highlights, the daily electricity demand profile (the location's 'load curve') of two given, randomly sampled electrified locations is all but dominated by electricity consumed from hammer mills. Figure A3 provides further load curves from other locations, and also provides details on alternative end-users' demand: mills are responsible for cumulative demand of 50-90% of rural locations' total electricity demand, and no other type of end-user admits a similar demand profile.[40]

[37] In other words, we allow for inequality in expected benefits of the trunk network across end-users.

[38] As discussed in greater detail in Section 3.3, free-riding in our context may take the form of waiting for another community member to move first, which we can represent as underreporting valuation.

[39] Such a mechanism could work as follows: if total reported valuation exceeds cost, the public project gets built. Each end-user then pays a fee equal to the difference between project cost and every other end-user's valuation. Only the pivotal user pays a non-zero fee, and truth-telling is a dominant strategy.

[40] In addition to the sheer dominance for locations' total electricity demand, mills' disproportionate benefits of electrification can likewise be confirmed: in focus group interviews with five millers in north-western and northern Zambia, cost reductions of approximately 80% against diesel-powered operations were reported. Such private gains would be even higher for those who previously milled traditionally.

If the Vickrey-Groves-Clark mechanism itself is realistic remains untestable without transaction-level data from each electrified location before, during and after electrification. However, in the absence of such data, we describe and discuss potential incentive schemes put in place by the utility that emulate such a mechanism in Sections 6.3 and 6.4 below. Anecdotal reports of back payments and kickbacks in electrified locations may be symptoms of how coordination failure was overcome, and who was involved.[41,42]

In line with theoretical mechanisms proposed above on how underprovision of local public goods can be overcome, we empirically test for the importance of pivotal end-users in ensuring the provision of the local trunk network: can the bimodality in electricity adoption be resolved by the presence of pivotal end-users in a given location at baseline? Do adoption and usage differ in such locations? Finally, are any benefits of electrification concentrated in locations with pre-existing pivotal end-users (here: hammer mills)?

Ours is by no means the first paper to empirically investigate the underprovision of local public goods. However, as Banerjee et al. (2007) note, although the available evidence appears "generally consistent with [collective action] theories" (ibid.), there is a "dearth of quality evidence" (ibid.). This paper asks what it takes to overcome the underprovision of local public goods. In the following, we introduce a promising empirical context, rich data and an empirical strategy that allows us to provide systematic empirical evidence required to test the theoretical predictions presented above.

3 Setting: Zambia's Rural Electrification Master Plan

3.1 Large-scale electrification expansion

In this paper, we study the local public good problem that originates from a large-scale, country-wide rural electrification expansion in a low-income country. The Zambian Rural Electrification Master Plan (REMP) set an objective of a 51% rural electrification rate by 2030, starting from approximately 3% rural electrification at baseline in 2008. Virtually non-existing rural access to and adoption of grid electricity led to the development of the REMP in the mid-2000s. The resulting rural electrification plan closely aligns with similar universal access programs implemented elsewhere in Sub-Saharan Africa

[41] An alternative interpretation of why inequality in expected benefits between miller and other end-users may help overcome underprovision is provided by Olson (1964), who argued provision was more likely if most benefits from a public good are captured by a small group with strong stakes in its provision.

[42] Bardhan et al. (2007) highlight yet another interpretation, by pointing to the role of pivotal end-users as 'elites' (Wade, 1994). However, they note that in their more general model of public good provision, inequality has an ambiguous effect on underprovision – a finding echoed by Banerjee et al. (2007).

around the same time.[43] Subsequently, a Rural Electrification Authority (REA) was established in 2003, and the first isolated electrification projects were reportedly implemented in 2006. Since around 2006, the Japanese development agency, JICA, provided extensive technical assistance around electrification and network expansion planning, which cumulated in the REMP, a sophisticated expansion plan arrived at by employing state-of-the-art network planning engineering tools and algorithms. According to REA records, at least $35m was spent on rural electrification projects until 2017.[44] With additional international funding becoming available, rural electrification accelerated from 2014 onwards (cf. Figure A12), reaching 18% electrified target locations, so called Rural Growth Centres (RGCs) by 2020.[45] Therefore, our study period focuses on the decade of rapid electrification progress from 2010 to 2020.

3.2 Legal mandate to electrify public infrastructure

A prominent feature of rural electrification in Zambia is the legal mandate for the Rural Electrification Administration (REA) to connect all forms of public infrastructure in locations selected for grid electrification. In the context of rural Zambia, REA's mandate mostly pertains to the eletrification of public schools and health facilities as the most common and widespread types of public infrastructure.[46,47]

This legal mandate has two implications for our empirical analysis: first, the original electrification expansion targets unelectrified places with such public infrastructure present at baseline. In fact, the presence of a public school or health facility in a loca-

[43] The REMP's timing and focus on rural electrification towards universal access is similar to, e.g. the Ethiopian Universal Electricity Access Program (2005), the Kenyan Rural Electrification Master Plan (2009), the Rwandan Universal Electricity Access Program (2017), or the Ghanaian National Electrification Master Plan (1990-2000, urban and rural). Similar National Electrification Plans (NEP) operate in many low-income countries. In general, these earlier plans and programs closely adhere to an universal access objective established by the Sustainable Development Goals (SDG 16 'Universal Access to Electricity') in 2005 and promoted heavily by relevant institutions such as the World Bank.

[44] The actual budget expenditure on REMP projects falls far short of the budget JICA deemed necessary to meet the plan's objectives by 2030, which explains the initially slow progress until 2012.

[45] As explained in greater detail in Section 5 and Appendix A.IV.1 below, the REMP defines a set of 1,200 locations as so-called Rural Growth Centres, to be electrified by 2030. RGCs are best thought of as rural locations with some form of economic potential, although exact selection criteria remain opaque.

[46] At baseline, approx. 14,400 (2,400) public schools (public health facilities) existed in Zambia.

[47] A second prominent feature of electrification in this context is the engineering-motivated search for dependable electric load (i.e. demand for electricity) in newly electrified locations Regarding the REA engineers' search for dependable ('anchor') load as source of stable electricity demand in rural locations, REMP engineering case studies confirm the beneficial load curves of primary schools in Zambia. The median primary school demands electricity throughout day and night, most likely due to a combination of indoor electricity demand from teaching during the day, and electricity demand from teachers' houses and school compound outdoor lighting at night. In contrast, other common public (or semi-public) buildings, such as churches, display a less favourable load profile to qualify as anchor load.

tion represents the smallest common denominator of what qualifies a given location to be included as Rural Growth Centre (RGC) in the REMP.[48] Second, since the school or health facility must receive a connection (while no such mandate to connect exists for any other potential end-user), the constrained-optimal solution that fulfils the legal mandate is to ensure grid arrival at the exact location of the school or health facility.

In practice, this logic can be confirmed throughout rural Zambia: medium-voltage transmission lines usually extend to the direct vicinity of a public school or health facility, and the location's (first and potentially only) transformer will be placed right next to the school building or centrally on the school (or health facility) compound.[49] As we confirm in Section 4 below, tracking school electrification equals the arrival of the grid in a given RGC, i.e. the potential for grid electrification of all (potential) end-users across the entire location. The 'only' remaining problem to be solved for the community is how to erect the local low-voltage trunk network that distributes grid access via local poles and low-voltage lines to individual end-users' service drops and metres.[50] This is the local public good problem described in Section 2 above.

3.3 Local electric trunk network for end-user access

From the authors' personal conversations with REA engineers, we understand the process of rural electrification in Zambia since the inception of the REMP as follows: REA extends medium-voltage lines to individual RGCs and takes responsibility of installing a transformer in a given location. This transformer will commonly be placed in the direct vicinity of a public school or health facility, as explained above. REA contractors then connect the school building and teacher houses (or: health facility building and nurses' and doctor's houses) to the transformer, install an electric meter and internal wiring. At this point, REA's legal mandate to electrify public infrastructure ends and the entire electrification project is handed over to the electric utility, ZESCO.[51]

[48] Since the expansion of universal primary education precedes the expansion of universal healthcare access in Zambia, almost all RGCs feature a public school, while some also feature a public health facility.

[49] A transformer denotes the sizeable piece of equipment that steps down medium- to low-voltage. Figure A2 (upper panel) shows a picture of a standard-sized Zambian transformer, next to a school.

[50] Any grid infrastructure (except for the equipment required to connect the school) that is situated beyond the transformer (e.g. low-voltage overhead lines, poles, service drops, electric metres, internal wiring, cf. Figure A2 for pictures) falls into the domain of the electric utility, ZESCO. REA routinely hands over electrification projects to ZESCO once their legal mandate in a given location is fulfilled.

[51] Towards the end of our sample in 2019, REA started implementing partial construction of the low-voltage trunk network in a selected handful of locations, once guidance from the policymaker was updated and matched with additional funding. Hence, REA is moving to a system where they provide partial low-voltage infrastructure. Focus group interviews in affected RGCs electrified under this scheme in 2022-2023 highlight confusion and uncertainty among interested (unelectrified) end-users in why a

The utility receives and manages any applications for individual end-user connections, e.g. for households, firms or other potential end-users. Any individual or group application involves standard engineering ground surveys and costing exercises to determine the location-specific fixed cost of erecting a local low-voltage trunk network, i.e. erecting additional poles, overhead lines, line drops to individual end-users, installation of meters and internal wiring. For example, as Figure A4 shows, the REMP estimated an approximate ZMK3m (USD1,000 in 2009) mean total fixed cost of connection based on a simple single-phase local low-voltage trunk network in a given rural area, representing between 3-6 months of average rural household income in the mid-2000s.

As remoter locations (which require longer medium-voltage lines to be reached) and less densely populated ones (which require longer low-voltage trunk network lines) get electrified over time, average total fixed cost of connection are bound to increase further. In addition, the weakening of the Zambian currency during our study period, which lost more than two thirds of its value against the US Dollar (including a redenomination from the First Kwacha (ZMK) to the Second Kwacha (ZMW) in 2012) inflated the cost of imported inputs such as transformers. Thus, without any subsidy scheme in place, the REMP's original concern with the extremely high fixed cost of connection faced by any new individual customers in rural Zambia (cf. Figure A4) still applies today.[52]

ZESCO, as a profit-maximising private enterprise under government ownership, employs realistic costing to local low-voltage trunk networks and individual connections. Thus, ZESCO engineers confirm that the more people connect, the cheaper the fixed cost gets, i.e. confirmation that the total cost curve of RGC-level electrification is similarly downward sloping as elsewhere, cf. Lee et al.'s (2020b) finding from rural Kenya, reproduced in Figure A5. Although ZESCO claims to calculate the fixed cost of connection for the first customer in a given location on the basis of the average fixed cost of connection of the entire location (and full adoption) to undermine free-riding behaviour by neighbours, we cannot confirm this practice to be implemented on the ground: customers report receiving varying quotes over time, while locations experience high aggregate population growth, settlement densification and in-migration (or lack of out-migration (Fried & Lagakos, 2017)) once the grid arrives. Therefore, ZESCO openly admits in conversations with the authors that one of their biggest concerns about rural electrification was free-riding by neighbours in the form of waiting. For most individual customers, the

certain partial trunk network was built, and if (or when) full trunk network construction may ensue.

[52]Furthermore, anecdotal evidence also confirms that applying for a connection with ZESCO and a ZESCO engineer showing up may take several months, if not longer. It is unclear if these delays are an endogenous response by ZESCO to avoid connecting further rural, loss-making customers to the grid, or if ZESCO faces other constraints in the supply of meters, wires, or sufficient engineering capacity.

returns to waiting and receiving a lower fixed cost quote once neighbours have already paid for large parts of the local trunk network appear large. In fact, as our bimodality result for Zambia in Figure A7 confirms, numerous locations receive a transformer that connects the school, while no local low-voltage trunk networks gets erected and no other potential end-user adopts.[53] Anecdotes of coordination failure and free-riding abound.

One implication of this empirical context, which appears common of rural electrification expansions across SSA, is that we can focus on schools and health facilities, and their electrification status, as effective ground-truth measure of grid access in a given location, although we require different data to measure actual adoption as evidence that a local low-voltage trunk network has been erected. Below, we introduce how to construct these measures and other data sources we rely on for the empirical analysis.

4 Data: tracking rural electrification in Zambia

4.1 Novel ground-truth measure of electric grid expansion

The study of rural electrification in low-income countries often suffers from a lack of accurate information on the exact placement and timing of the low-voltage grid while it is being rolled out. In Zambia, as the engineers tasked with designing a rural electrification plan noticed themselves at the time, this lack of information appears to have been particularly acute: no comprehensive database of the universe of low-voltage lines, transformers and service drops appears to have existed in the mid-2000s, leading to a fundamental lack of data on both the extent of rural electrification at baseline in 2008 and the subsequent grid roll-out until the end of our sample in 2020.

To fill this data gap, we present a novel measure to track the status and progress of rural electrification across time and over space, which provides us with a cross-validated, geo-identified annual village-level dataset of electrification in Zambia from 2008 to 2020. The various steps required to arrive at this novel measure are briefly described below, while comprehensive detail on the data acquisition, cleaning, geo-identification and merging is provided in Appendix A.III.

In particular, we exploit a feature in the Ministry of General Education's administrative data that requires all public school headmasters across Zambia to report annual school-level statistics and characteristics. School headmasters are asked to fill out a short form which includes a question on their school's access to electricity, differentiat-

[53] A ground-truthing exercise to northwestern Zambia in January 2024 confirmed that out of two dozen RGCs inspected in person by the authors, less than a dozen had a low-voltage trunk network, while the majority featured a single transformer placed next to the school without any trunk network.

ing between grid and off-grid access, and an additional question if this connection was functioning. We understand this measure as a form of effective 'ground truth' since, first, public schools were explicit, primary targets of rural electrification (as described in Subsection 3.2 above). Second, headmasters are often the most educated and knowledgable person in rural locations, making them ideal sources of accurate information on when major pieces of new infrastructure such as the electric grid (which also affect school operations) arrive. The extraction, cleaning and panelisation of the annual report-derived headmaster answers on school electrification (and, thus, grid arrival at the village) are described in Appendix A.III.1.

The resulting school-year-level dataset requires geo-identification to be matched to rural growth centres (RGCs), the unelectrified spatial units similar to larger villages and towns of particular interest for which electrification targets were specified. Given the REA's legal mandate to electrify all pieces of public infrastructure in RGCs reached by the electric grid, the electric transformer required to step-down transmission (medium) voltages to usable distribution (low) voltages would ideally be placed in direct proximity of public infrastructure. In addition to the legal mandate, as JICA and REA engineers verified from field surveys, schools, and in particular the teachers houses (provided to teachers directly in the vicinity of the school buildings), appear to qualify as sources of stable daily 'anchor load' in villages, making them prime end-points of low-voltage distribution networks (and virtually the first customer in a village). Since public schools and rural health centres (or smaller health posts) are by far the most common features of public infrastructure in rural Zambia, a school or health facility reporting access to the grid can be safely assumed to mean that a transformer was placed essentially next to the school.[54] Put differently, as natural target points for grid extension, the headmaster reporting grid access *de facto* translates into a transformer being present. Such a confirmed presence of a transformer in a given location provides the necessary precondition for any wider village-level grid electrification once an electric trunk network from the school to other end-customers and the wider community is erected.[55]

After tedious geo-identification of 14,380 primary schools to find all permissible matches with the 1,200 RGCs, especially the 784 inconsequential ones, we obtain a panel dataset of school-level electrification from which the year of each electrified RGC's

[54] This crucial insight is confirmed by REA engineers, as well as by on-the-ground inspections of a random subset of RGCs and schools by the authors.

[55] Figure A10 provides some visual validation, plotting schools (and health facilities), rural growth centres and the most recent information on medium- and low-voltage distribution grid infrastructure derived from construction records as of 2023.

first connection can be determined.[56] The derivation of RGC-electrification status from several, geo-identified schools in a given RGC is described in Appendix **??**.

Our measure of school-derived grid electrification provides the first reliable, annual source of grid electrification across the near-universe of populated places in rural Zambia. We are not aware of any previous work that made use of this novel measure to arrive at geo-identified, cross-validated village-level information on the availability of the grid – short of a geo-identified full history on transformer location and commissioning from the respective utilities.[57] Figure 3 provides the resulting graphical representation of electrification progress across rural Zambia from 2008 to 2020, whereas Figure A11 provides the same measure for the subset of locations that qualify as (wards containing) inconsequential locations in the sense of the Rural Electrification Master Plan (REMP), which form the main units of interest for our empirical strategy below.

Our novel methodology to track electrification at high-frequency at the village-level in low-income country settings by relying on and integrating various administrative datasets is easily applicable to other contexts with similar administrative reporting systems in place. Furthermore, its accuracy can be validated with several alternative measures that may be expected to have higher precision while suffering from imperfect spatial or temporal coverage. As Table A2 confirms, schools' reporting of access to the electric grid at endline (2020) or over time (2008-2020) correlates highly and statistically significantly with: first, the electrification time and status of the subset of REA projects for which historical construction records are available (both in pooled (column (1)), or difference-in-differences (column (2)) specifications); second, with the presence of a ZESCO (or REA) transformer in 2023 within two kilometres (column (3)); third, with the electrification status, either defined as electrification indicator (column (4)) or percentage (column (5)), of 3,400 rural health facilities derived from the Ministry of Health's Health Monitoring Information System (HMIS) data for the year 2017; fifth, with HMIS-derived electrification, either as indicator (column (6)) or percentage (column (7)), in difference-in-differences specifications employing all three years of available HMIS electrification data (2005, 2012 and 2017).

Overall, our school-reported electrification measure performs remarkably well against these alternative forms of ground-truth data from administrative and project records,

[56] The rather involved geo-identification procedure is explained in detail in Appendix A.III.2 below.

[57] Recent ground-truthing efforts by ZESCO, the Zambian electric generation, transmission and distribution utility and the Rural Electrification Administration are beginning to assemble a full status quo cross-section of geo-identified low- and medium-voltage infrastructure. However, this new geographic information system database (cf. Table A2, column (3)) still lacks historical records on infrastructure commissioning or replacement to effectively track grid expansion at high temporal and spatial resolution.

albeit with the unique benefit of covering almost the full universe of rural growth centres at annual frequency since at least 2008.[58]

4.2 Rich baseline data across rural Zambia

While the Rural Electrification Master Plan was concluded in 2008 and implemented starting in 2009, no meaningful funding was released or allocated to it until 2012. The *de facto* start of rural electrification under the master plan is also confirmed in electrification data: Figure A12 confirms that only from 2012 onwards did the electric grid reach new rural growth centres, and the plan's implementation accelerated with further funding becoming available in 2014 and 2017. Hence, for the purposes of our empirical analysis, we define the period before 2012 as baseline, noting that several data sources are available to provide rich baseline information across rural Zambia.

Three unusually rich sources of baseline information and pre-existing conditions are available: first, geo-identified population census and household survey data. In particular, we employ the 100 % sample of the Population and Housing Census (PHC) 2010, as well as the Living Conditions Monitoring Survey (LCMS) from 2010, which both provide detailed information on household-level information such as access to electricity, energy source used for lighting, energy source used for cooking and standard durable asset holdings, such as appliances, housing characteristics and other proxies for household real wealth (cf. Young (2012)). Individual-level schedules also include information on education, gender, educational attainment, migration and occupational choice for the entire adult population.[59]

Second, we use a geo-identified cross-section of 58,500 points of interest collected in 2010 as a by-product of the PHC 2010, that were made available only much later and separately from any original PHC 2010 data products. This dataset contains coordinates and a classification of any given 'point of interest', which includes features as diverse as grain mills, churches, commercial entreprises, administrative buildings, wells, bus stops, banks or sewage systems. To the best of our knowledge, the points of interest

[58]As Table 1 and Figure 6 highlight, 742 out of 784 inconsequential, non-off-grid electrified rural growth centres can be matched to a public school that existed at baseline in 2008. The remainder of RGCs either only obtained a public school later, the RGC does not qualify as rural according to the 2010 Census, or the RGC does not coincide with the nearest larger settlement that would host a school.

[59]We also hold a 100% sample of the earlier Population and Housing Census (PHC) 2000, but geo-identification beyond the level of the ward (i.e. admininstrative level three in Zambia beyond provinces and districts) does not appear possible due to missing maps of enumeration areas. In this paper, we merely use the PHC 2000 microdata to confirm that wards containing RGCs that were electrified before 2020 do not have statistically different population growth rates than wards containing RGCs that were not yet electrified by 20202, cf. Table 2.

dataset provides the most accurate and holistic source of ground-truth on geo-identified pre-existing conditions across rural Zambian settlements and locations. In our empirical analysis, we test explicitly for the possible role any of these points of interest features, such as mills, churches or commercial entreprises may play in determining the effects of rural electrification in locations with when compared to locations without such features. The dataset contains, for example, 14,191 churches, 13,513 agricultural operations, 8,382 wells, 6,003 commercial operations, 2,360 mills and 1,073 administrative buildings. Figure A13 provides a correlation matrix of different types of points of interest within a two kilometre spatial radius around our relevant sample of 784 inconsequential RGCs. One relevant insight, for example, is that mills are more likely to co-locate in RGCs that also host a commercial operation (such as a shop), a church and a well. It is for this reason that we provide robustness of our results below with respect to any pre-existing points of interest that have empirical support (i.e. churches, farms, commercial, administrative, well), and also highlight results when restricting to subsamples that only feature one of the points of interest, instead of multiple.

Third, the full Population and Housing Census 2010 also includes unique information on each building and structure in Zambia, totalling 2.1m structures in 2010. We use this data source to provide an alternative measure of productive capacity pre-existing at baseline that may be relevant for the effect of electrification in a given locatin. In particular, we focus on the subset of buildings determined by two key variables: occupancy of structure and structure type. We focus on non-residential buildings that are part of commercial operations as an alternative measure of productive capacity and high expected benefits of electrification in a given location: we obtain non-residential buildings from the 'occupancy' variable (which includes residential, non-residential and vacant occupancy), and (part of) commercial buildings from the 'type' variable (which includes, for example, flat, house, commercial, makeshift or mobile types).

4.3 Rich outcome data across rural Zambia

Furthermore, we have access to rich outcome data on outcomes from six geo-identified rounds of the National Labour Force Surveys (LFS) towards the end of our sample in the years 2017-2022, one Living Conditions Measurement Survey (LCMS) towards the middle of our sample in 2015, and various remotely-sensed outcome datasets spanning the entire study period, such as nighttime luminosity (VIIRS, 2012-2020), built-up area (GHSL, 2005/2010/2015/2020), vegetation (NDVI, 2008-2020) and population estimates (WorldPop, 2000/2010/2020) at fine spatial resolution over time.

To generate panel data at the level of rural growth centres, we overlay the respective remotely-sensed outcome variable raster layers with two kilometre buffer polygon areas. This provides balanced panels of observations for nighttime luminosity, built-up area, vegetation or greenery and population estimates for cleanly defined spatial units. Beyond the obvious measure of nighttime luminosity as a measure of electrification and intensity of electricity adoption or usage, built-up area estimates measure construction activity and location growth, while vegetation and greenery indices provide objective measures of agricultural productivity in the vicinity of electrified rural locations. Finally, population estimates provide an alternative measure of in-/out-migration and fertility responses, compared to more accurate census and household survey estimates that do, however, suffer from arguably worse spatial match with the underlying locations of interest.[60]

Given the high frequency of nighttime luminosity and vegetation indices (which we aggregate to annual observations to abstract from issues around seasonality), both outcome measures enable event study-type analyses and provide relevant empirical tests on parallel trends pre-treatment across rural growth centres. Despite their advantages in terms of panel observations and clean spatial identification, these remotely-sensed and processed outcome variables fall short of providing more detailed measures of economic choices and changes on the ground. Therefore, we also construct outcome variables from both the LFS rounds at endline and the LCMS round at midline. In particular, we construct the same variables of interest from the pooled (quarterly) LFS rounds from 2017 to 2022 as those described above for the PHC 2010, and form a long difference (or: two-period panel) dataset merging the subset of enumeration areas included in the LFS rounds to the same locations that were enumerated as part of the PHC 2010. To match this panel of enumeration areas to rural growth centres, we rely on a spatial intersection of RGC coordinates with PHC 2010 enumeration area polygons. Any enumeration area that intersects with a two kilometre buffer around a given RGC will be included in the two-period (2010 vs 2017-2022) panel for our long difference analysis to study the adoption of electrification and its development effects over a long time horizon.

The coverage of all samples over time is provided in Figure 4. As we explain below, we employ this rich suite of data to ask a succession of empirical questions to understand the rural electrification experience, and in particular the potential bimodality in adoption across rural Zambian locations better.

[60] PHC 2010 enumeration areas, for example, do not neatly map into spatial distance buffers around rural growth centre coordinates, since larger RGCs may contain multiple census enumeration areas, whereas smaller ones may share an enumeration area with neighbouring locations (that are either not part of any RGC, or fall into a separate one).

5 Empirical strategy

5.1 Electrification packages and inconsequential locations

Rural electrification is famously expensive due to the long line distances and low electric loads involved.[61] Hence, any scarce funds should be expected to be carefully targeted by policymakers and engineers to maximise expected returns for the utility, or expected economic and/or political benefits for the planner. As discussed in Section 3 above, Zambia's Rural Electrification Master Plan (REMP) is no exception to such a notion of endogenous allocation of rural grid electrification, given that locations are explicitly ranked by their expected demand potential (and, thus, economic return).[62]

We circumvent this problem of endogenous allocation that would most likely introduce bias away from zero in econometric estimates of the effects of grid electrification by exploiting a peculiar feature of the REMP: 784 of the total 1,200 Rural Growth Centres (RGCs) are deemed to be of relatively low priority, but are nonetheless predicted to receive grid access earlier than their own ranking would predict since they lie along the way between the existing grid and highly ranked locations of either economic or political significance.[63] As Figure 6 documents, out of the total of 1,200 RGCs, 180 packages are formed around 180 high-priority RGCs.[64] 784 RGCs are then selected to be electrified as part of the 180 packages, while the remaining 230 RGCs were de-selected for grid electrification, and instead intended to receive off-grid electricity access.[65,66] The planner operationalised this notion of electrification in groups of villages (i.e. 'branches' of the distribution network) by forming so-called electrification packages: bundles of RGCs that are intended to get electrified together, formed of a high-priority RGC as target and other RGCs that fall between said target and an existing electric substation.[67]

---

[61] Rural locations' characteristically low population density, which qualifies them as 'rural' in the first place, makes them more likely: to require higher mileage of medium-voltage transmission lines to be reached, to need higher mileage of low-voltage distribution line per connected end-user, and to have lower cumulative demand than comparably sized urban neighbourhoods, which all inflate cost.

[62] Section 3 highlights the (almost) linear mapping from expected population size in 2030 to expected electricity demand, from which location ranking for electrification timing is derived in the plan.

[63] cf. Appendix A.IV.1 for details on the creation of RGCs, their function and characteristics.

[64] cf. Appendix A.IV.2 for details on the methodology of RGC prioritisation in the REMP.

[65] cf. Appendix A.IV.3 for details on the bundling of RGCs into packages, and package prioritisation.

[66] In practice, we find that the intended off-grid electrification for these 236 RGCs did not materialise as planned: for the 224 off-grid RGCs for which we have verifiable electrification information, 143 RGCs remain unelectrified as of 2020, whereas 28 RGCs had already obtained off-grid access at baseline in 2008 and 22 RGCs had already obtained grid access at baseline in 2008. 31 off-grid RGCs received new grid access by 2020, while no RGC received new off-grid access. Therefore, we also show robustness of results to including off-grid RGCs in the analysis of inconsequential (grid) RGCs below.

[67] REA internally refers to electrification packages as 'electrification projects'. An electric 'substation' commonly denotes the large piece of infrastructure required to step down high- to medium-voltages.

A schematic drawing of this idea is presented in Figure 5: the sketch highlights the existing substation (orange circle), the target location of the given package, i.e. the highest ranked RGC around the substation (red circle), as well as the resulting inconsequential RGCs (blue circles) that are of lower ranking but are accidentally electrified early since they happen to lie on a path between substation and highly ranked RGC. In practice, Figure A14 provides one example of how the resulting setting of project packages can look like: for expositional purposes we plot a subset of packages (numbers 6 to 10) in Zambia's Central Province, highlighting each package's highest ranked RGC (black circle around coloured dot), the inconsequential RGCs in that package (coloured dots), and un-selected RGCs that are allocated to off-grid solutions and exit our sample (triangles). Ranking numbers are displayed as labels next to each RGC, confirming that inconsequential RGCs are of vastly lower priority (i.e. a higher ranking number) than the highest ranked RGC, confirming that they should have indeed only been electrified much later were it nor for their status of lying on a path to a highly ranked RGC. Interestingly, Figure A14 highlights how the electrification roll-out that resulted from the REMP's ranking and package bundling algorithm does approximate the experimental ideal of inconsequentially electrified RGCs in Figure 5 rather well.[68]

In our empirical analysis, we therefore restrict the sample to the 784 inconsequential (non-off-grid) rural growth centres. Since all 1,200 RGCs included in the plan are scheduled to get electrified eventually, the inconsequential RGCs provide the subset of RGCs from which some get 'accidentally' electrified earlier than their rank would dictate since they happen to lie on a path between the baseline grid and a highly ranked location further afield.[69] Table 1 provides a summary of this quasi-experimental variation implicit in the Rural Electrification Master Plan. It also shows that 742 RGCs (out of 784 inconsequential RGCs) can be geo-identified and successfully matched to school-derived ground-truth information on electrification over time: of these, 391 RGCs remain unelectrified, whereas 351 RGCs received grid access by 2020.

Early-treated RGCs are observationally equivalent to not-yet-treated RGCs from the perspective of the original plan: Table 2 shows that the original plan estimated that early vs late inconsequential RGCs have similar levels of households, predicted demand and priority ranking. In baseline household data matched to RGCs from the PHC 2010, however, early-treated RGCs have weakly lower wealth index levels and higher Gini coefficients than not-yet-treated RGCs, although they appear to follow parallel

[68]We produce three dozen maps in the style of Figure A14 for all parts of the country, and confirm throughout that planned package electrification is remarkably close to the quasi-experimental ideal.

[69]cf. Appendix A.IV.4 for further details on the concept of inconsequentially electrified locations.

pre-trends in terms of their overall population growth.

Reassuringly, we show in event study designs below (cf. Section 6) that this notion of parallel trends across early-electrified, inconsequential RGCs and not-yet-electrified, inconsequential RGCs appears to hold in terms of their pre-treatment mean nighttime luminosity and their mean vegetation reflectance levels (both in and around RGCs).[70]

5.2 Triple difference interaction with presence of pivotal end-users

Beyond the staggered, dynamic difference-in-differences analysis, which essentially averages treatment effects over a bimodal distribution of adoption across locations (cf. Section 2.2), our interest revolves more around the determinants of bimodality: as discussed in Sections 2.3 and 3.3, our hypothesis that pivotal end-users may resolve the local public good problem calls for a triple difference design, comparing the difference-in-differences of inconsequential locations electrified by 2020 (before and after) to locations not-yet-electrified by 2020 (before and after) separately for locations with pre-existing pivotal end-users to those locations without. We hypothesise that it is the locations with pivotal end-users, such as mills, present at baseline that drive electrification effects (including adoption, usage and development effects).

Proposed originally by Gruber (1994), but frequently employed in empirical analysis since (Olden & Møen, 2022), our preferred triple difference specification is as follows:

$$Y_{itp} = \beta ActualElect_{itp} + \delta ActualElect_{itp} * PivotalUser_{it_0p} + \gamma_i + \lambda_t + \nu_{ct} + \epsilon_{itp} \tag{1}$$

where i denotes a rural growth centre, t denotes year, c denotes the constituency that contains RGC i and t_0 denotes the baseline year of 2010 in which we measure presence of a pre-existing $PivotalUser$ in location i. γ and λ denote standard location- and time-fixed effects, respectively, whereas ν denotes constituency-year dummies that control for time-varying effects of the relevant parliamentarian's efforts in affecting the release of funds for electrification in their constituency. All standard errors are clustered at the level of the electrification package (denoted by p), where several RGCs are bundled into a single package p that will be electrified together, as discussed in Section 5.1.

In core results below, we report the difference-in-differences estimates next to the triple difference estimates, where we report both the 'level' (i.e. the difference-in-differences effect for locations with a mill), and the triple difference 'interaction' term (i.e. the

[70] Extremum nighttime luminosity and vegetation reflectance such as the maximum reflectance are noisier, and parallel pre-trends hold in some, but not all specifications.

additional effect in the difference-in-differences for locations with a mill, compared to locations without a mill). The identification assumptions for this procedure are weaker than requiring parallel trends in both underlying difference-in-differences: instead we only have to rely on assuming parallel ratios (Olden & Møen, 2022).

Fortunately, the event study plots discussed below highlight for the placebo (or pre-treatment time) periods that for some outcomes and in several specifications, even parallel trends in difference-in-differences across mill status would hold. Reassuringly, parallel ratios appears to hold across the board (cf. Figures 9, 10, A21 and A22).

Our estimation strategy takes three forms, dictated by the available frequency of outcome data: first, in line with the specification in Equation 1 we report standard two-way fixed-effects regressions with interaction terms by pre-existing characteristics for all remotely-sensed outcomes, irrespective of their available temporal frequency. Second, for remotely-sensed outcomes available at high frequency (e.g. annual or even below), we report state-of-the-art dynamic staggered difference-in-difference event studies for both in aggregate and separately for the interaction with pre-existing characteristics and without (i.e. essentially a graphical representation of a triple difference event study).[71] Third, for outcomes derived from households and individuals, we run a long triple-difference between baseline (PHC2010) and a pooled endline (called 'LFS2020', which represents the pooled LFS2017-LFS2022 samples) data, for the subset of locations in which any pooled endline survey was conducted.[72]

The identifying variation of the above triple difference designs is as follows: out of a total of 742 inconsequential RGCs (cf. Table 1), 186 RGCs (25%) contain a mill within two kilometres distance, whereas the remaining 556 RGCs do not. Of these 186 locations with pre-existing mills in 2010, 99 will get electrified by 2020 according to the schools data, whereas 87 remain unelectrified at endline. In contrast, of the 556 locations without a pre-existing mill, 252 locations become electrified by 2020.[73] We present the results of our triple difference empirical strategy below.

[71] Average treatment effects on the treated aggregated over all group-times for these dynamic, staggered event study triple-difference results are available upon request, but generally confirm the (naive) triple difference effects obtained via estimating two-way fixed effects with an interaction term.

[72] No location constraints arise from the PHC2010, for which we hold a 100% sample. In contrast, each LFS provides a cross-section of randomly drawn locations, which we pool to form a single endline.

[73] Hence, among electrified locations, those with mills are marginally overrepresented (28% vs. 22%).

6 Results

Before we focus specifically on how the effects of electrification may differ as a function of pivotal end-users' presence in a given location, we ask how electrification affects locations' outcomes on average.

We can confirm the prevailing findings in the literature (cf. Lee et al. (2020a) for an excellent review) that document zero or mixed effects: in Table 3, columns 1 and 3 report results from a standard two-way fixed-effects estimator. On average, electrification in inconsequential RGCs in Zambia leads to weakly positive nighttime luminosity effects that are marginally significant or insignificant, depending on the measure of electrification. In some sense, this result is reassuring for two reasons: first, since at least one end-user in all of our (electrified) RGCs reports electrification, i.e. the public school itself, a sharp zero effect would seem counterintuitive. Public schools in Zambia usually provide teacher accommodation in the form of small teacher houses on the public school compound, which get electrified irrespective of any location-wide low-voltage trunk network being built or not. Second, the average effect conflates potentially bimodal effects across two common types of locations, which we test for more explicitly below.

6.1 Triple difference by presence of pivotal end-user

We test explicitly for the effects of electrification in the presence of pre-existing pivotal end-users of electricity. Our hypothesis is that pre-existing pivotal users may be crucial to determine the erection of the local trunk network, i.e. the resolution of the local public good problem, leading to widespread electricity adoption and usage, as well as resulting development effects.

Our key result is presented graphically in Figure 7: locations with pre-existing pivotal end-users, mills in our context, show dramatically larger treatment effects of electrification on mean and maximum nighttime luminosity. In fact, these locations display luminosity responses that are three to four times larger than the average response across all locations, which highlights that effects in locations without pre-existing pivotal end-users are consistently null. This stark result points to fundamental differences in the electrification experience across locations with and without a mill, despite both receiving technically the same initial electrification shock (i.e. a transformer being placed next to a public school). Table 3 in columns 2 and 4 provides the same point estimates as Figure 7 in Panel B using the per cent of schools electrified as treatment measure. Panels A and C confirm that the same effect holds with alternative definitions of location electrification status, either as indicator if any school reports grid arrival, or once all schools in a given location report grid arrival, and it survives numerous robustness checks (cf. Section 6.2 below). These results highlight that mills matter greatly for determining the effect of electrification in rural Zambia.

We can repeat the above analysis for several other outcome variables, and can confirm strong and robust effects of electrification in locations with pre-existing mills, and null or insignificant and small effects in locations without. For example, Figure A15 highlights how the effect of electrification on nighttime luminosity appears in a 500 metre buffer around the RGC (i.e. mostly the school as location of choice for the transformer, cf. Section 3.3), and we cannot reject that the initial lighting up of the area around the school differs between locations with and without mills. However, nighttime luminosity responses in one and two kilometres distance from the RGC centroid are starkly positive in locations with mills, but insignificant and/or very small in those without.[74]

An even starker finding emerges for built-up area in Figure A16: the effect of electrification on the share of a location's area covered in built structures is positive and statistically significantly different from zero within one, two and three kilometres radius around RGCs with a mill, but insignificant throughout around RGCs without one. A complementary finding is provided in Figure A17, namely that mean vegetation reflectance in RGCs with a mill is weakly negative within a one kilometre radius, but positive and significant in the outer distance 'halo' rings around the settlement, for example from one to three and three to five kilometres. One interpretation of these results is that, first, agricultural activity in centre of the location is replaced with built-up structures, whereas the outskirts of electrified locations with mills see more intense, or higher-productivity agricultural activity.[75] Since we suspect the mill to adopt electricity, and given the cost advantage of mills powered by electricity, increases in agricultural productivity appear plausible.

Using an imperfect population proxy that has the advantage of being consistently geo-identified to the same underlying area, we can also confirm that successful electrification may halt out-migration from rural areas (Fried & Lagakos, 2017): while electrified locations with pre-existing mills see no significant change in their measure of mean population (i.e. population per two-kilometre-radius around RGC area), locations without

[74]As we discuss in Section 6.2 below, a two-kilometre radius around the RGC centroid (i.e. the school) most likely covers a large part if not the entirety of the RGCs spatial extent.

[75]An alternative interpretation of the latter result would be that the cutting of trees for use as sources of energy is reduced, cf. Table 4 which highlights switches out of wood as source of cooking.

mills do experience statistically significantly less population.[76,77]

For the subset of remotely-sensed outcomes that are available at high frequency (e.g. annual or higher), we report state-of-the-art Chaisemartin and D'Haultfoeuille (forthcoming) dynamic staggered difference-in-difference event studies, both for aggregate difference-in-differences, and separately for the interaction with and without pre-existing characteristics.[78] Figure 9 confirms that aggregate dynamic mean and maximum nighttime luminosity effects are small or insignificant (upper panel). In contrast, locations with pre-existing mills are positive and marginally statistically significant, compared to insignificant effects over time for those locations without. The corresponding event study for vegetation greenery in the area surrounding the electrified location is presented in Figure 10, which highlights that any positive (mean) vegetation effects around electrified villages are purely driven by electrified locations with mills. Interestingly, this result also highlights the long time-of-exposure required for some effects to materialise (Collinson et al., 2024): only in year five after grid arrival does the surrounding area become statistically significantly greener. Hence, given the short post-period for most of our sample, locations with mills may still not yet have exhausted the adoption and usage of electricity despite a local low-voltage trunk network clearly being present early on (cf. Figure 9 discussed above).

Finally, we present results of the long difference analysis comparing locations' electrification experience over approximately one decade from baseline (PHC2010) to endline (LFS2020). This analysis allows us to investigate outcomes (aggregated to the RGC-level) derived from individual- or household-level responses. Table 4 and Figure 8 confirm, first, that adoption and usage of electric power are confined to electrified locations with mills; second, that any development effects of electrification are likewise confined to such locations; third, that the effects are highly gendered, with structural transformation among women driving effects, in line with existing literature on rural electrification (Dinkelman, 2011; Dinkelman & Ngai, 2022). For example, electrified locations with mills are 16% more likely to adopt electric lighting, are 11% less likely to use solar

[76] The remotely-sensed population proxy employed here, WorldPop raster information for 2010, 2015 and 2020, avoids the potential confounding effects of being unable to estimate 'population per location' from the PHC2010-LFS2020 long difference, since not all enumeration areas intersecting with the spatial extent of the RGC would be sampled. Thus, some RGCs would have their entire location covered in the long-difference analysis, others none, and some only parts.

[77] Although confirmed in other contexts, such as rural Ethiopia, this triple difference finding is conflating several effects: in electrified locations without mills, the school receives electricity which increases total enrolment (Figueiredo Walter & Moneke, 2024), suggesting some in-migration irrespective of the local low-voltage network being erected; however, electrification may be expected to reduce fertility.

[78] This setup provides a graphical representation of a triple difference event study. It also adheres to the presentation of our core result in Table 3. Upper (lower) panels correspond to columns 1/3 (2/4).

lamps as alternative to the grid, are 6% more likely to adopt electric cooking and 12% less likely (albeit insignificantly) to cook using wood. In terms of broader development effects, electrified locations with mills see large structural transformation responses: individuals report large drops in agricultural employment, small, insignificant increases in manufacturing employment, and large, statistically significant increases in services employment. These results are mostly driven by females, for whom they are also more precise (cf. Table 4, columns 7-9). Another development proxy commonly studied in the electrification literature (Lipscomb et al., 2013; Lee et al., 2020b) is educational attainment: although relative enrolment overall does not change, the shares of the population that are literate (+14%) or that have completed primary education (+2%) show stark positive effects in electrified locations with mills, pointing towards a mixture of improved educational attainment and skill-biased in-migration.

It is important to note that, throughout our long-difference analysis, locations without mills do not appear to experience significantly positive effects of electrification. In other words, any development effects appear confined to those locations with pivotal end-users present at baseline, which also see dramatically higher adoption and usage of electricity beyond the initial transformer. We interpret these results as being consistent with fundamental differences in the electrification experience across these two types of villages. We conclude that the local public good problem around the erection of the local low-voltage trunk network appears to be the most likely explanation. However, we provide more evidence on the exact mechanisms which support this interpretation in Section 6.3 below. Further evidence confirming the robustness of our core findings is presented below.

6.2 Robustness of triple difference results

Our core triple difference results are highly robust to various permutations of the empirical analysis, including alternative treatment variable definitions, alternative sample restrictions, alternative controls or alternative measures for pivotal end-users.

In particular, we show that our main result in Table 3, Panel B, is robust to alternative definitions of the treatment variable, either as indicator for any school reporting grid arrival (Panel A), or a more extreme indicator if all schools in a given RGC are reporting grid arrival (Panel C).

Likewise, we can construct the panel of schools that inform RGC-level electrification status in three different ways: as balanced panel of all schools that existed at baseline in 2008 (preferred specification), as balanced panel for all schools that exist in 2020, with electrification values back-filled as zero for those school-years in which the school does not report (i.e. did not exist yet), or as unbalanced panel of all schools that ever exist. Since school construction may respond endogenously to electrification (or more likely: school and grid construction occur at the same time) during our study period, we prefer the most conservative balanced panel of only pre-existing schools at baseline. Table A4 confirms that results are qualitatively and quantiatively unchanged with less strict school panels underlying the electrification measure.

A recurring issue in the analysis of place-based policy is uncertainty about the exact spatial extent of each RGC. Our approach attempts to remain as agnostic as possible, and let the data inform our choice of a reasonable spatial extent. To this end, we show in Figure A15 for nightlights, Figure A16 for built-up area, Figure A17 for vegetation and Figure A18 for remotely-sensed population how the triple interaction effect slowly peters out across space. Moving from circular buffers of 500 metre distance to the RGC up to a distance of five kilometres in each direction, most outcome variables become statistically insignificant beyond two kilometre distance from the RGC. The recurring pattern across the various outcome variables reassures our choice of two kilometre distance buffers as a reasonable spatial extent of RGCs in rural Zambia.

An alternative complication arises from the fact that the RGCs, as their name suggests, were supposed to be rural locations. Either due to mismeasurement, incomplete data, human oversight or political interference, the REMP proposes several RGCs for electrification that appear urban or peri-urban even at baseline. Our preferred approach restricts the sample of RGCs to only those that had below 1,500 people per square kilometre population density at baseline in 2010, a common classification of 'urban' settlements. However, in Table A5 we confirm robustness to three alternative sample definitions of 'rural' locations: population restrictions based on surrounding ward population, or two other maximum density restrictions.

Similarly to the issue of non-rural RGCs, our main analysis focuses on the subset of inconsequential RGCs, while we do not expect either high-priority ('consequential') RGCs or off-grid RGCs to adhere to the quasi-experimental ideal of inconsequentially-electrified locations. Nonetheless, the sign and magnitude of treatment effects in the full sample is still of interest: in Table A6 we confirm that inclusion of either consequential or off-grid RGCs does not materially affect results. Put differently, the planners do not appear to have predicted economic potential of electrification well, at least in the sense of predicting accurately how pivotal end-users would affect electrification outcomes.[79]

[79] Given the small sample sizes of the pools of consequential and off-grid inconsequential RGCs, we lack the degrees of freedom to perform analysis for each subsample alone using our main specification.

One concern that pertains to our empirical strategy is the possibility that other spatial shocks or treatments are co-determining outcomes since they may share the spatial or temporal signature of rural electrification in Zambia. Moneke (2023) highlights how large infrastructure expansions often arrive bundled, and that complementarities may exist. Therefore, we control for time-varying phenomena such as road access, main road access (as key determinants of market access) and rainfall patterns (as key determinant of agricultural productivity and seasonality). Table A7 confirms that our results are qualitatively unchanged for market access (despite the large drop in sample due to only three time periods being available), and quantitatively unchanged for rainfall.

Our core estimates of dynamic, staggered difference-in-differences results shown above can also be performed in a more agnostic fashion without assuming any non-parametric structure or constraining the sample of pre- and post-treatment group-time observations to be included. We show event study plots of such 'brute force' analysis in Figure A21 for nighttime luminosity, which are also available for all vegetation index event studies. Results are qualitatively unchanged from our preferred event study specifications.

Finally, this paper focuses on pivotal end-users and their potential contribution to resolving a local public good problem in the form of the erection of the local low-voltage trunk network. The REMP engineers themselves pointed out the likely relevance of grain mills as pivotal, and the evidence on mills' load curves confirms suspicions on their disproportionately high expected benefits from electrification. However, the points of interest dataset allows us to agnostically test for dozens of other categories of potential pre-existing characteristics of settlements that may ensure a trunk network gets constructed. In Table A8, we test for interaction effects of alternative points of interest, such as commercial entreprises (e.g. shops, workshops, etc.), large farms, churches or places of worship, administrative buildings and many more.[80] With the exception of administrative buildings, which we discuss in Section 6.3 below, no other of the numerous categories of points that have empirical support across rural Zambia are significantly affecting electrification's effect on nighttime luminosity or other outcomes of interest.

6.3 Mechanism: resolving a local public good problem

Why does electrification work in some places, but not in others, and what is the underlying determinant of bimodal adoption? In this paper, we argue that rural electrification poses a local public good problem in grid-electrified villages due to the high fixed cost

[80]The correlation matrix of points of interest in RGCs represented in Figure A13 provides the full list of point of interest categories.

of last mile infrastructure, the local low-voltage trunk network required to connect end-users. As we show in Section 6.1 above, pivotal end-users with disproportionate expected benefits of electrification, such as grain mills, play a crucial role in ensuring widespread adoption. Although we lack detailed construction information of local poles and overhead lines to explicitly test for the mill solving free-riding and the coordination failure that makes the village fall short of collectively arranging for the local trunk-network to be erected, we present a series of separate pieces of empirical evidence below that confirm individually and collectively that pivotal end-users may indeed drive electrification outcomes via resolution of the underprovision of the local low-voltage trunk-network.

Our evidence confirming this mechanism rests on six separate pieces of evidence. First, we confirm that bimodality in adoption among grid-connected locations is present in Zambia, too, and that it only affects grid-connected locations, not unconnected ones (cf. Figure A7). However, once controlling for pre-existing pivotal end-users (or 'productive capacity') in a given location, the bimodality in adoption among electrified locations disappears almost fully: Figure A8 compares adoption among electrified places with pre-existing pivotal end-users to all other remaining unelectrified and electrified locations, with the former group of pivotal end-users accounting for most of the 'full' adoption mode, whereas the latter group of locations without such pre-existing pivotal end-users account for most of the 'zero' adoption mode.

Second, we highlight that the effect of mills in driving electrification's outcomes follows an inverted U-shape in distance between the mill and the point of grid arrival: mills that are far, but not too far away from the point of grid arrival (i.e. the transformer, mostly located on public school premises) matter most (cf. Figure A6). Locations in which a mill is located around one kilometre from the electric transformer ('the grid') generate some of the largest nighttime luminosity responses following grid arrival. The resulting inverted U-shape in distance between mill and transformer is depicted in Figure A6. That mills located further away from the transformer have stronger effects on electrification than mills directly near the transformer (or those further than two kilometres away, i.e. beyond most settlements' spatial extent) adds another piece of evidence suggestive of the mill being indeed pivotal for a local electric trunk network to emerge.

Third, we find that mills matter most for electrification's outcomes in larger settlements. This finding resonates closely with the underprovision of public goods literature spawned by Olson (1964), who point to coordination being harder to attain in locations with a larger number of agents present. Quadruple interaction effects with RGC population size confirm this insight (results available upon request), and so does inspection of post-electrification scatter plots of mean nighttime luminosity residuals across pre-existing RGC statistics (such as maximum load, population density, population or electricity adoption) by RGC's pre-existing mill status: Figure A20 confirms that residuals for electrified locations by pre-existing mill status (red dots in upper-right panel) have more positive residual values in higher population density RGCs, while no such increasing underprediction of the effects of electrification as population density increases appears to exist for places without mills.

Fourth, if pivotal end-users indeed help overcome the underprovision of the local low-voltage trunk network, one may expect their role to be minimal in locations where underprovision was not an issue to begin with. For example, some locations may warrant public provision of a trunk network, ensuring its provision irrespective of pivotal users. Indeed, Figure A8 shows that RGCs with public administration offices present at baseline also see positive electrification effects.[81] In line with REA's mandate to connect all public infrastructure (when present), we inspect the possible substitutability of mills with alternative public provision of trunk networks in places that host important government buildings. Table A9 neatly confirms that places without a pre-existing mill see large effects due to a pre-existing administrative building, whereas places without administrative buildings see large effects due to mills.[82]

Fifth, beyond public administration buildings, of the numerous pre-existing features included in the points of interest database, the only other feature that seems to work in the same way as mills (and does also matter in their absence) is an indicator for the resolution of another local public good problem: locations with sewage systems appear to see large electrification effects (results available upon request). It appears reassuring that communities that solved one public good provision problem, a local sewerage system, would also manage to solve another, a local electric trunk network.[83]

Sixth, the most compelling suggestive evidence is provided by ZESCO, the utility in charge of building the local low-voltage trunk network, itself: ZESCO offers large end-users that seek a grid connection two possible connection modi, one for an individual 'dedicated' connection and one for a group scheme connection. Whereas the former quotes the full fixed cost of setting up all required infrastructure between the transformer and the large end-user (i.e. a mini-trunk network for this single user), the group scheme offers a highly subsidised pro rata connection cost based on a full trunk network

[81] A similar result holds for the very few RGCs that host a traditinal chief's palace or a court house.

[82] Analogously, we can compare RGCs with only a single school to those with two schools, where the latter may warrant a trunk network to be publicly provided to connect both schools. We confirm that mills are strongly significant in the former type of location, but lose their explanatory power in the latter.

[83] Further suggestive evidence points to market places and clusters of commercial activity to function as similarly pivotal end-users as mills. Among types of mills, it is specifically electrifiable hammer mills, less so sawing mills (usually located outside of villages) or traditional grain mills, that drive effects.

costing for the entire location (assuming eventual full adoption), minus a generous 30% subsidy, divided by the total number of group scheme applicants. The group scheme, however, mandates the individual applicant to coordinate the formation of the group. In other words, an interested miller would be enticed by the ZESCO district engineer handling their case to find a minimum number of additional end-users in their location to be connected as a group, with dramatically high-powered incentives for the miller to do so. For example, recent cost quotes for a miller in northern Zambia offered an individual connection at a total cost of ZMW200,000, whereas the miller's cost share of the group scheme, subject to enticing enough neighbours to sign up, was ZMW5,000. Such extremely high-powered incentives add a design feature that ensures the miller as 'pivotal' in the original sense of being instrumental in resolving the underprovision of the local low-voltage network. We discuss implications of the assembled suite of suggestive evidence on the underlying mechanism and further discussion of the utility's particular scheme below.

6.4 Discussion

Our findings highlight the need to understand the origins of pre-existing pivotal end-users (in the form of farmers investing in a grain mill) better. However, conditional on such pivotal agents to be present in a rural location, what could explain the stark heterogeneity in results? The existing literature on the underprovision of local public goods in similar contexts points towards coordination failure and free-riding, e.g. in dyke maintenance (Cornes, 1993), irrigation systems (Bardhan, 2000; Bardhan & Dayton-Johnson, 2002b), or common property resources more generally (Bardhan et al., 2007).

In fact, ample anecdotal evidence from the Zambian utility in charge of electricity distribution, ZESCO, and several farmers, millers and school headmasters (the latter being, in most villages, the most educated and knowledgable person) confirms that free-riding in the form of waiting is rampant. To overcome underprovision and free-riding, the utility is operating an incentive scheme to identify and encourage pivotal users (which are profitable end-users for the utility once the local trunk network is in place, but who are loss-making (under finite time horizons) if the utility were to provide the local trunk network) to address and resolve local underprovision and coordination failures.

Similar to the historical precedent from the Rural Electrification Administration in the United States between 1935 to 1960 where the formation of local rural farmer co-operatives was encouraged and incentivised with subsidised federal loans (REA, 1960, 1981), ZESCO offers rural millers (and similarly pivotal end-users) who request a quote for a local connection two separate cost quotes: an individual contract for a dedicated line (with minimal trunk network to bridge the distance between transformer and mill), and a group-scheme contract with equal cost-sharing of the total cost of a full trunk network, including a generous subsidy.[84] The conditions to be offered a group-scheme vary, but usually involve coordinating local support in the form of a large part of the location's households and other end-users connecting, too.

The implied incentive structure for the pivotal end-user is stark: whereas a dedicated, individual line may cost multiples of annual business income, the group scheme shares cost equally across many end-users. Unless the mill is situated very close to a school or health facility, the cost of the latter is usually an order of magnitude lower, and group schemes for electrification abound in rural Zambia. Our paper highlights, that pivotal end-users such as grain millers are indeed very likely to be the driving force behind overcoming coordination failures and resolving the underprovision of local public goods.

7 Conclusion

This paper provides new evidence on how local public good problems of last-mile infrastructure can be resolved. To this end, we characterise the provision of the local electric trunk network required for last-mile connections as a local public good problem: once erected, connections to the local trunk network are imperfectly excludable and non-payment for electricity is empirically common, both in our context of rural Zambia and elsewhere around the world. Furthermore, local electricity usage is imperfectly rivalrous since the local electric trunk network is designed by expert engineers to meet peak local demand, while global supply is unaffected by local variation in individual consumption.

We then document a new empirical regularity across grid-connected locations in rural Zambia: household connections are bimodal conditional on prior grid arrival – with locations either experiencing near-complete electrification or none at all. This bimodality in last-mile electricity adoption – where the local trunk network is erected in some locations but not others – holds more widely across Sub-Saharan Africa countries.

In line with the theoretical literature on how the underprovision of local public goods can be overcome, we highlight the importance of pivotal end-users for the provision of the local trunk network: household and end-user adoption of electricity is concentrated in villages with grain mills, i.e. pivotal agents with disproportionately high expected benefits of electricity. Finally, we provide evidence that any causal development effects

[84] A 'dedicated' line suggests excludability, which appears questionable in this context, cf. Table A1: almost half of rural enumeration areas report that only some or no households pay for electricity supply.

of rural electrification are concentrated in locations where pivotal agents are present at baseline, and where the local public good problem gets resolved.

Our setting provides insights into the development effects of last-mile electrification: electricity acts as a potent amplifier of pre-existing initial conditions. Whereas grid arrival fails to create such initial conditions, electrification can greatly boost productivity and development in places where they exist already. This insight opens up interesting avenues for future research into the optimal expansion of the grid, the origins of conditions conducive to local public good provision, and welfare implications of widespread failure to overcome underprovision of local public goods in low-income countries.

References

Alesina, A., & La Ferrara, E. (2000). Participation in heterogeneous communities. *The Quarterly journal of economics*, *115*(3), 847–904.

Allcott, H., Collard-Wexler, A., & O'Connell, S. D. (2016). How do electricity shortages affect industry? evidence from India. *The American Economic Review*, *106*(3), 587–624.

Banerjee, A. (2004). Who Is Getting the Public Goods in India? Some Evidence and Some Speculation. In *India's Emerging Economy: Performance and Prospects in the 1990s and Beyond*. The MIT Press.

Banerjee, A., Iyer, L., & Somanathan, R. (2005). History, social divisions, and public goods in rural India. *Journal of the European Economic Association*, *3*(2-3), 639–647.

Banerjee, A., Iyer, L., & Somanathan, R. (2007). Chapter 49: public action for public goods. In T. P. Schultz & J. A. Strauss (Eds.).

Bardhan, P. (2000). Irrigation and cooperation: An empirical analysis of 48 irrigation communities in South India. *Economic development and cultural change*, *48*(4), 847–865.

Bardhan, P., & Dayton-Johnson, J. (2002a). Inequality and conservation on the local commons: A theoretical exercise. *The Economic Journal*, *112*(481), 577–602.

Bardhan, P., & Dayton-Johnson, J. (2002b). Unequal irrigators: Heterogeneity and commons management in large-scale multivariate research. In *The drama of the commons* (pp. 87–112). Washingon, DC: The National Academies Press.

Bardhan, P., Ghatak, M., & Karaivanov, A. (2007). Wealth inequality and collective action. *Journal of Public Economics*, *91*(9), 1843–1874.

Bensch, G., Kluve, J., & Peters, J. (2011). Impacts of rural electrification in Rwanda. *Journal of Development Effectiveness*, *3*(4), 567–588.

Bergstrom, T., Blume, L., & Varian, H. (1986). On the private provision of public goods. *Journal of public economics*, *29*(1), 25–49.

Buera, F., Hopenhayn, H., Shin, Y., & Trachter, N. (2023). Big push in distorted economies. *Washington University in St. Louis mimeo.*

Burgess, R., Greenstone, M., Ryan, N., & Sudarshan, A. (2020a). The consequences of treating electricity as a right. *Journal of Economic Perspectives*, *34*(1), 145–169.

Burgess, R., Greenstone, M., Ryan, N., & Sudarshan, A. (2020b). Demand for electricity on the global electrification frontier. *London School of Economics mimeo.*

Burlig, F., & Preonas, L. (forthcoming). Out of the darkness and into the light? Development effects of rural electrification. *Journal of Political Economy.*

Chaisemartin, C., & D'Haultfoeuille, X. (forthcoming). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics.*

Clarke, E. H. (1971). Multipart pricing of public goods. *Public Choice*, *11*(1), 17–33.

Collinson, M., Dinkelman, T., Twine, W., & Wittenberg, M. (2024). Leaping or creeping up the energy ladder: Electricity diffusion in rural Africa. *Notre Dame U mimeo.*

Cornes, R. (1993). Dyke maintenance and other stories: Some neglected types of public goods. *The Quarterly journal of economics*, *108*(1), 259–271.

Cornes, R., & Sandler, T. (1996). *The theory of externalities, public goods, and club goods*. Cambridge, Cambridge University Press.

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

Dinkelman, T., & Ngai, L. R. (2022). Time use and gender in Africa in times of structural transformation. *Journal of Economic Perspectives*, *36*(1), 57–80.

Figueiredo Walter, T., & Moneke, N. (2024). Complementarities in public infrastructure: Evidence from school electrification in Zambia. *University of Oxford mimeo.*

Fried, S., & Lagakos, D. (2017). Rural electrification, migration and structural transformation: Evidence from Ethiopia. *Carleton College mimeo.*

Grimm, M., Hartwig, R., & Lay, J. (2013). Electricity access and the performance of micro and small enterprises: Evidence from west africa. *The European Journal of Development Research*, *25*(5), 815–829.

Groves, T. (1977). Incentives in teams. *Econometrica*, *41*(4), 617–631.

Groves, T., & Ledyard, J. (1977). Optimal allocation of public goods: A solution to the "free rider" problem. *Econometrica*, *45*(4), 783–809.

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American economic review*, *84*(3), 622–641.

Kassem, D. (2018). Does electrification cause industrial development? Grid expansion and firm turnover in Indonesia. *London School of Economics mimeo*.

Kitchens, C., & Fishback, P. (2015). Flip the switch: The impact of the REA 1935–1940. *The Journal of economic history*, *75*(4), 1161–1195.

Lee, K., Brewer, E., Meyo, F., Miguel, E., Podolsky, M., Rosa, J., & Wolfram, C. (2014). Barriers to electrification for "under grid" households in rural Kenya. *NBER Working Paper Series*.

Lee, K., Brewer, E., Meyo, F., Miguel, E., Podolsky, M., Rosa, J., & Wolfram, C. (2015). Barriers to electrification for "under grid" households in rural Kenya. *Development Engineering*, *1*, 26–35.

Lee, K., Miguel, E., & Wolfram, C. (2020a). Does household electrification supercharge economic development? *Journal of Economic Perspectives*, *34*(1), 122–144.

Lee, K., Miguel, E., & Wolfram, C. (2020b). Experimental evidence on the economics of rural electrification. *Journal of Political Economy*, *128*(4), 1523–1565.

Lewis, J., & 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*.

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

Moneke, N. (2023). Can Big Push infrastructure unlock development? Evidence from Ethiopia. *University of Oxford mimeo*.

Murphy, K., Shleifer, A., & Vishny, R. (1989). Industrialization and the big push. *The Journal of Political Economy*, *97*(5), 1003–1026.

Olden, A., & Møen, J. (2022). The triple difference estimator. *The econometrics journal*, *25*(3), 531–553.

Olson, M. (1964). Agriculture and the depressed areas. *American Journal of Agricultural Economics*, *46*(5), 984–988.

Peters, J., & Vance, C. (2011). Rural electrification and fertility – Evidence from Côte d'Ivoire. *Journal of Development Studies*, *47*(5), 753–766.

Proffen, C. (2024). How does electrification matter for child development? Evidence from Ethiopia. *Goethe-Universität Frankfurt mimeo*.

Redding, S. J., & Turner, M. (2015). Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics*, *5*, 1339–1398.

Rud, J. P. (2012). Electricity provision and industrial development: Evidence from India. *Journal of Development Economics*, *97*(2), 352–367.

Rural Electrification Administration, U. S. (1960). Rural lines, USA: The story of the Rural Electrification Administration's first twenty-five years, 1935-1960.

Rural Electrification Administration, U. S. (1981). Rural lines, USA: The story of cooperative rural electrification, *1962-2022-476*.

Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics*, *36*(4), 387–389.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance (New York)*, *16*(1), 8–37.

Wade, R. (1994). *Village republics : Economic conditions for collective action in South India*. San Francisco, CA, ICS Press.

Young, A. (2012). The African growth miracle. *Journal of Political Economy*, *120*(4), 696–739.

8 Figures

Figure 1: Bimodal household connections in electrified villages

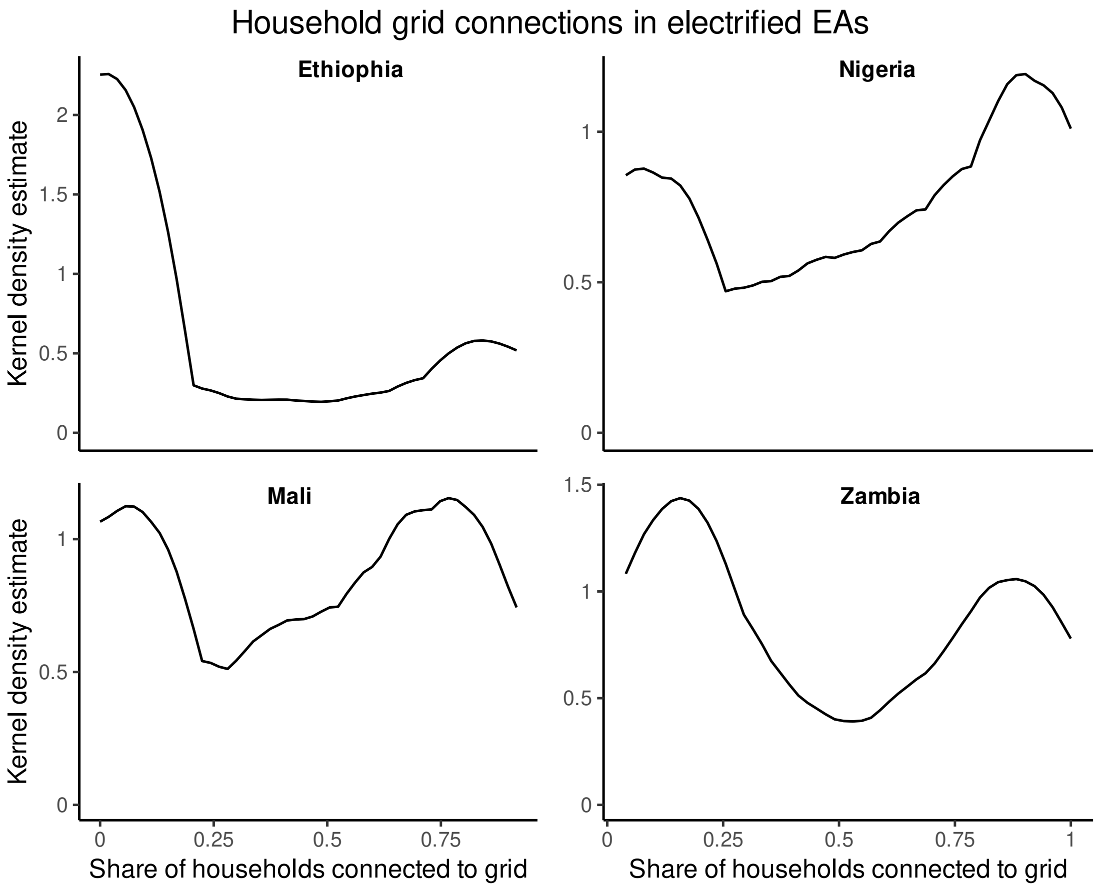


Figure 2: Grain mills as pivotal end-users of electricity

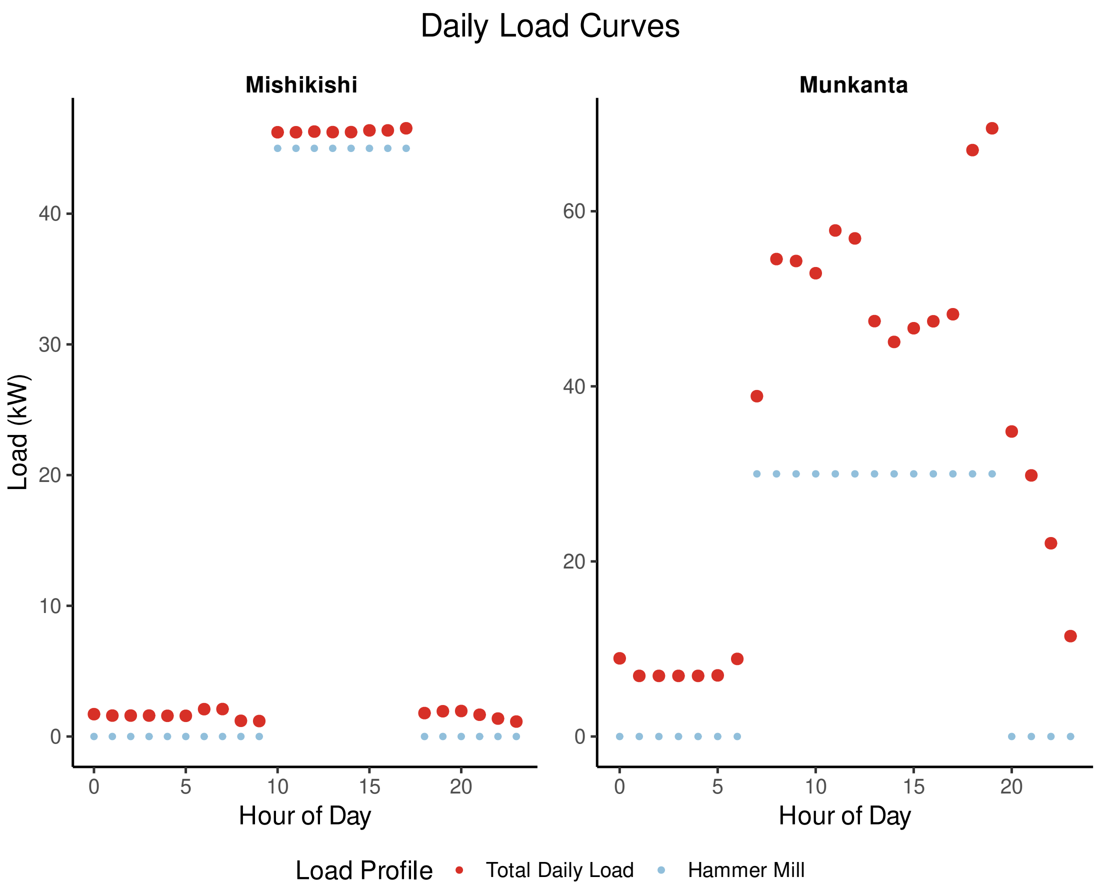
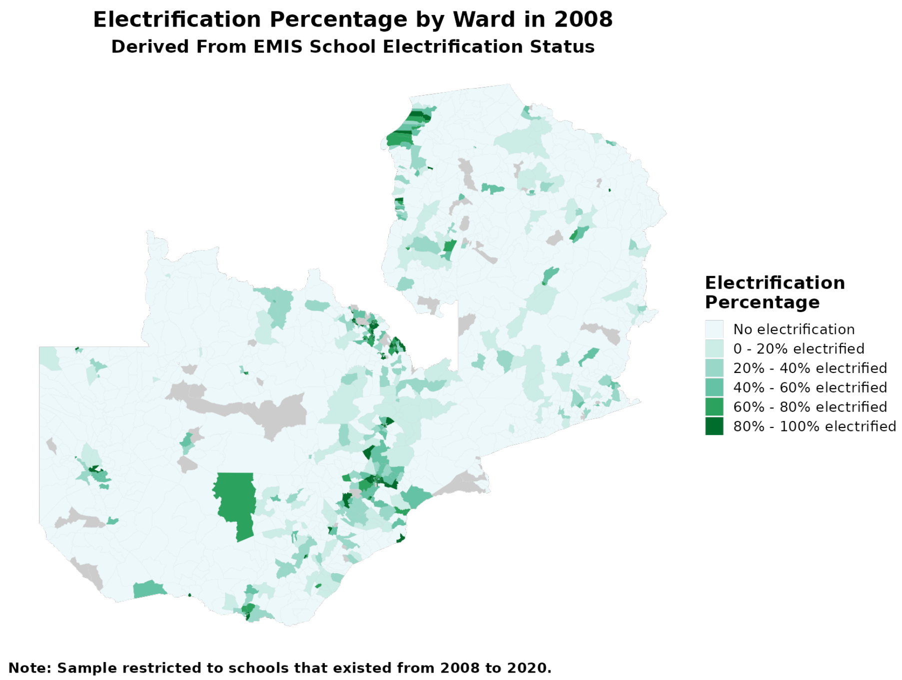


Figure 3: Rural electrification in Zambia: 2008 vs 2020 (derived from schools)



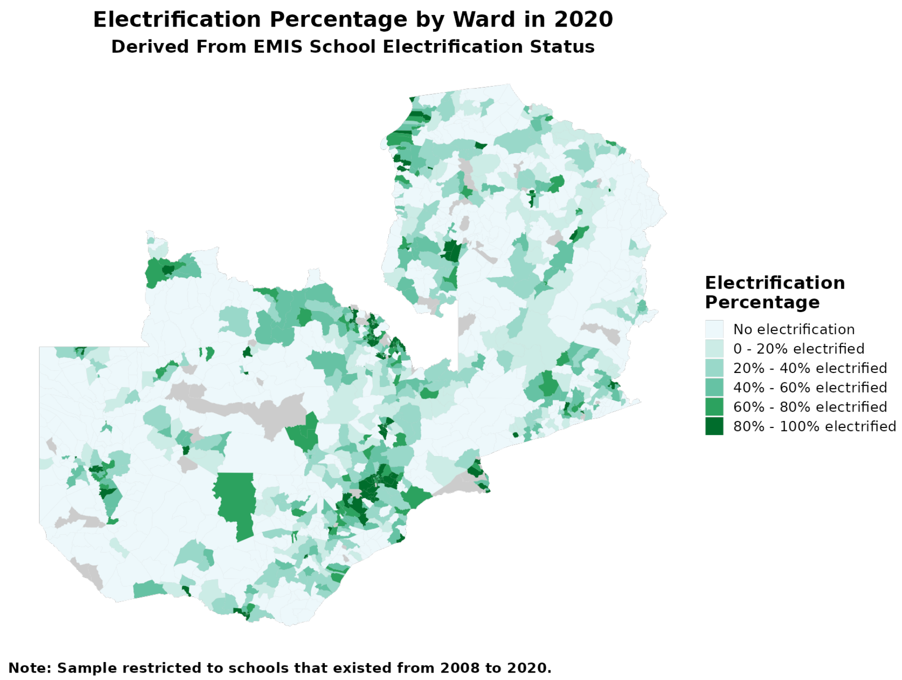


Figure 4: Sample coverage across years and datasets

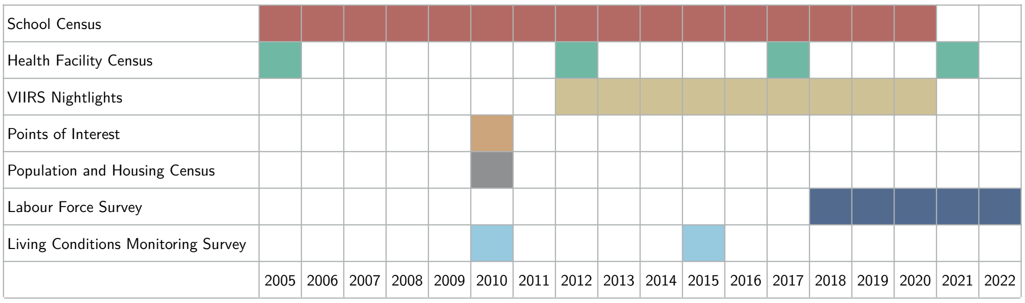


Figure 5: Inconsequential rural growth centres (RGCs) in electrification packages

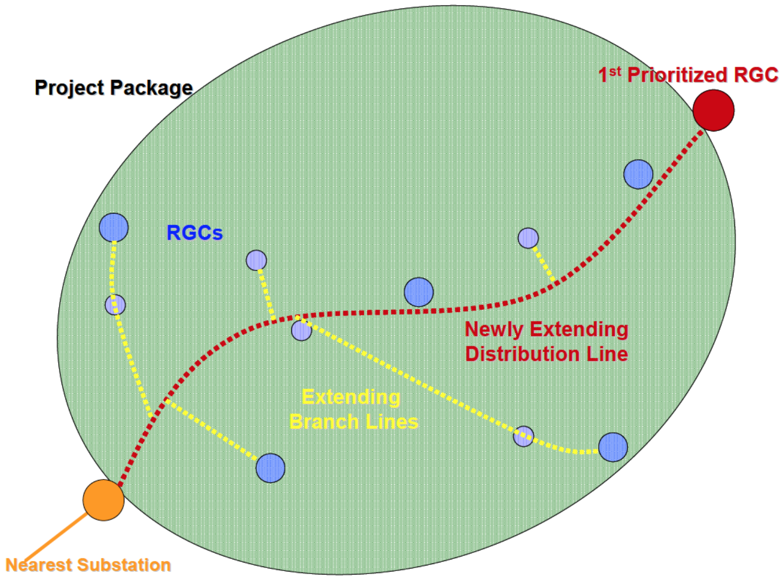


Figure 6: Electrification treatment status across (inconsequential) RGCs by 2020

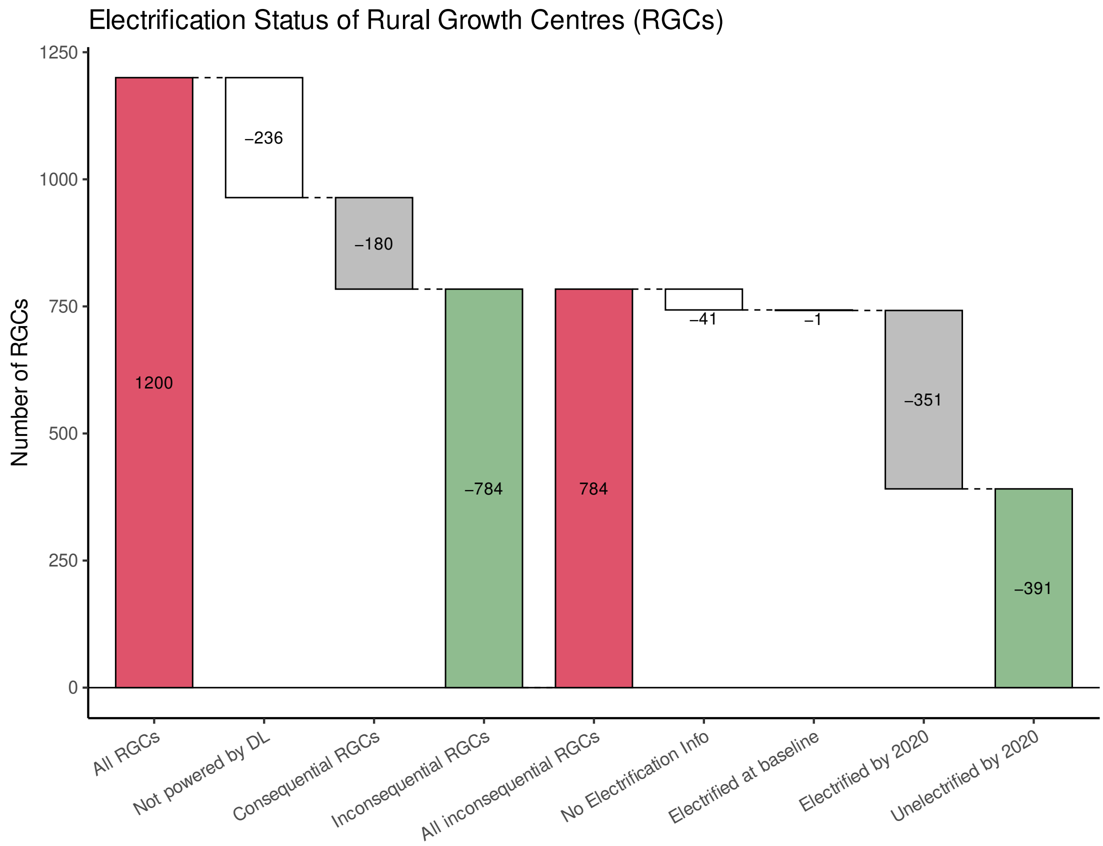


Figure 7: Locations with mills see adoption after grid arrival

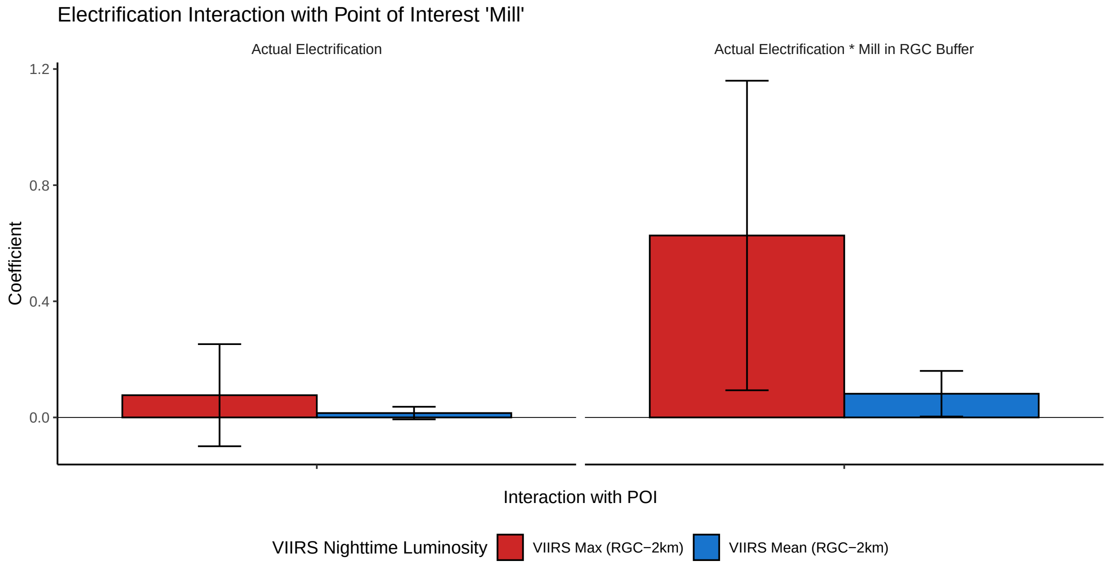


Figure 8: Locations with mills see development after grid arrival

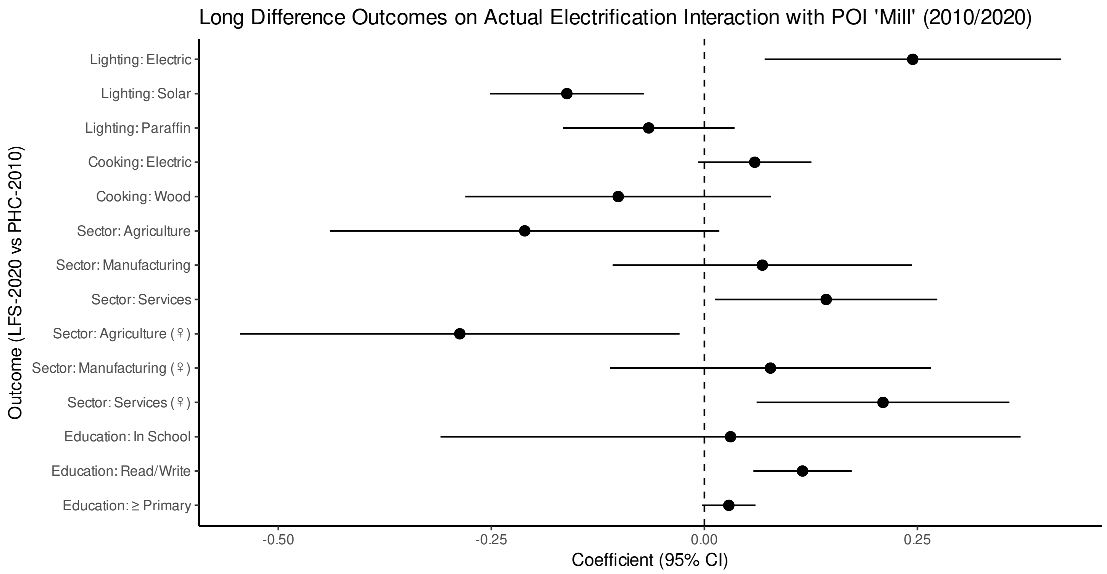


Figure 9: Event studies of nightlights on electrification: overall, with and without mill

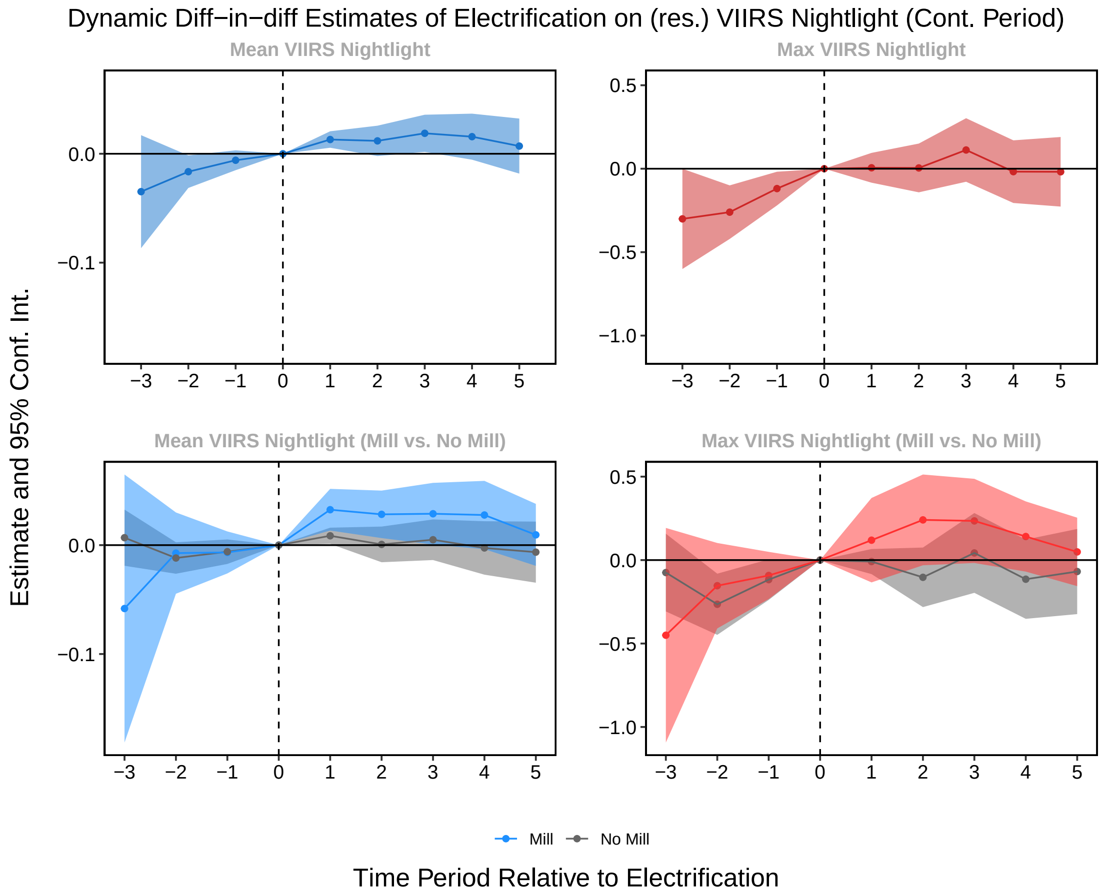
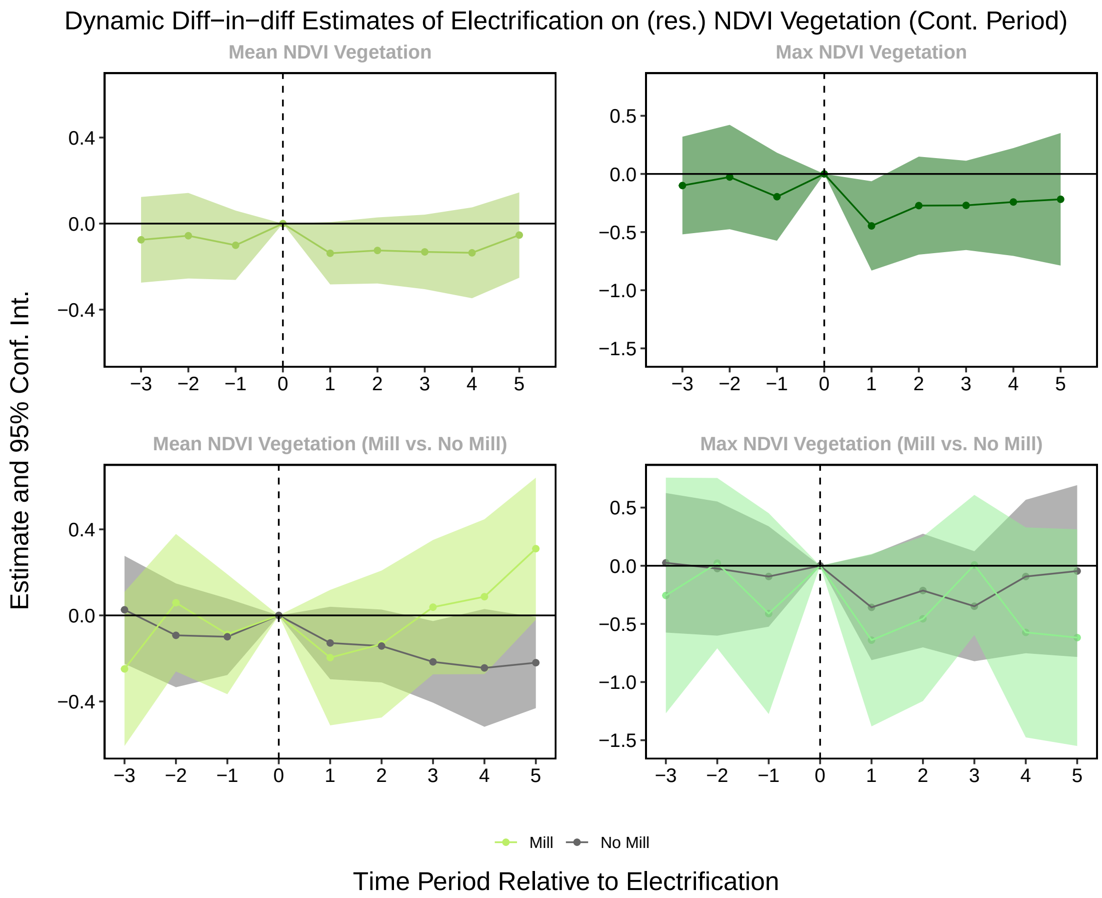


Figure 10: Event studies of vegetation greenery in surrounding area (3km-5km halo ring around RGC) on electrification: overall, with and without mill



9 Tables

Table 1: Treatment Status by Rural Growth Centre (RGC) with School Match

	Not treated by 2020	Treated by 2020	All observations	
Sample	RGC with school	RGC with school	RGC with school	Any RGC
All RGCs	654	481	1,135	1,200
– offgrid	168	56	224	236
– consequential	92	77	169	180
– inconsequential	**391**	**351**	**742**	**784**

Note: RGC with school denotes RGCs that can be geo-identified to actual electrification status from nearby school with high confidence. RGC-level electrification status cannot be derived (with confidence) for subset of RGCs without nearby schools.

Table 2: Summary Statistics at the RGC-level by Treatment

	Not treated by 2020		Treated by 2020		
	Mean	Std. Dev.	Mean	Std. Dev.	Test
Number of households	354	316	383	396	F= 1.064
Predicted demand	405452	347556	439515	436421	F= 1.239
RGC ranking	594	297	592	316	F= 0.006
Wealth index	0.026	0.013	0.024	0.0062	F= 2.881*
Gini coef. (all)	0.43	0.14	0.41	0.12	F= 3.816*

Note: Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01.

Table 3: Nightlights on Actual Electr. and Mill Interaction (2012-2020)

	VIIRS Mean Nightlights		VIIRS Max Nightlights	
	(1)	(2)	(3)	(4)
	Panel A - Electrification Dummy (any school electrified)			
Actual Elect. (any)	0.02*	0.00	0.11	0.02
	(0.01)	(0.01)	(0.07)	(0.07)
Actual Elect. (any) * Mill in RGC		0.04**		0.36**
		(0.02)		(0.14)
	Panel B: Electrification Percentage			
Actual Elect. (pct)	0.04**	0.02	0.23**	0.08
	(0.01)	(0.01)	(0.10)	(0.09)
Actual Elect. (pct) * Mill in RGC		0.08**		0.63**
		(0.04)		(0.27)
	Panel C - Electrification Dummy (all schools electrified)			
Actual Elect. (ind)	0.02	0.00	0.15	0.02
	(0.02)	(0.01)	(0.12)	(0.13)
Actual Elect. (ind) * Mill in RGC		0.06		0.59*
		(0.04)		(0.32)
Year FE	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓
Const-Year trends	✓	✓	✓	✓
R^2	0.96	0.96	0.88	0.88
No. of Years	9	9	9	9
No. of RGCs	742	742	742	742
No. of Const-Years	1215	1215	1215	1215
Obs.	6678	6678	6678	6678

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. RGC-level actual electrification derived from schools existing since 2008. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC.

Table 4: Long Difference Outcomes on Actual Electrification Interaction with POI 'Mill' (2010/2020)

	Light: Elec.	Light: Solar	Cook: Elec.	Cook: Wood	Empl.: Agr.	Empl.: Man.	Empl.: Ser.	Empl.: ♀-Agr.	Empl.: ♀-Ser.	Edu.: Enrol.	Edu.: R/W	Edu.: Prim.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Act. Elect. (pct)	-0.08^{**}	0.05	-0.00	-0.02	0.01	0.00	-0.01	0.02	-0.02	-0.09	0.02	-0.01
	(0.03)	(0.04)	(0.01)	(0.04)	(0.08)	(0.06)	(0.04)	(0.09)	(0.05)	(0.12)	(0.02)	(0.01)
Act. Elect. * Mill	0.24^{***}	-0.16^{***}	0.06^{*}	-0.10	-0.21^{*}	0.07	0.14^{**}	-0.29^{**}	0.21^{***}	0.03	0.12^{***}	0.03^{*}
	(0.09)	(0.05)	(0.03)	(0.09)	(0.12)	(0.09)	(0.07)	(0.13)	(0.07)	(0.17)	(0.03)	(0.02)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Prov-Year trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.83	0.79	0.81	0.92	0.72	0.62	0.79	0.69	0.80	0.56	0.84	0.70
No. of Yrs	2	2	2	2	2	2	2	2	2	2	2	2
No. of RGCs	231	231	231	231	231	231	231	231	231	231	231	231
No. of Prov-Yrs	20	20	20	20	20	20	20	20	20	20	20	20
Obs.	462	462	462	462	453	453	453	452	452	462	458	459

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. RGC-level actual electrification derived from schools existing since 2008. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC centroid. Empl. denotes share of population working in respective sector of economy (Agr. - agriculture, Man. - manufacturing, Ser. - services). Fem. Agr. denotes share of females working in respective sector. Edu.: Enrol. denotes share of population that is enrolled in school at time of enumeration. Edu.: R/W denotes share of population that can read and write a basic sentence (i.e. is literate). Edu.: Prim. denotes share of population that completed at least primary school.

Online Appendix (not for publication)

A.I Appendix: Additional figures

Figure A1: Share of Non-Payment across Enumeration Areas (LCMS-2015)

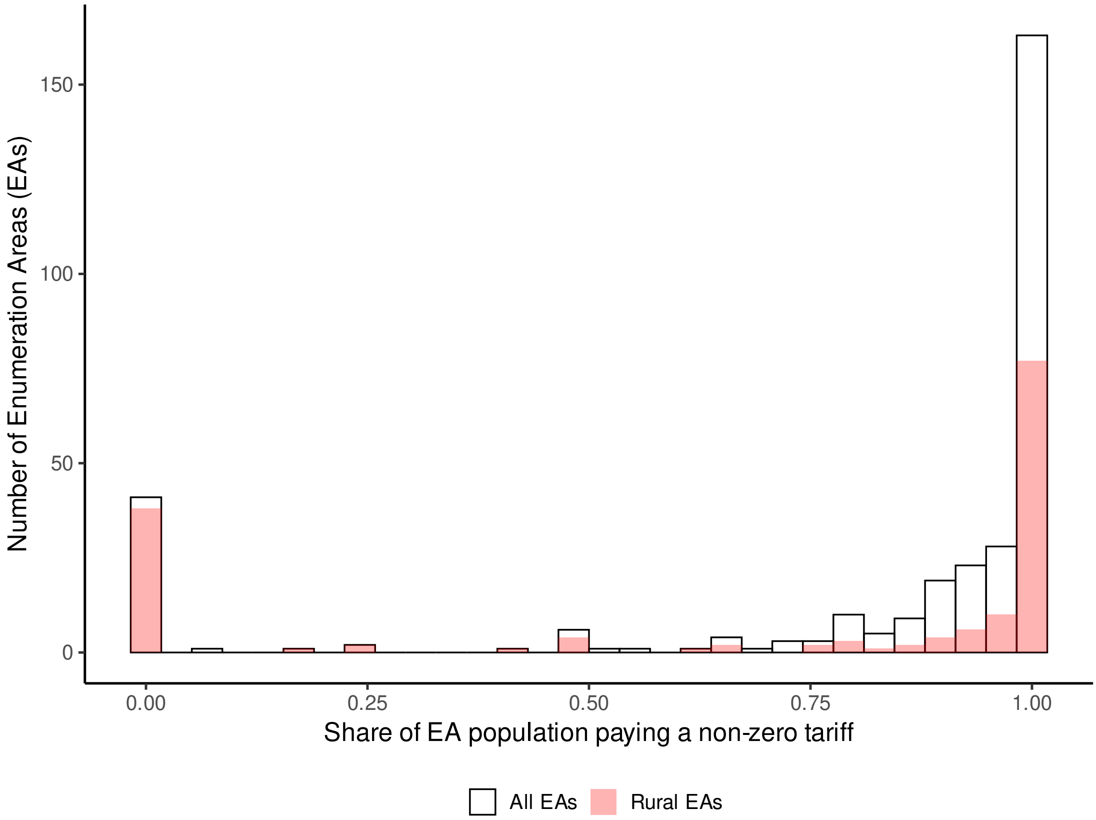


Figure A2: Pictures of local electric trunk network in rural Zambia

(a) Transformer under medium-voltage line, spanning low-voltage trunk network

(b) Low-voltage trunk network pole and end-user connections

Figure A3: Grain mills as pivotal end-users compared to alternative end-uses (REMP)

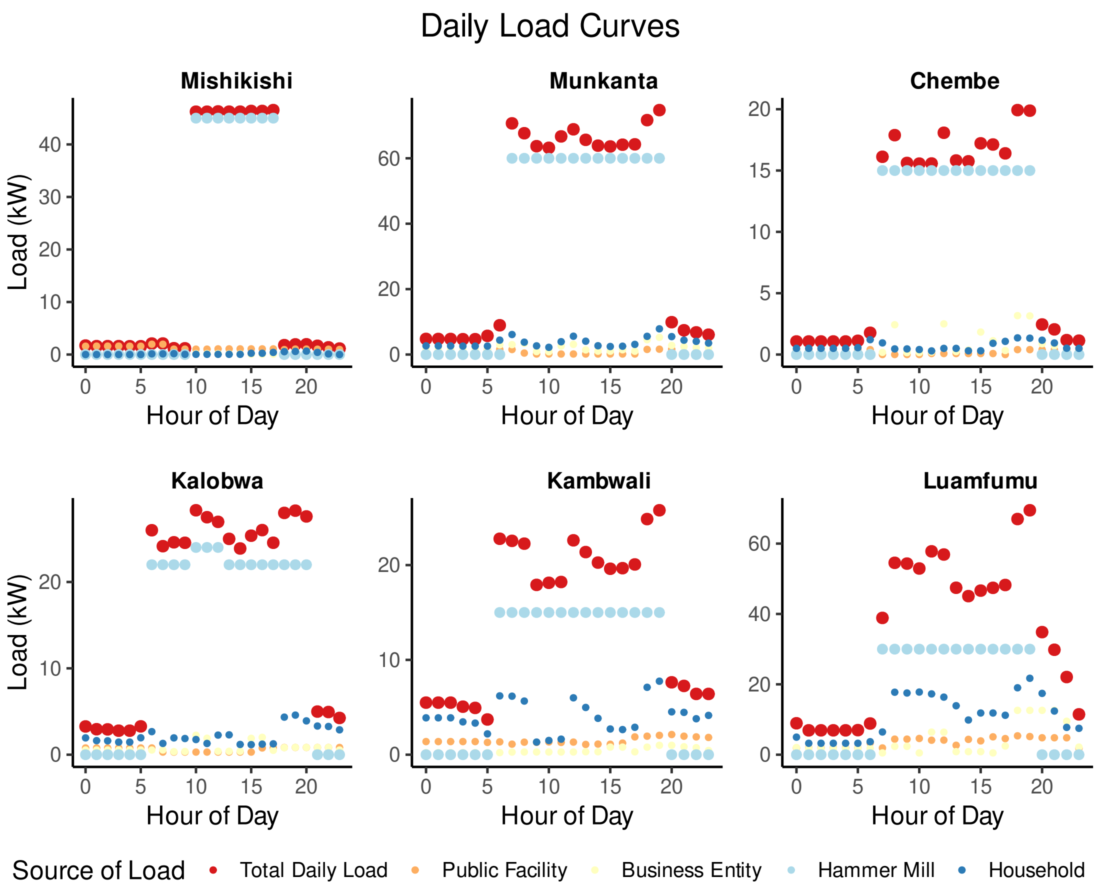


Figure A4: High fixed cost of 'last mile' electrification (REMP)

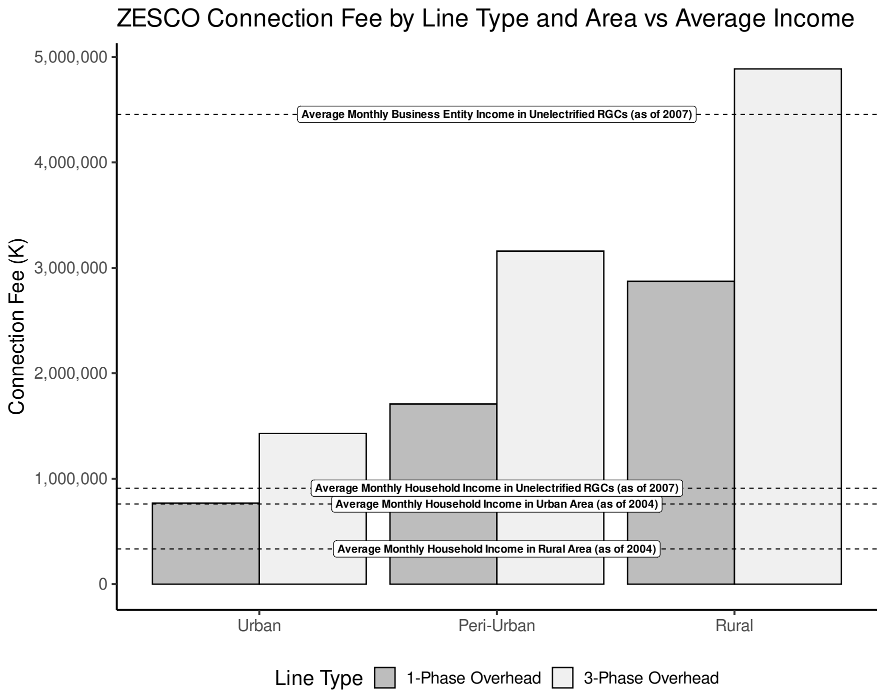


Figure A5: Lee et al.'s (2020b) estimates of downward sloping cost curves in rural Kenya

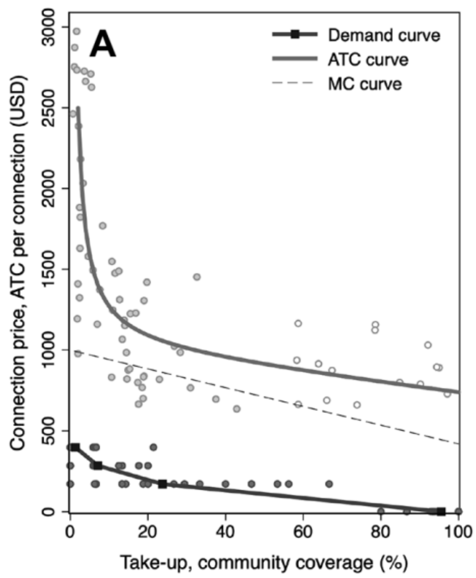
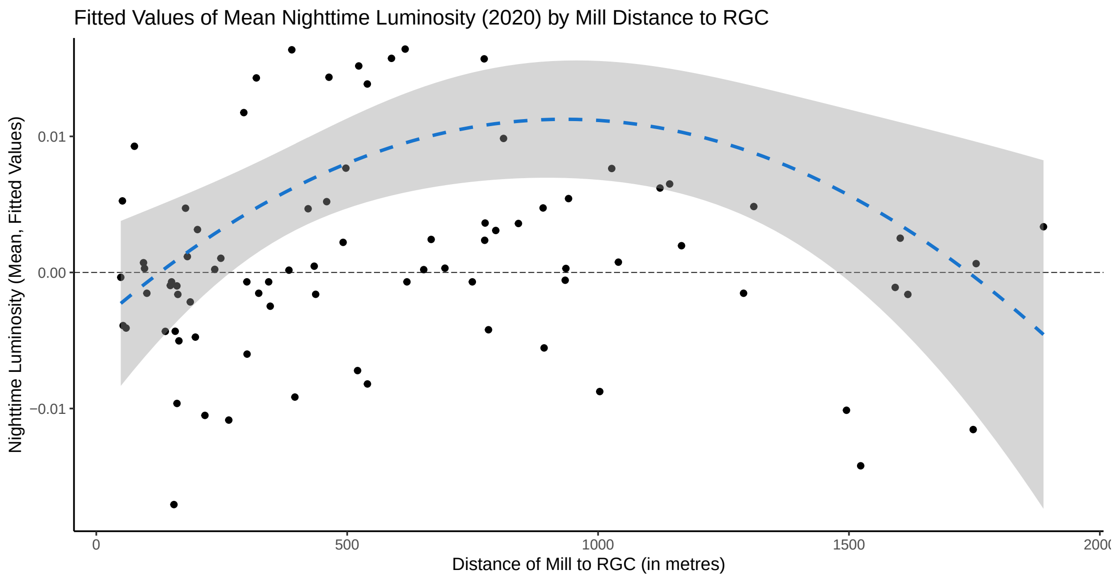


Figure A6: Nighttime luminosity by distance between RGC and mill



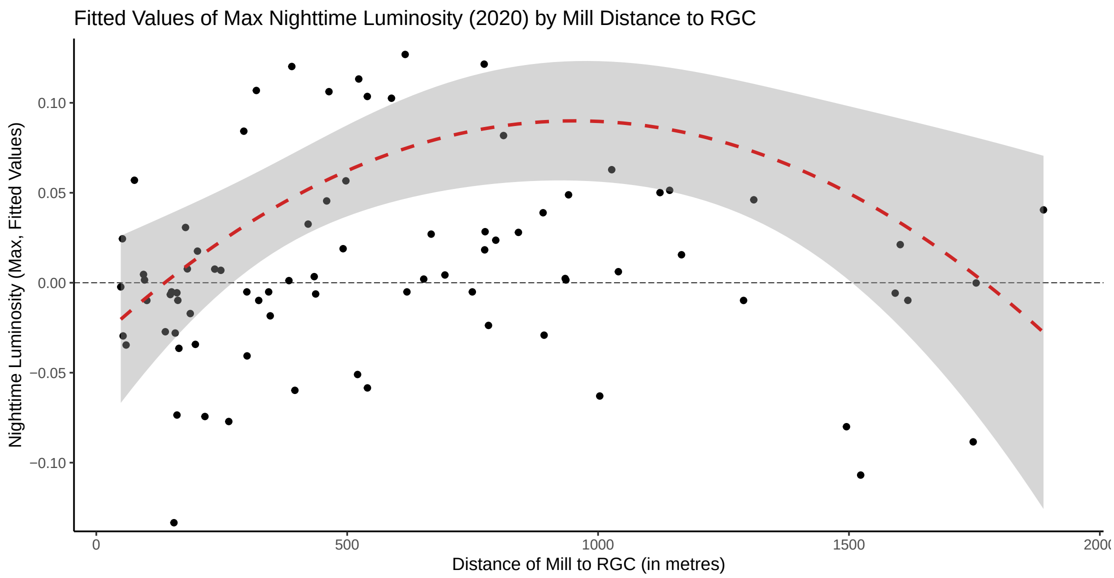
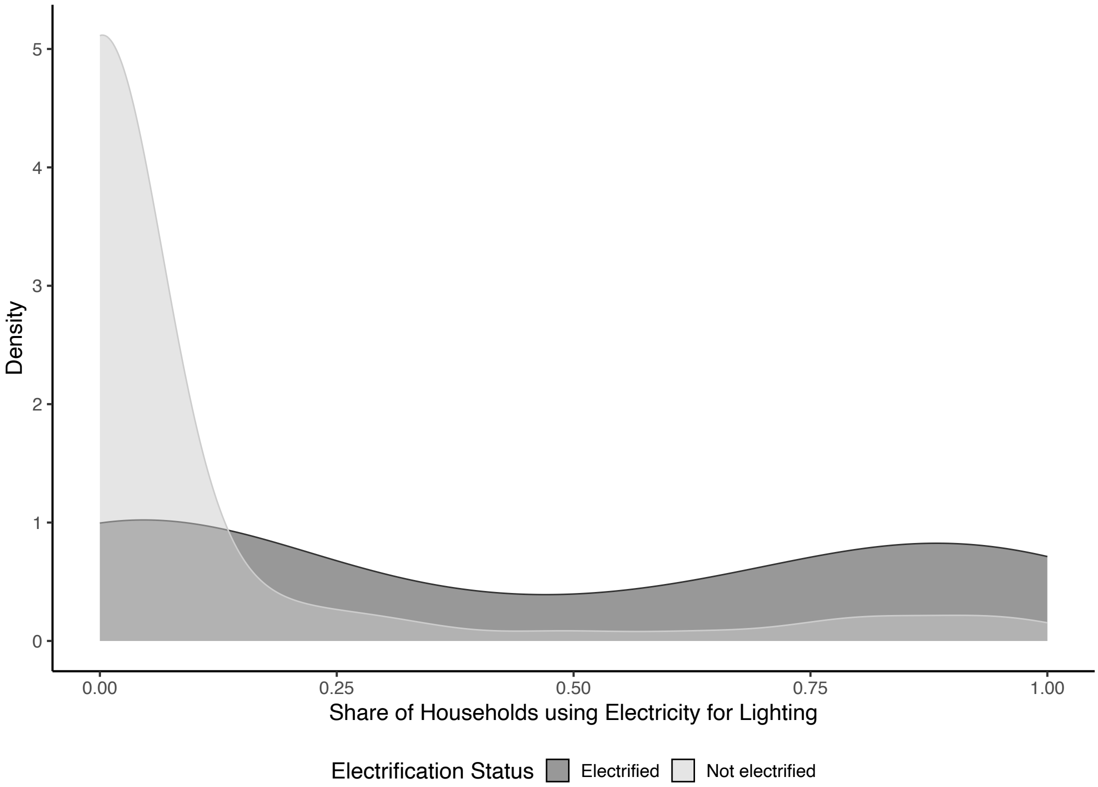
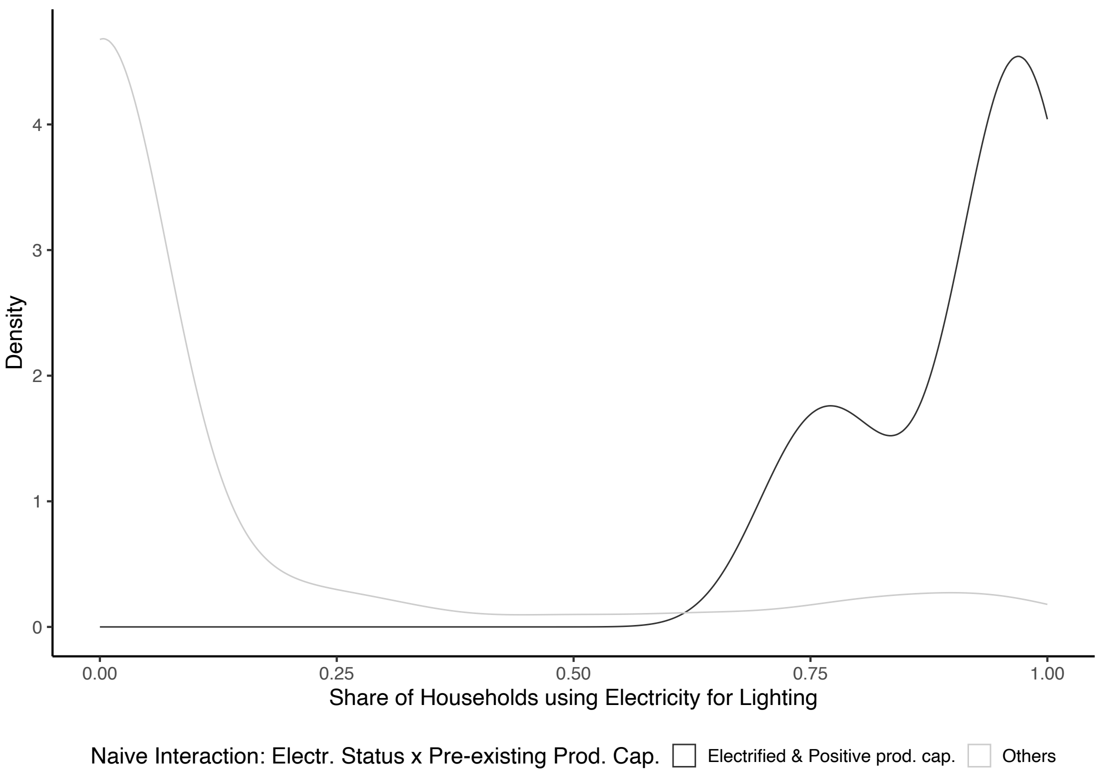


Figure A7: Do households in electrified locations adopt?



Note: Sample restricted to schools that existed from 2008 to 2020. Electrification status is a dummy variable that equals one when at least 80% of schools in ward are electrified.

Figure A8: Do households in electrified locations with productive capacity adopt?



Note: Sample restricted to schools that existed from 2008 to 2020. Productive capacity is a dummy variable that equals one when the share of non-residential buildings is positive in ward. Electrification status is a dummy variable that equals one when at least 80% of schools in ward are electrified.

Figure A9: Electricity access and GDP per capita, cross-country (2016)

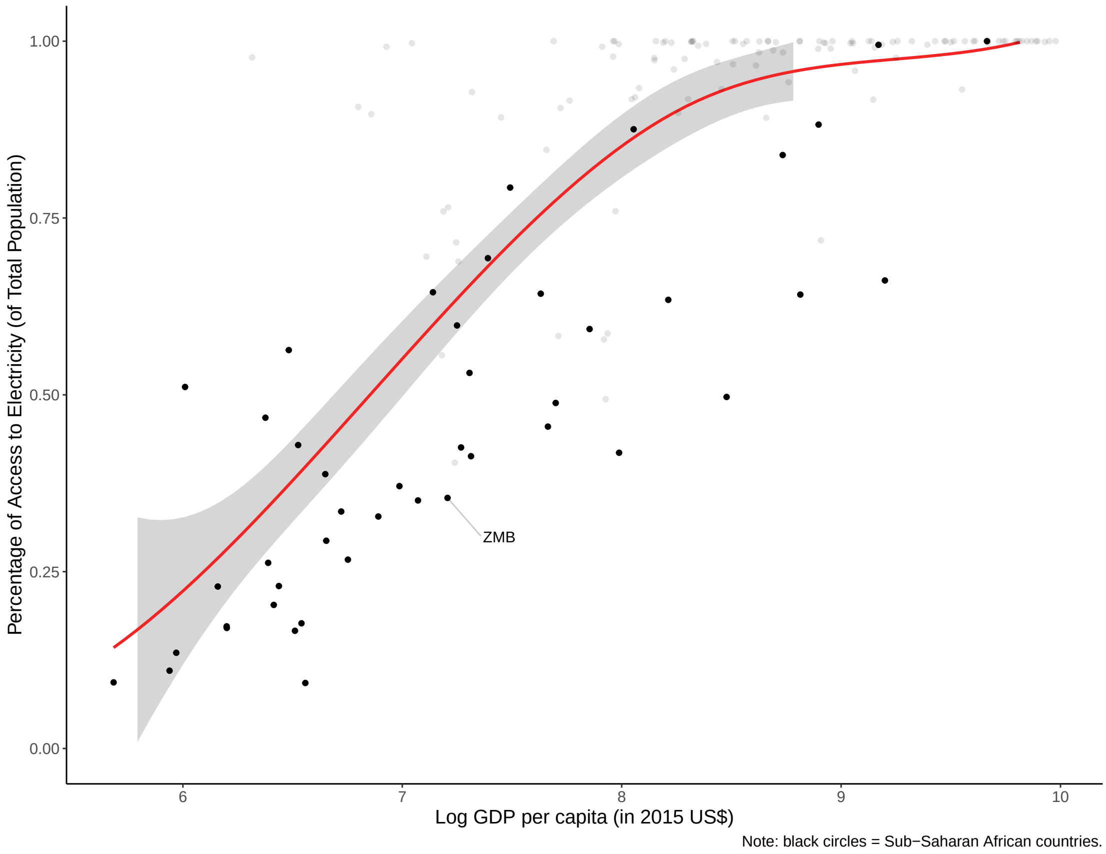


Figure A10: Cross-validation of school- and health facility-derived electrification with ZESCO medium-voltage lines (2023) across RGCs

Figure A11: Rural electrification in Zambia: 2008 vs 2020 (derived from schools), sample of rural wards with inconsequential RGCs

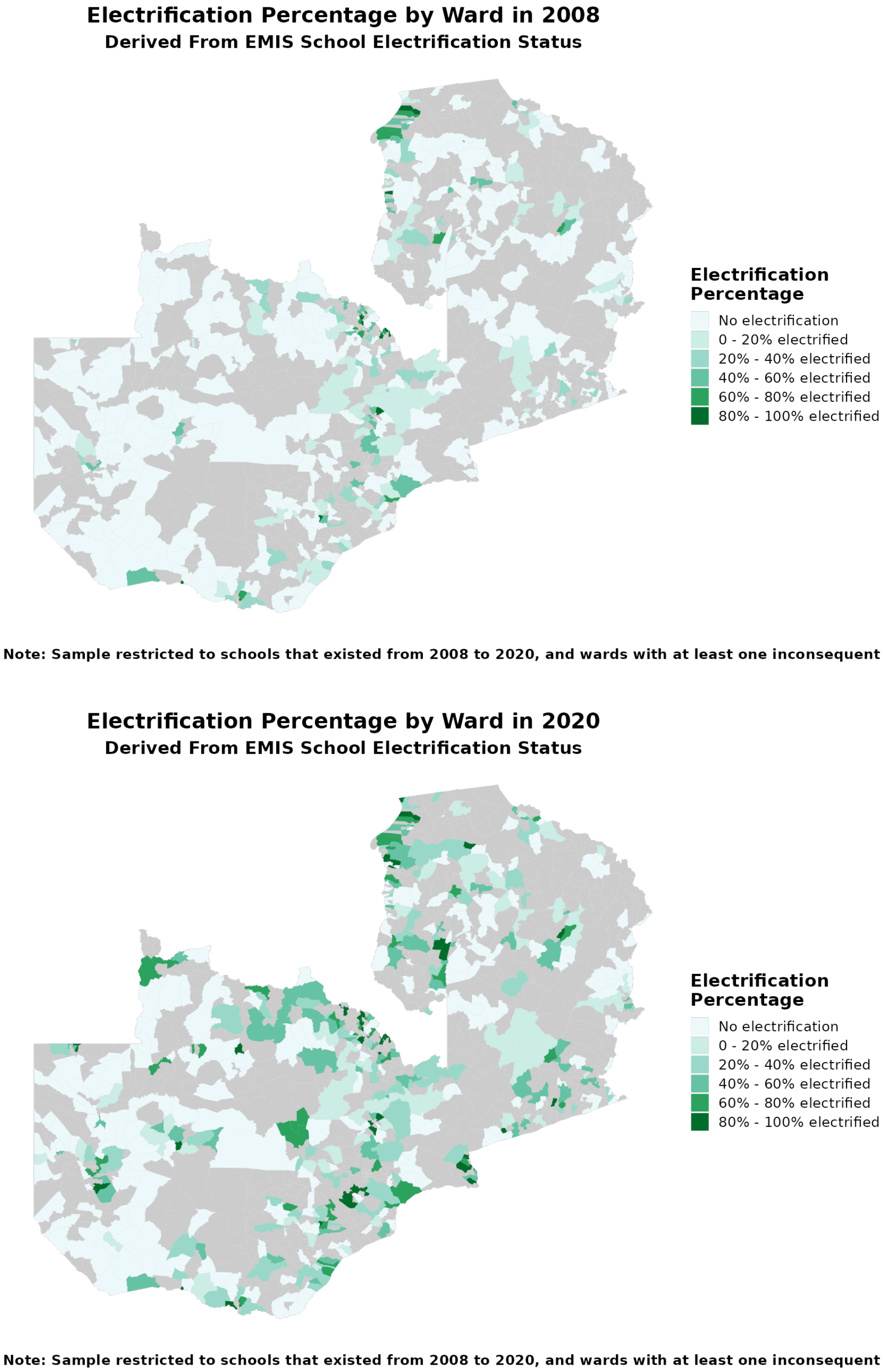


Figure A12: Percentage of electrified RGCs

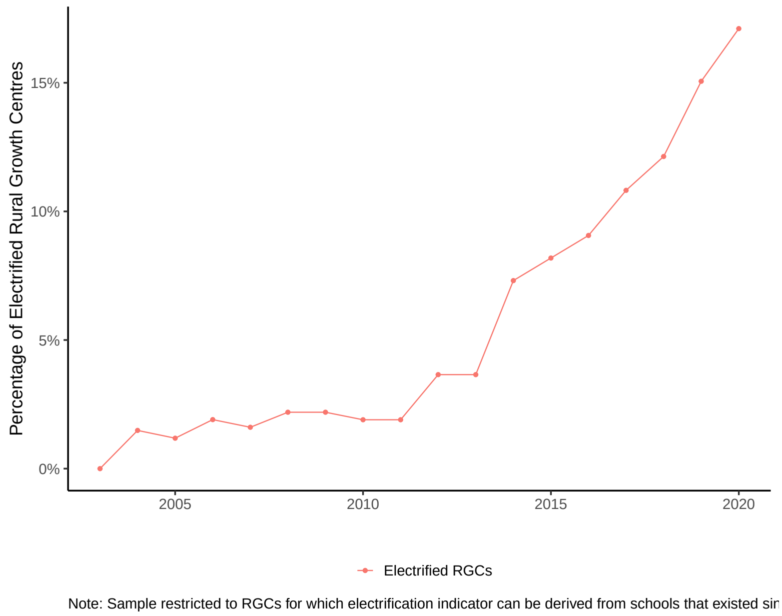


Figure A13: Correlation matrix of Points of Interest (2010) in inconsequential RGCs

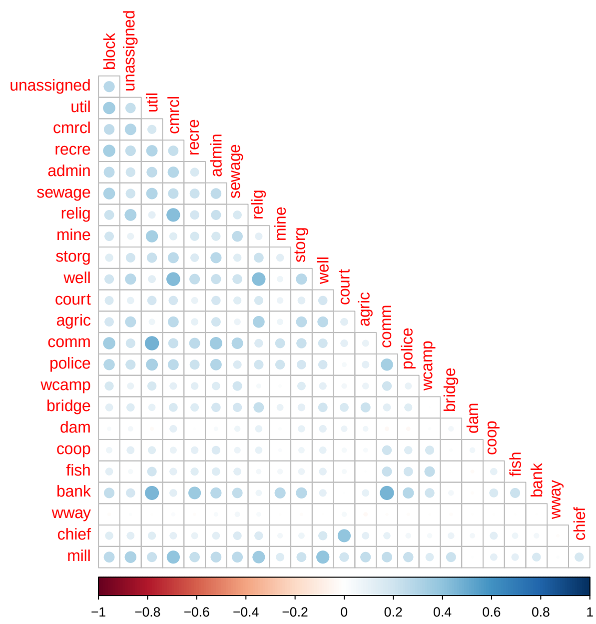


Figure A14: Inconsequential units in practice: packages #21–#25, Eastern Province

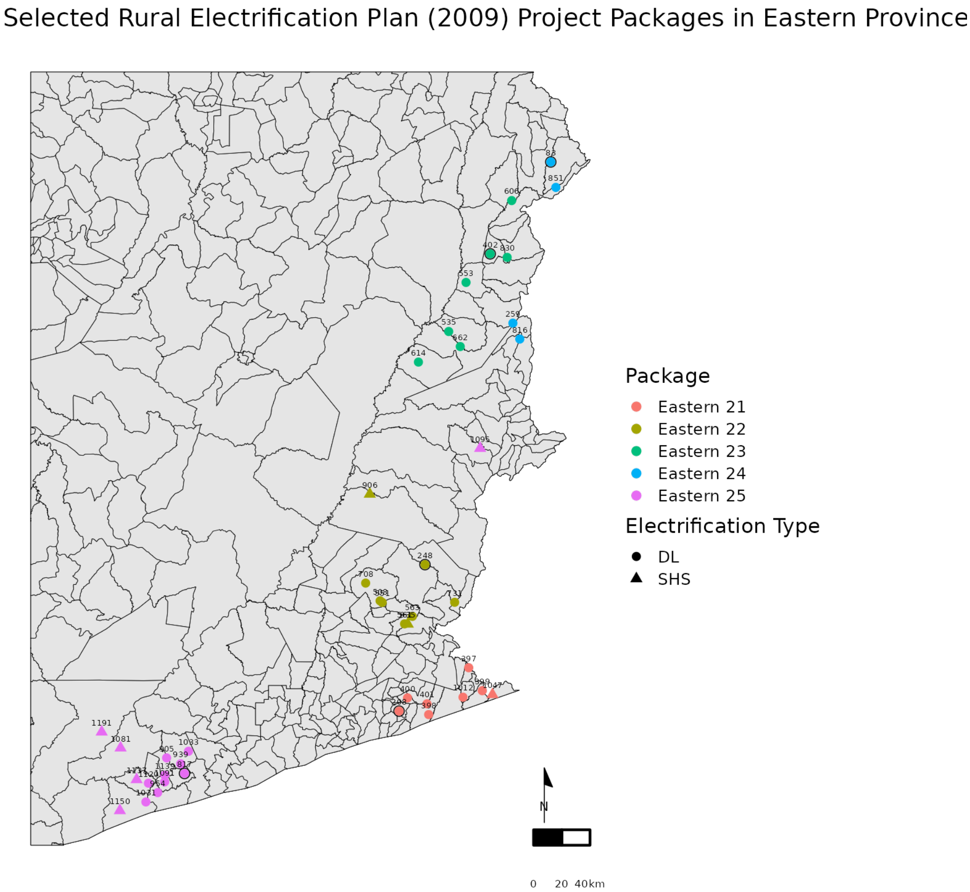
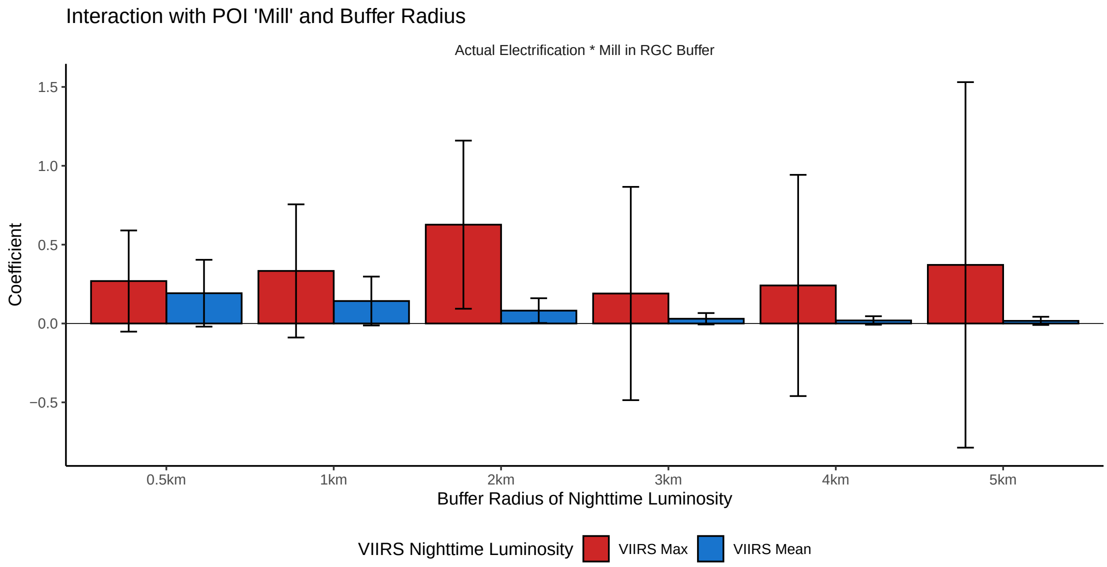


Figure A15: Effect of electrification on nighttime luminosity

(a) Nighttime luminosity in electrified places with mills



(b) Nighttime luminosity in electrified places without mills

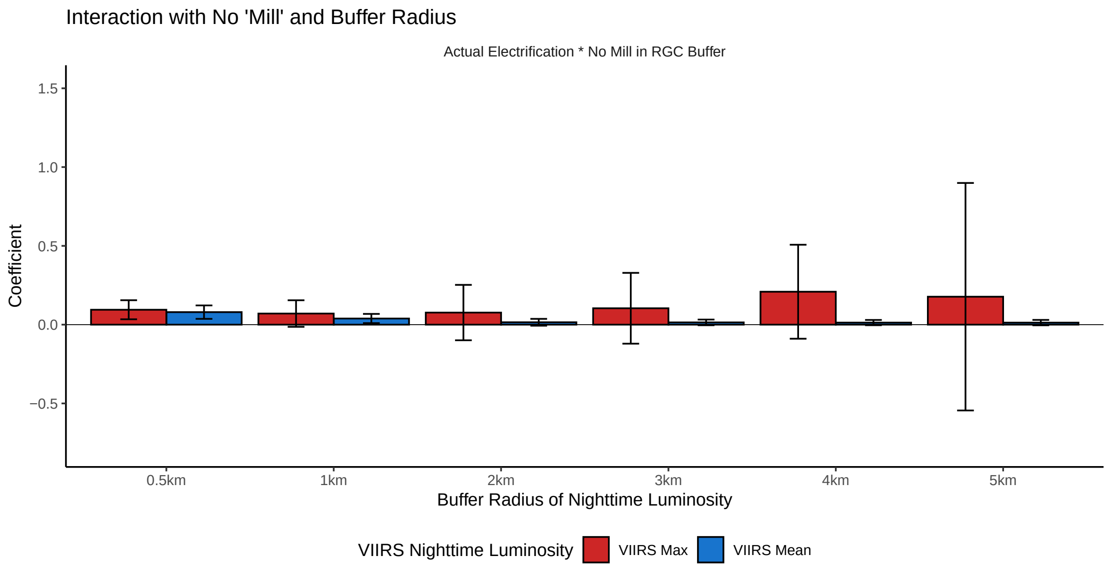
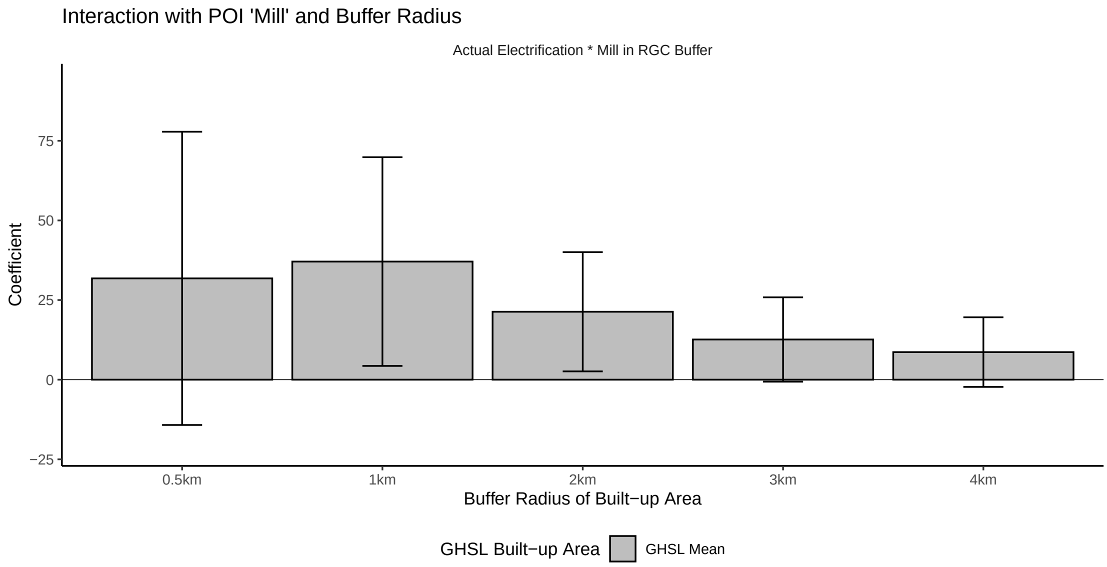


Figure A16: Effect of electrification on built-up area

(a) Built-up area around electrified places with mills



(b) Built-up area around electrified places without mills

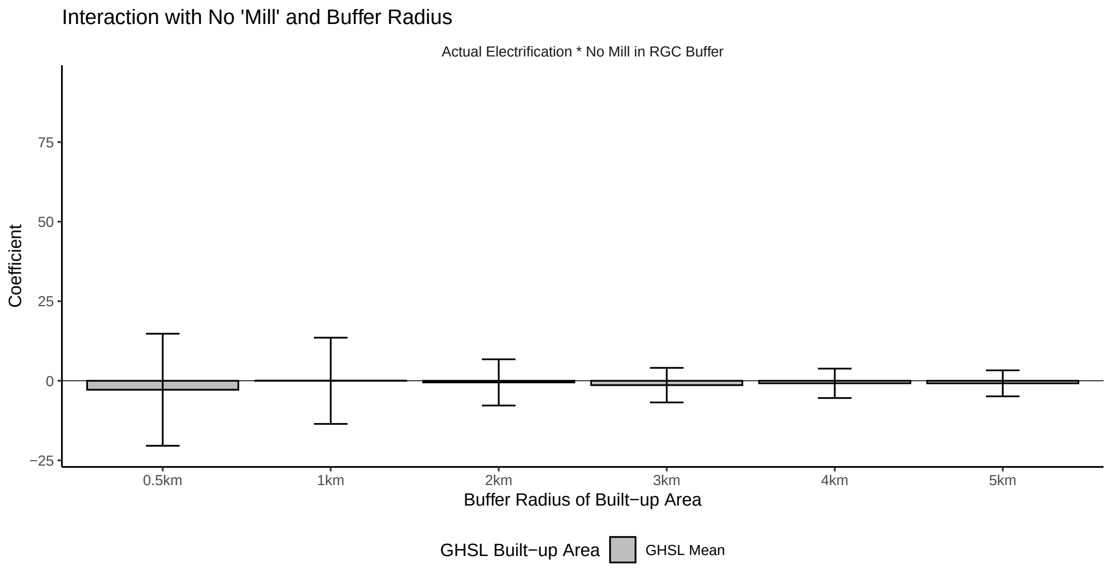
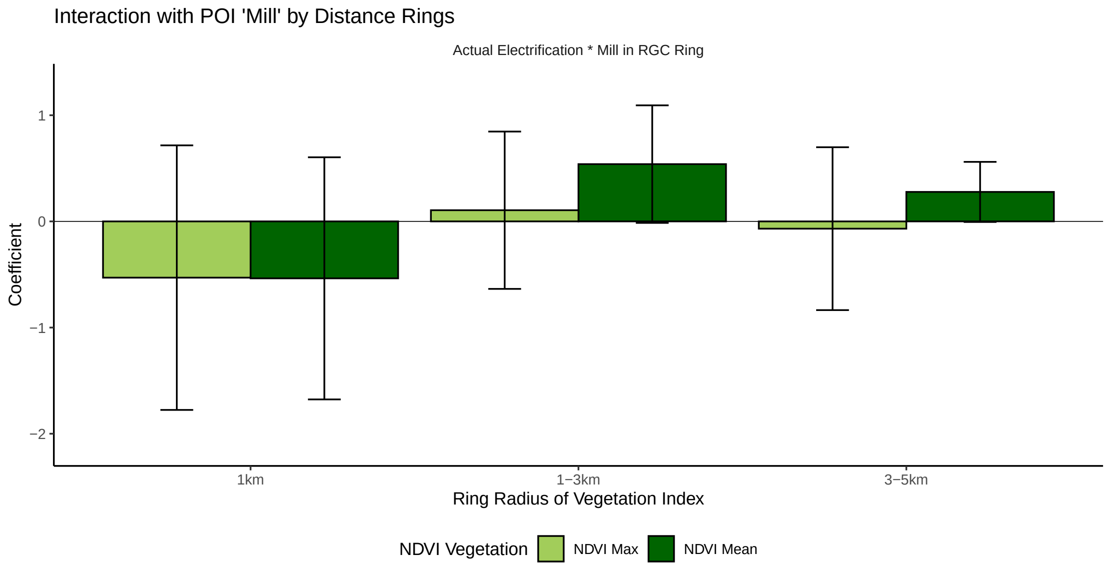


Figure A17: Effect of electrification on vegetation

(a) Vegetation around electrified places with mills



(b) Vegetation around electrified places without mills

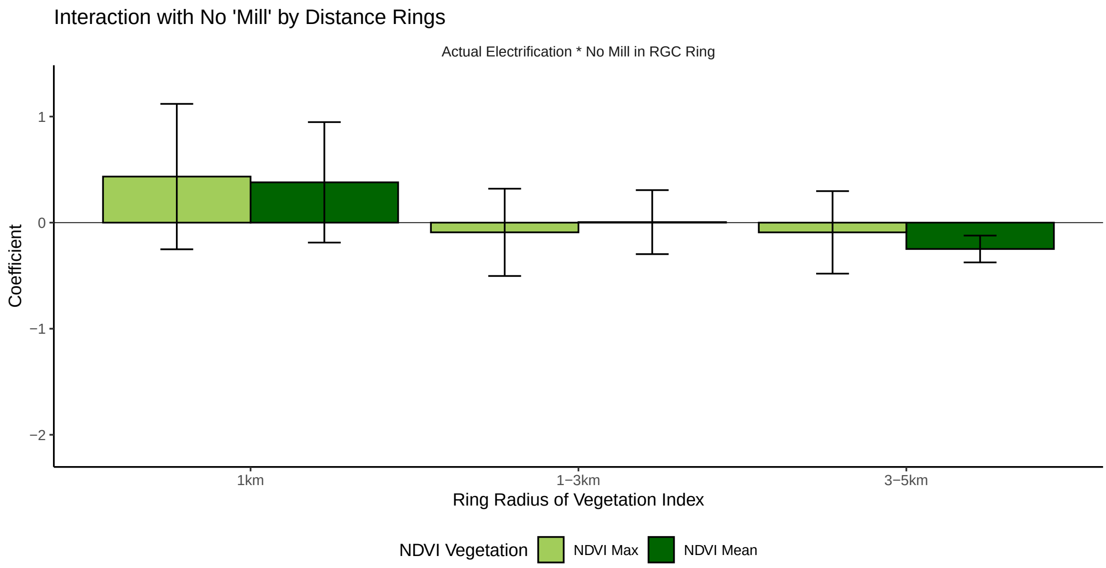
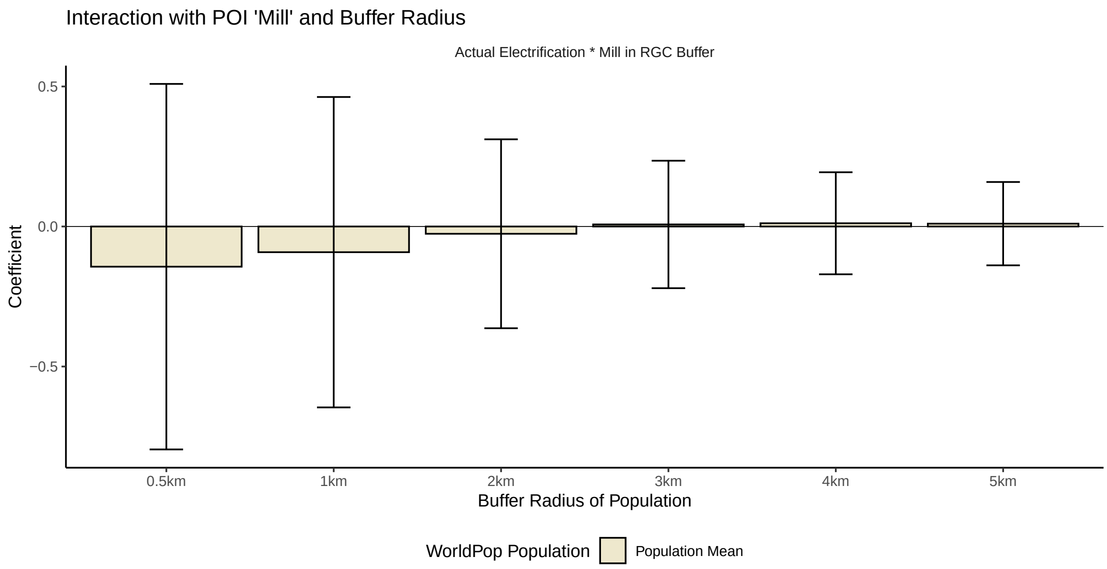


Figure A18: Effect of electrification on remotely-sensed population

(a) Population around electrified places with mills



(b) Population around electrified places without mills

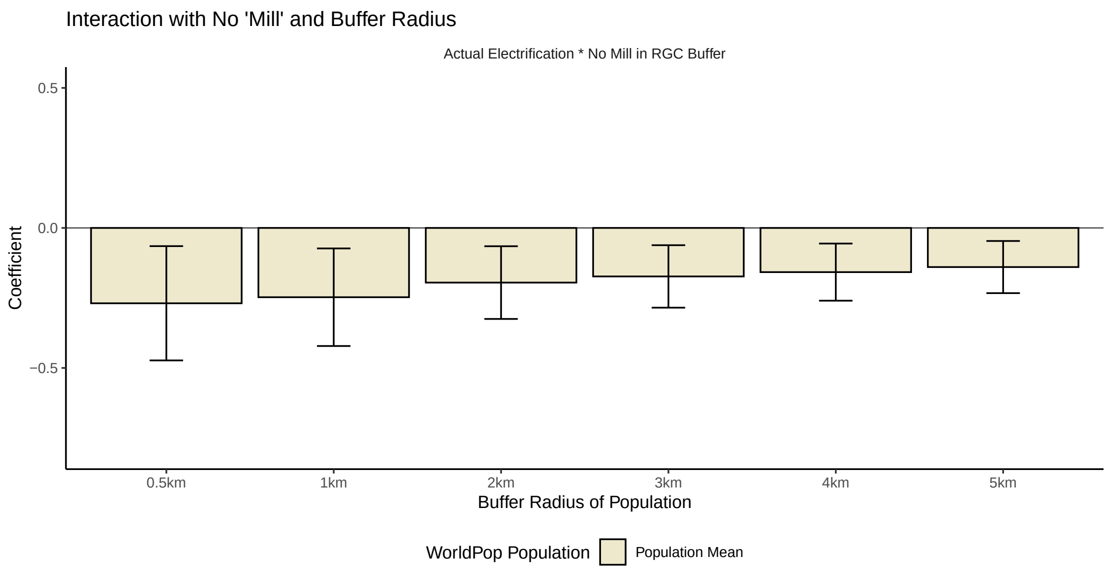


Figure A19: Distribution of Rural Growth Centres (RGCs) by mill status at baseline

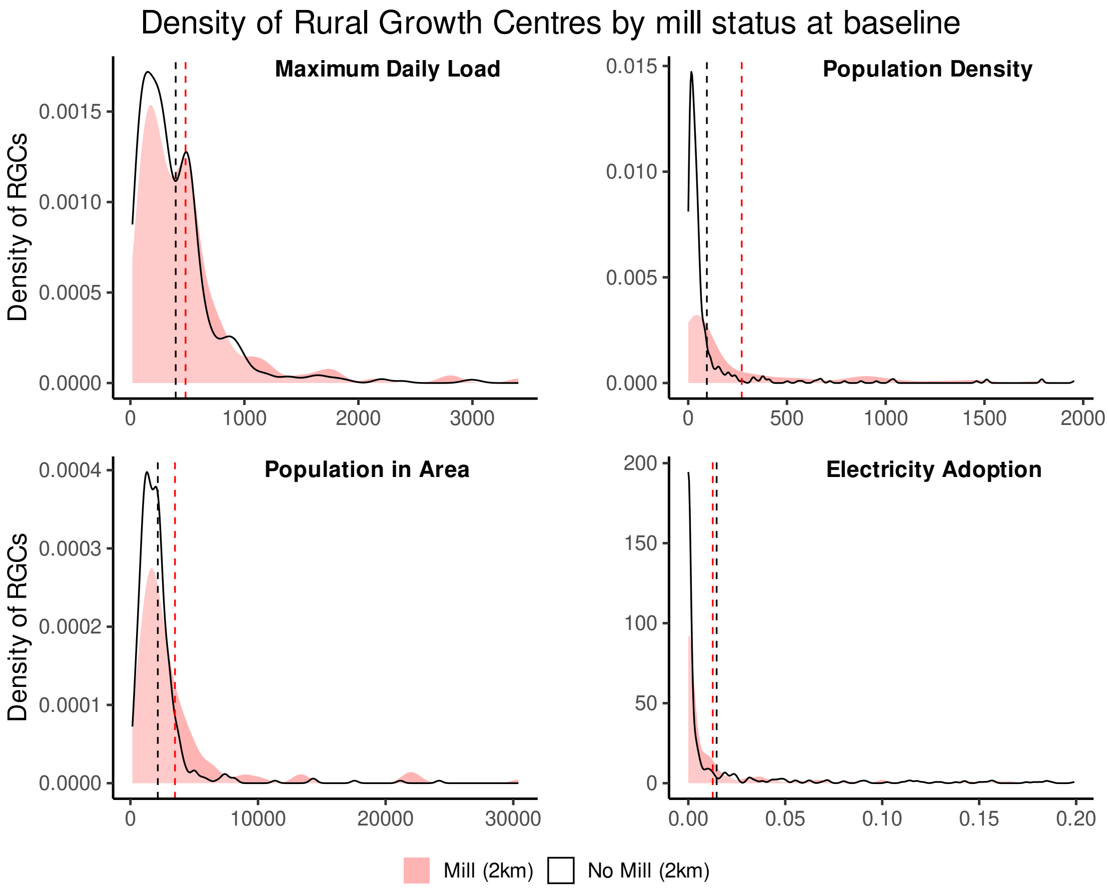


Figure A20: Predictive power of variation in pre-existing Rural Growth Centre (RGC) statistics for mean RGC-level post-electrification treatment residuals, by mill status

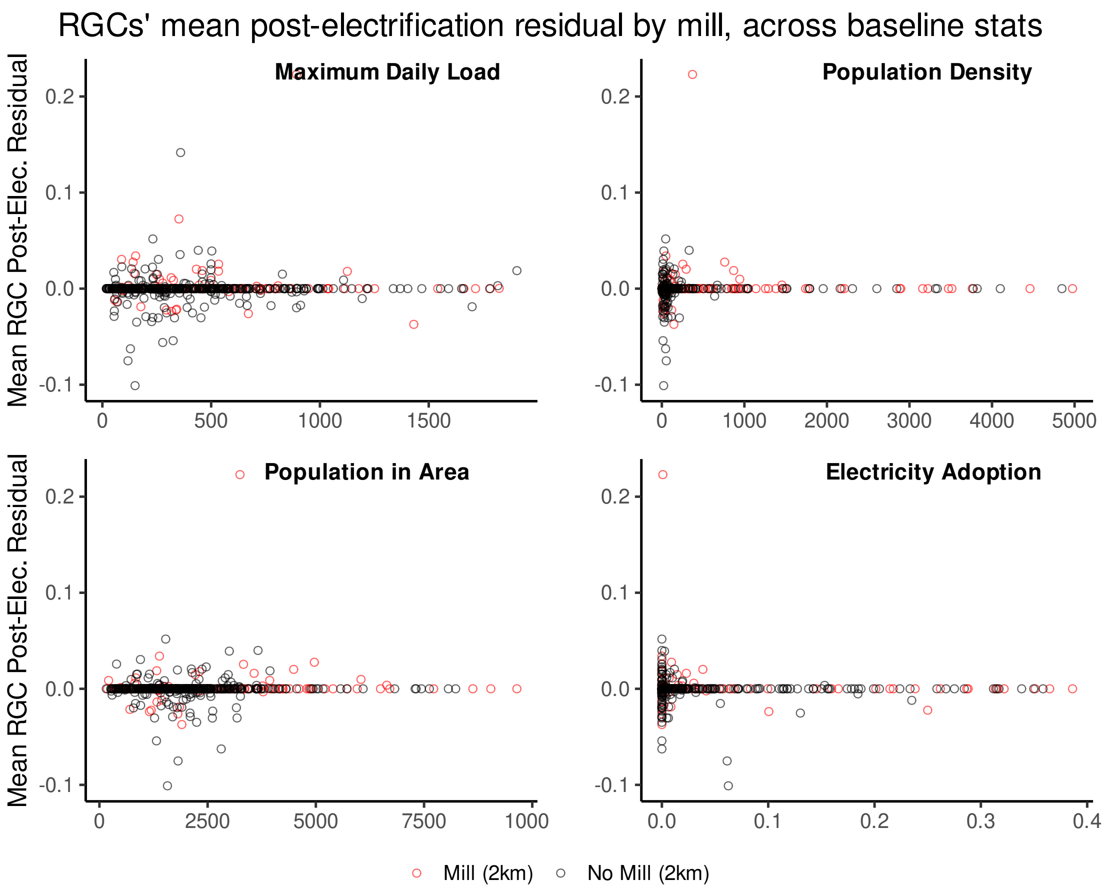


Figure A21: Event studies of nightlights on electrification: overall, with and without mill (brute force)

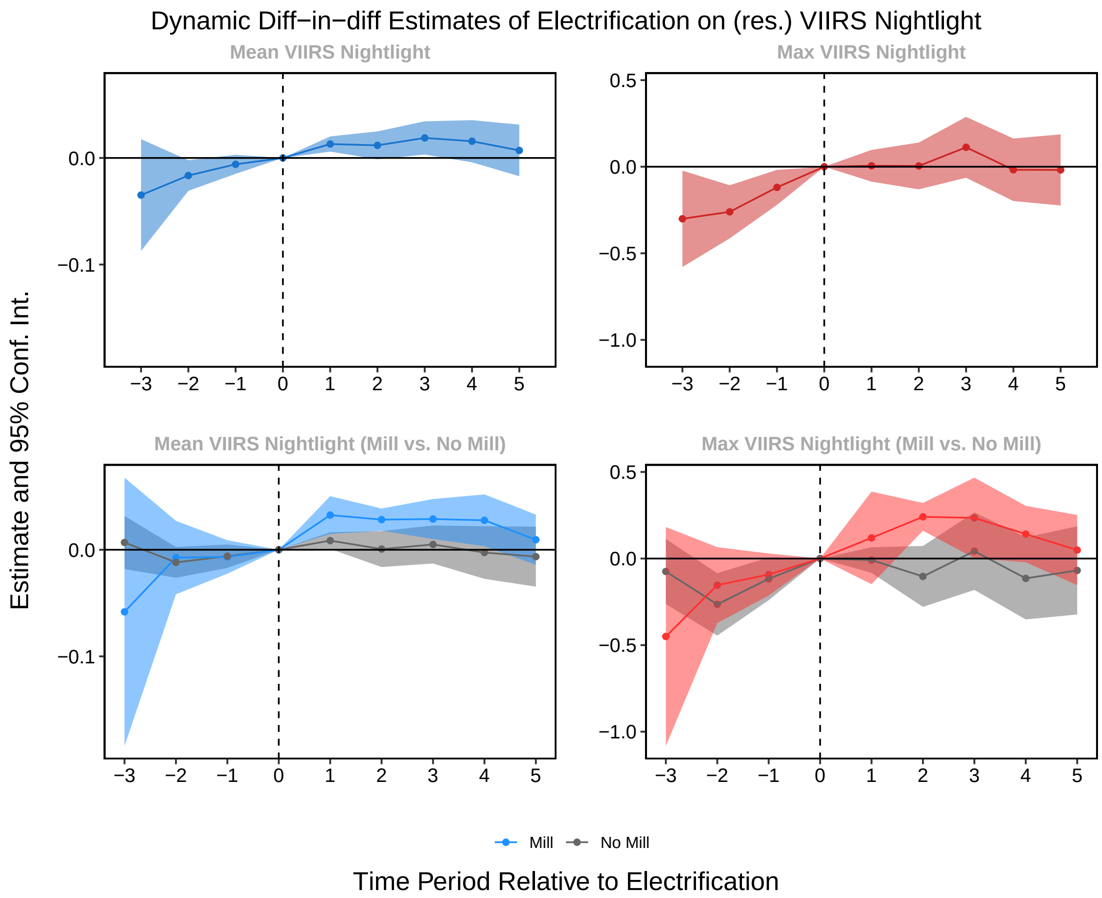
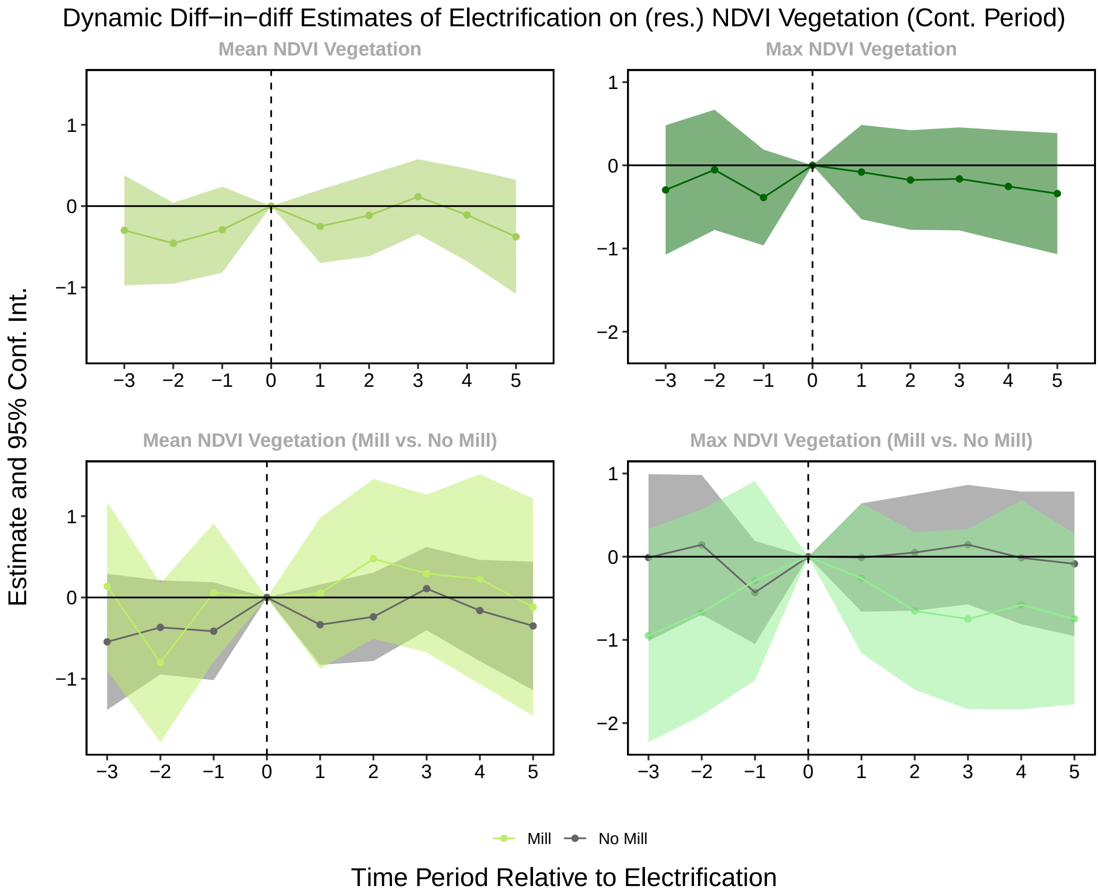


Figure A22: Event studies of vegetation greenery in RGC (3km buffer) on electrification: overall, with and without mill



A.II Appendix: Additional tables

Table A1: Non-Payment of Electricity Usage Across Zambia (LCMS-2015)

	Rural EAs		Urban EAs		All EAs	
	N	%	N	%	N	%
Every household pays	76	49.7	86	50.9	162	50.3
Some households pay	39	25.5	80	47.3	119	37.0
No household pays	38	24.8	3	1.8	41	12.7
Total	153	100.0	169	100.0	322	100.0

Includes all LCMS-2015 enumeration areas (EAs) where at least one respondent answers questions on electricity usage and payment. Share of EA not paying for electricity derived from dividing count of those paying for electricity divided by those stating grid access. Urban/rural status of EA derived from PHC 2010 ward-level urban/rural status, in line with remainder of analysis.

Table A2: School-Reported Electrification vs Electrification Status from Alternative Sources (REA Project, ZESCO/REA Transformer and Health Facility-Reported)

	REA Project		ZESCO	Health Facility			
	OLS	DD	OLS	OLS	OLS	DD	DD
School Electrification	0.53***	0.23***	0.63***	0.72***		0.52***	
	(0.05)	(0.03)	(0.04)	(0.03)		(0.05)	
School Electrification (pct)					0.94***		0.66***
					(0.05)		(0.07)
School/RGC FE		✓				✓	✓
Year FE		✓				✓	✓
R^2	0.21	0.74	0.40	0.53	0.46	0.80	0.79
Obs.	2137	2137	743	701	701	1710	1710

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Sample of schools from inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. School Electrification = 1 for all schools/RGCs with any schools reporting electrification. REA Projects' electrification status is reported at the school-level, including universe of project records from 2006-2020. ZESCO Transformers' electrification status is reported as single cross-section in 2020 at the RGC-level (including a minority of REA transformers not yet handed over to ZESCO). Health Facility Electrification is reported at the RGC-level, last updated in 2017. Difference-in-difference results for health facilities reported for all years with health facility reporting (2005, 2012, 2017).

Table A3: Nightlights on Actual Electrification across Specs. (2012-2020)

	VIIRS Mean Nightlights			VIIRS Max Nightlights		
	OLS	DD	DD	OLS	DD	DD
Actual Elect. (pct)	0.51***	0.03**	0.04**	1.96***	0.30***	0.23**
	(0.12)	(0.01)	(0.01)	(0.40)	(0.08)	(0.10)
Year FE		✓	✓		✓	✓
RGC FE		✓	✓		✓	✓
Const-Year trends			✓			✓
R^2	0.06	0.94	0.96	0.05	0.83	0.88
No. of Years		9	9		9	9
No. of RGCs		742	742		742	742
No. of Const-Years			1215			1215
Obs.	6678	6678	6678	6678	6678	6678

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. RGC-level actual electrification derived from schools existing since 2008. Const-Year trends denotes constituency-year dummies.

Table A4: Nightlights on Actual Electr. and Mill Interaction (2012-2020): Robustness to Alternative School Panels (Endogenous School Construction)

	VIIRS Mean Nightlights			VIIRS Max Nightlights		
	(1)	(2)	(3)	(4)	(5)	(6)
Actual Elect. (pct)	0.02	0.02	0.02	0.08	0.07	0.08
	(0.01)	(0.01)	(0.01)	(0.09)	(0.08)	(0.08)
Actual Elect. (pct) * Mill in RGC	0.08**	0.06*	0.05	0.63**	0.42*	0.33
	(0.04)	(0.03)	(0.03)	(0.27)	(0.22)	(0.28)
Year FE	✓	✓	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓	✓	✓
Const-Year trends	✓	✓	✓	✓	✓	✓
R^2	0.96	0.96	0.96	0.88	0.88	0.88
No. of Years	9	9	9	9	9	9
No. of RGCs	742	748	757	742	748	757
No. of Const-Years	1215	1215	1212	1215	1215	1212
Obs.	6678	6673	6679	6678	6673	6679

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. Models (1), (2), (5), (6) use RGC-level actual electrification derived from schools existing since 2008. (3) and (7) use actual electrification derived from schools existing in 2020. (4) and (8) use actual electrification derived from all schools. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC.

Table A5: Nightlights on Actual Electr. and Mill Interaction (2012-2020): Robustness to Rural Sample Restriction

	VIIRS Mean Nightlights				VIIRS Max Nightlights			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actual Elect. (pct)	0.02	0.02	0.00	0.01	0.08	0.08	0.04	0.07
	(0.01)	(0.01)	(0.01)	(0.01)	(0.09)	(0.09)	(0.08)	(0.09)
Actual Elect. (pct) * Mill in RGC	0.08^{**}	0.08^{**}	0.03^{**}	0.03^{*}	0.63^{**}	0.63^{**}	0.33^{*}	0.36^{*}
	(0.04)	(0.04)	(0.02)	(0.02)	(0.27)	(0.27)	(0.18)	(0.19)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓	✓	✓	✓	✓
Const-Year trends	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.96	0.96	0.88	0.93	0.88	0.88	0.59	0.78
No. of Years	9	9	9	9	9	9	9	9
No. of RGCs	742	742	681	715	742	742	681	715
No. of Const-Years	1215	1215	1143	1197	1215	1215	1143	1197
Obs.	6678	6678	6129	6435	6678	6678	6129	6435

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. Model key: (1) and (5) include only RGCs with a 2010 population below 18849.56 (baseline specification). (2) and (6) include only RGCs with a 2010 population below 20000. (3) and (7) include only RGCs with a 2010 population below 1500. (4) and (8) include only RGCs either marked as rural, or with a 2010 population density below 100 people per square kilometer. RGC-level actual electrification derived from schools existing since 2008. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC.

Table A6: Nightlights on Actual Electr. and Mill Interaction (2012-2020): Robustness to Consequential and Off-Grid RGCs

	VIIRS Mean Nightlights				VIIRS Max Nightlights			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actual Elect. (pct)	0.02	0.01	0.01	0.01	0.08	0.08	0.07	0.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.09)	(0.08)	(0.09)	(0.07)
Actual Elect. (pct) * Mill in RGC	0.08**	0.06**	0.06**	0.08**	0.63**	0.36*	0.43**	0.45*
	(0.04)	(0.03)	(0.03)	(0.04)	(0.27)	(0.18)	(0.19)	(0.25)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓	✓	✓	✓	✓
Const-Year trends	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.96	0.96	0.96	0.96	0.88	0.84	0.83	0.87
No. of Years	9	9	9	9	9	9	9	9
No. of RGCs	742	1135	966	911	742	1135	966	911
No. of Const-Years	1215	1251	1233	1233	1215	1251	1233	1233
Obs.	6678	10215	8694	8199	6678	10215	8694	8199

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Model key: (1) and (5) use sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan (REMP) (baseline specification). (2) and (6) include all RGCs, including both consequential and solar home system RGCs. (3) and (7) include inconsequential grid and off-grid RGCs in the REMP. (4) and (8) include consequential and inconsequential grid RGCs in the REMP. RGC-level actual electrification derived from schools existing since 2008. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC.

Table A7: Nightlights on Actual Electr. and Mill Interaction (2012-2020): Robustness to Time-Varying Controls (Roads, Main Roads, Rainfall)

	VIIRS Mean Nightlights				VIIRS Max Nightlights			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Actual Elect. (pct)	0.02	−0.03	−0.04	0.01	0.08	0.02	−0.20	0.06
	(0.01)	(0.04)	(0.04)	(0.01)	(0.09)	(0.27)	(0.16)	(0.09)
Actual Elect. (pct) * Mill in RGC	0.08**	0.17	0.19*	0.08**	0.63**	1.06	1.27**	0.61**
	(0.04)	(0.10)	(0.10)	(0.04)	(0.27)	(0.86)	(0.51)	(0.27)
Road Control		✓				✓		
Main Road Control			✓				✓	
Mean Precipitation Control				✓				✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓	✓	✓	✓	✓
Const-Year trends	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.96	0.95	0.95	0.96	0.88	0.87	0.87	0.88
No. of Years	9	3	3	9	9	3	3	9
No. of RGCs	742	742	742	742	742	742	742	742
No. of Const-Years	1215	405	405	1215	1215	405	405	1215
Obs.	6678	2226	2226	6678	6678	2226	2226	6678

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. RGC-level actual electrification derived from schools existing since 2008. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC. Road and Main Road Controls denote an indicator values that takes the value of one if at least one road or main road situated within 2km of RGC.

Table A8: Nightlights on Actual Electr. and Mill Interaction (2012-2020): Robustness to Alternative POI

	VIIRS Mean Nightlights						VIIRS Max Nightlights					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Actual Elect. (pct)	0.04**	0.02	0.02	0.04**	0.03**	0.01	0.23**	0.08	0.08	0.23**	0.25**	0.08
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.10)	(0.09)	(0.18)	(0.11)	(0.11)	(0.14)
Actual Elect. (pct) * Mill in RGC		0.08**						0.63**				
		(0.04)						(0.27)				
Actual Elect. (pct) * Commercial in RGC			0.06						0.42			
			(0.04)						(0.32)			
Actual Elect. (pct) * Agriculture in RGC				−0.00						0.00		
				(0.02)						(0.15)		
Actual Elect. (pct) * Religion in RGC					0.02						−0.03	
					(0.02)						(0.16)	
Actual Elect. (pct) * Admin in RGC						0.19**						1.21*
						(0.09)						(0.66)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Const-Year trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.96	0.96	0.96	0.96	0.96	0.96	0.88	0.88	0.88	0.88	0.88	0.88
No. of Years	9	9	9	9	9	9	9	9	9	9	9	9
No. of RGCs	742	742	742	742	742	742	742	742	742	742	742	742
No. of Const-Years	1215	1215	1215	1215	1215	1215	1215	1215	1215	1215	1215	1215
Obs.	6678	6678	6678	6678	6678	6678	6678	6678	6678	6678	6678	6678

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. RGC-level actual electrification derived from schools existing since 2008. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC. Commercial in RGC denotes an indicator value that takes the value of one if at least one commercial Point Of Interest (POI) situated within 2km of RGC. Agriculture in RGC denotes an indicator value that takes the value of one if at least one agricultural POI situated within 2km of RGC. Religion in RGC denotes an indicator value that takes the value of one if at least one religious POI situated within 2km of RGC. Admin in RGC denotes an indicator value that takes the value of one if at least one administrative POI situated within 2km of RGC.

Table A9: Nightlights on Actual Electr. and Mill Interaction (2012-2020): Robustness to Mill vs Admin POI

	VIIRS Mean Nightlights						VIIRS Max Nightlights					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Actual Elect. (pct)	0.02	0.01	0.14	0.01	−0.01	0.14^{**}	0.08	0.03	0.41	0.08	−0.15	1.10^{**}
	(0.01)	(0.01)	(0.14)	(0.02)	(0.02)	(0.07)	(0.09)	(0.09)	(0.32)	(0.14)	(0.19)	(0.44)
Actual Elect. (pct) * Mill in RGC	0.08^{**}	0.02^{*}	−0.10				0.63^{**}	0.31^{***}	−0.16			
	(0.04)	(0.01)	(0.12)				(0.27)	(0.11)	(0.36)			
Actual Elect. (pct) * Admin in RGC				0.19^{**}	0.25	0.04				1.21^{*}	1.81	−0.34
				(0.09)	(0.16)	(0.11)				(0.66)	(1.14)	(0.53)
Sample	full	no admin	admin	full	no mill	mill	full	no admin	admin	full	no mill	mill
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RGC FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Const-Year trends	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.96	0.95	0.99	0.96	0.97	0.98	0.88	0.80	0.99	0.88	0.91	0.93
No. of Years	9	9	9	9	9	9	9	9	9	9	9	9
No. of RGCs	742	652	90	742	556	186	742	652	90	742	556	186
No. of Const-Years	1215	1170	513	1215	1152	693	1215	1170	513	1215	1152	693
Obs.	6678	5868	810	6678	5004	1674	6678	5868	810	6678	5004	1674

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. All standard errors clustered at electrification package-level. Sample of inconsequential, non-urban, non-solar home system Rural Growth Centres (RGCs) included in Rural Electrification Master Plan. RGC-level actual electrification derived from schools existing since 2008. Mill in RGC denotes an indicator value that takes the value of one if at least one mill situated within 2km of RGC. Commercial in RGC denotes an indicator value that takes the value of one if at least one commercial Point Of Interest (POI) situated within 2km of RGC. Agriculture in RGC denotes an indicator value that takes the value of one if at least one agricultural POI situated within 2km of RGC. Religion in RGC denotes an indicator value that takes the value of one if at least one religious POI situated within 2km of RGC. Admin in RGC denotes an indicator value that takes the value of one if at least one administrative POI situated within 2km of RGC.

A.III Appendix: Additional details on electrification data

In the following, we describe how employ administrative data collected by public school headmasters across Zambia to construct an objective measure of village-level electrification by year, which allows us to accurately determine the arrival time of the electric grid across locations.

A.III.1 EMIS database extraction and cleaning

The raw data from the Ministry of General Education's Education Management Information System (EMIS) consists of a Microsoft Access database for each year from 2004 to 2020. Each of these relational databases, in turn, contains many different data tables as well as the encoding of value labels for all variables. The data cleaning process is described below.

Step 1: Data export from Microsoft Access to Microsoft Excel. First, we export the following data tables from the Access database for each year from 2004 to 2020 to excel: `gdb_class`, `gdb_dropouts`, `gdb_finance`, `gdb_Ownership`, `gdb_Readmit`, `gdb_repeater`, `gdb_school`, `gdb_student`, `gdb_summary`, `gdb_TeacherDepart`. Additionally, we export the following value label tables for all years: `a_accred`, `a_Age`, `a_AgencyFound`, `a_AgencyRunning`, `a_Constituency`, `a_district`, `a_DropOutReason`, `a_EdLevel`, `a_Finance`, `a_genderStatus`, `a_location`, `a_ownership`, `a_Readmit`, as well as `a_region`, `a_TeacherDepart`, `a_ward`, and `a_zone`.

Step 2: Data import. Second, we import each of the data tables into statistical software. Within a given year we merge records using the variable "code" as the unique school identifier. Next, we import the labels associated with the values of each variable from the following label tables. In cases where value labels change across years, we harmonize these. Finally, we append the data from all years.

Step 3: Cleaning school identifiers. Third, we clean school identifiers. In principle, there are three sets of variables that can be used to identify schools: (i) school number, (ii) school code, (iii) school location (district) and school name. We begin by examining variation of school code within school number and ensure that changes in code within number do not reflect changes in school identity, judging from the wealth of information we have about each school in each year. Next, we examine variation in school location and school name within school number. To this end, we first harmonize province and district names across EMIS years such that they are aligned with the 2010 population census which featured 72 districts across 9 provinces. Second, we clean up the spelling of school names. Then we ensure that changes in school location and name within school number do not represent changes in school identity. Finally, we check for the presence of duplicates comparing different school numbers in the same location that are associated with similar school names to each other. In the resulting dataset each entry is uniquely identified by school number and year.

Step 4: Cleaning student and teacher numbers. Fourth, we clean enrollment and teacher information. We ensure that the total number of students reported is equal the sum of students reported each gender and grade. Furthermore, we correct enrollment and teacher numbers for obvious data entry errors. Finally, we check the resulting student-teacher ratio for outliers and inconsistencies.

Step 5: Cleaning school coordinates. The EMIS data contains GPS coordinates of a subset of schools that were collected by the Ministry of Education in 2007. We leverage additional school location information from other initiatives that collected school coordinates over the years and merge these to our dataset. The sources of these additional coordinates are: ZEMA, FSDZ, Ministry of Health, Innovations for Poverty Action (community health worker project), DFID and the World Bank (ZEEP). Each of these sources is a list of schools with GPS coordinates. Using a combination of automatic and manual methods, we match these to our data using district, school name and school number. This leaves us with up to seven sets of coordinates for each school.

Step 6: Adding electricity information. In the last step, we add information on the source of power in each school in each year. To this end, we export the table `gdb_waterelec` and the corresponding value labels from the Access database, import them into statistical software. The available information allows us to generate an indicator for each of the following sources of power – grid, generator, solar, hydro – in each year and school. Finally, we merge the resulting data to our main data set (step 5).

A.III.2 EMIS school geolocation

Additionally, we have geographic locations information for each school, namely the province, district, constituency and ward as reported by the school headmasters. For 4,598 out of 14,395 schools (31.94%), the school headmaster also registered a GPS coordinate for their school, which allows us to overlay with the 2010 Zambian census shapefile and provides a sanity check against the reported district and ward information.

To fill in the missing coordinates, we obtained six additional sources for school coordinates [85]. However, we found that some of these coordinates did not perfectly match

[85] These sources are the Zambia Environmental Management Agency (ZEMA), Financial Sector Deepening Zambia (FSDZ), Community Health Workers (CHW), Ministry of Health (MOH), Department for International Development (DFID), and World Bank (WB)

with the reported coordinates. To geolocate these schools, we first scored the six sources on 'quality', which was a measure of how often any of these sources' coordinates corresponded with a school based on satellite imagery. Then, a given set of coordinates was compared to all others, and assigned a score based on whether it was within a 5.55 km wide box of the other coordinate. Suppose there were three coordinates in a row, A - B - C, each 4 km apart. B would have a score of 2, whereas A and C would have a score of 1, since A and C are close to B but not close to each other. When there is a tie, we take the coordinate that is of the highest quality.

Schools were assigned into one of five cases. Out of 14,395 schools, 4,729 (32.85%) of them were case 5 schools where there were three or more coordinates available, and the highest scoring coordinate (or highest quality coordinate in the case of a tie) was taken. 1,179 (8.19%) were case 4 schools, where only two coordinates were available, and were close to each other, so the higher quality coordinate was taken. 361 (2.51%) were case 3 schools, where there were two coordinates but they were not close to each other, so the higher scoring coordinate was taken. 4,010 (27.86%) schools were case 2 schools where we only had one coordinate.

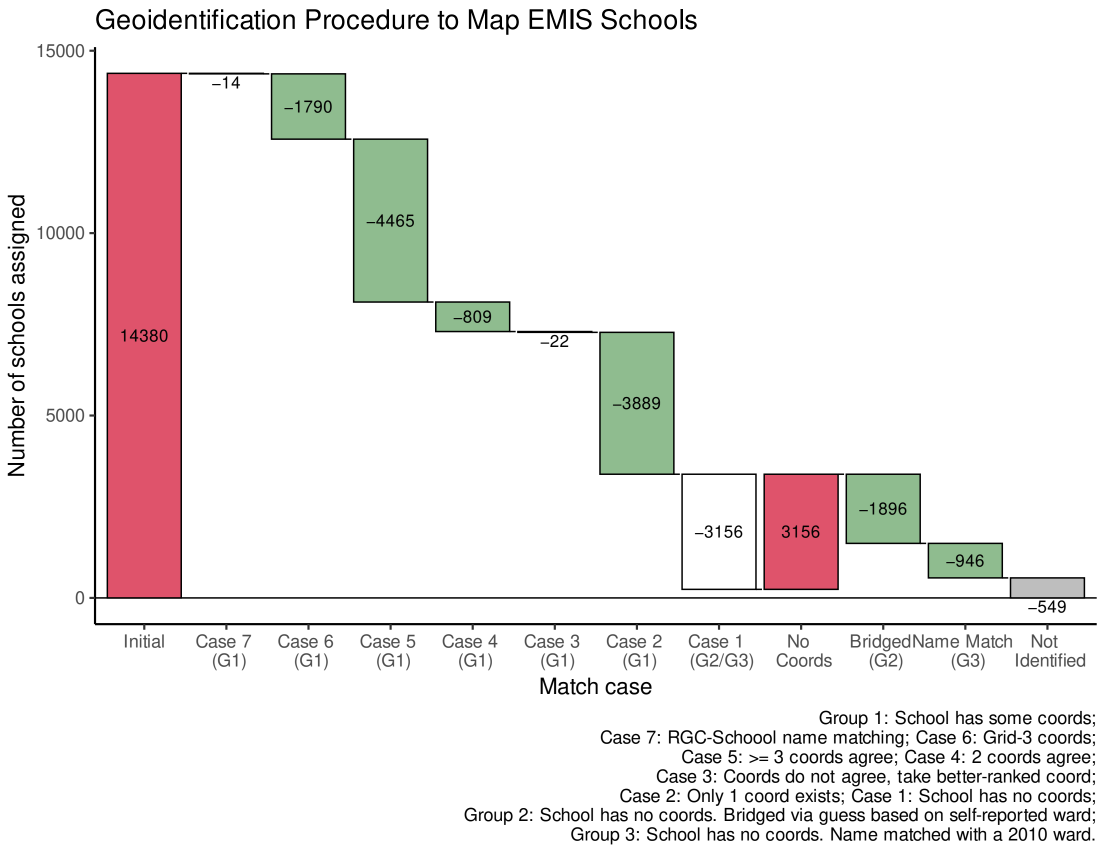
4,116 (28.59%) case 1 schools had no coordinates available. For these, we will have to rely on the self reported geographic locations (wards). However, for the schools with coordinates, the self reported locations often differ from where the coordinates say they are. This could be due to changing ward and constituency boundaries, or misreporting. Therefore, to get a best guess of where the schools without coordinates are, we generate a statistical bridge to match the self reported ward with the modal actual ward.

For example, suppose there are 12 schools that self-report to be in Freedom ward, Livingstone constituency. 9 schools have coordinates, and we would expect them to report that they are in Freedom ward, but only 5 of them do so. Since the remaining 4 schools do have coordinates, we use the coordinates first and assign them the wards the coordinates say they are in. However, for the remaining 3 schools with no coordinates, we will assume that they are in Freedom ward, since the modal ward for a school self-reporting to be in Freedom ward is also Freedom. If the modal ward was a different ward, we would use that.

So, we have for each self-reported ward, our best guess of where the ward would likely be had there been coordinates. Therefore, of the 4116 case 1 schools with no coordinates, we first bridged them based on the self-reported ward that they reported to be in. These are the Bridged 1 schools, of which we could geolocate 1,896 (60.1%) schools. This method allows us to partially account for administrative boundary changes across the 2000 and 2010 census years.

Some schools self-reported to be in wards that did not exist in 2010. For example, if a school was in Freedom ward, but reported itself as being in Fredom ward, then it would not be geolocated in bridged 1. Furthermore, schools with no coordinates whatsoever would also not be geolocated in bridged 1. Therefore, after fixing these typos, we would then be able to match these self-reported locations with locations in the 2010 census shapefile to geoidentify them to a ward. There are 946 such bridged 2 schools.

In the end, we fail to match (and thus geo-identify) 549 (3.8%) of schools to a ward. The entire procedure is summarised in the waterfall diagram below.



The EMIS dataset geo-identified above comprises school-year observations for 11,224 schools. It features an unbalanced panel structure and the source of electrification, i.e., grid, solar, hydro, generator, or no power, is specified for each year.

The school electrification data features a unbalanced panel structure with annual information on the school's source of electricity. We focus on grid vs any other option.

A.IV Appendix: Additional details on empirical strategy

A.IV.1 Rural growth centres

In February 2009, JICA, the Japanese development organisation, submitted an extensive report that details the Rural Electrification Master Plan (REMP) for Zambia. This plan reflects an overarching effort to increase the rate of electrification in rural Zambia from 3% to 51% by the year 2030. Implementation strategy is centred around electrification of rural growth centres (RGC), which are defined as rural localities with a high concentration of residential settlements and are supposed to represent centres of rural economic activity. Each RGC serves surrounding villages in its "catchment area", allowing rural residents to sell their agricultural and handcrafted products, as well as purchase groceries or access public services (e.g. school, health facility, postal office or administrative functions). Another significant function of rural growth centres is processing maize into flour form using electrified grain mills (so-called 'hammer mills') for making 'Nsima', the Zambian staple food.

JICA predicted two main channels of economic impact of gaining stable access to electricity for RGCs. The first is increased usage of electric refrigerators at shops and health facilities. Electric refrigerators offer considerable benefits over the currently used paraffin refrigerators in terms of efficiency and reliability, which reduces the risk of spoiled food or vaccines. The second channel is a growing supply of electrified hammer mills for maize flour production, and consequently, lower milling fees for the villagers. Unelectrified RGCs generally have maize mills with capacity of about 15kW, powered by privately-owned diesel generators, and electrification via the local low-voltage distribution network is expected to significantly increase production capacity.

A.IV.2 Priority ranking of RGCs

In the JICA report, a total of 1,200 rural growth centres are identified for electrification planning and ranked based on forecasted potential daily max power demand in 2030 at each location.[86] To forecast potential demand for unelectrified RGCs based on 2006 survey data of consumption trends in electrified RGCs, the first step is estimating an hourly average daily load curve per unit of facility for four different types of electricity consumers: (1) public facilities (2) business entities (3) hammer mills (4) households. Then the report forecasts the number of consumers in 2030 by type in the 1,200 unelec-

[86] Among unelectrified RGCs, BOMA (district centres) are give priority over the other RGCs, and all other RGCs are ranked by the size of potential demand in 2030.

trified RGCs using 2006 numbers. Specifically, the numbers of public facilities, business entities, and households are assumed to grow at a constant annual rate of 2.9%, same as the population growth rate announced in "Population Projection Report" published by Central Statistics Office in November 2003. The number of hammer mills in 2030 is estimated using the forecasted number of households in 2030 divided by 179, which is the unit hammer mill service ratio (i.e. every hammer mill serves an average of 174 households). All types of consumers are assumed to be 100% electrified in 2030,[87] and the forecasted number of consumers is multiplied by the unit daily load curve for each type of consumers to create the daily load timetables in any given RGC. The total potential daily load curve for a RGC in 2030 is obtained by summing up daily load curves across all types of customers and the maximum value (i.e. daily peak demand) is used as a design capacity of electrification facilities and the sole parameter to determine electrification priority of the 1,200 RGCs.

A.IV.3 Project package creation and prioritisation

Subsequent to eletrification priority ranking, the 1,200 unelectrified RGCs are grouped into 180 project packages based on geographical proximity and existing infrastructure. Each package contains one highly ranked RGC that serves as the main candidate to be powered by distribution line extension from the nearest substation, and other unelectrified RGCs with lower priority that fall along the potential electrification route are also included in the same package. Package size varies between 1 and 22, with an average of 9 RGCs per package.

Projects with numerous unelectrified RGCs are further divided into components to be electrified via either transmission/distribution line extension or stand-alone electrification modes (e.g. solar home system, mini-hydro power development, or diesel generator) in order to cut down inefficient distribution line construction and optimize resource utilisation. JICA conducted case studies for all project packages to determine the optimal electrification method for each RGC, which resulted in the 180 packages being further divided into 835 project components. The case study relies on "Unit Life Time Cost in Net Present Value (US$/kWh)" to determine the best way to power any given RGC,[88]

[87] The Rural Electrification Authority (REA) and Department of Energy (DoE) concede that assuming full electrification of facilities in 2030 seems to result in overestimation of potential power demand. However, since REA and DoE are planning to extend electricity access to the catchment areas of these 1,200 RGCs after 2030, they decided to apply 100% electrification rate in the forecase in order to allow for extra supply margin on the design capacity.

[88] The unit life time cost in net present value is calcuated as $\frac{NPV(\text{Construction/Initial cost (\$)+Total lifetime operation \& maintenance cost (\$)})}{\text{Total amount of electricity consumable during lifetime (kWh)}}$. The electrification mode with

and concludes that 79.9% of RGCs, which corresponds to 94.5% of households, are scheduled to be electrified by transmission/distribution line extension, whereas 19.8% of RGCs fall under SHS electrification and only 0.3%, or 4 RGCs are planned for mini-hydro electrification.

With the optimal case of segmentation and electrification mode determined, JICA ranked the 180 project packages in electrification priority based on financial indicators (e.g. Financial Internal Rate of Return (FIRR) and Economic Internal Rate of Margin (EIRR)). To estimate the financial indicators, JICA assumed a series of parameters.[89]

A.IV.4 Inconsequentially electrified locations

To provide more detail on why the earlier-than-planned electrification of such inconsequential rural locations could be deemed plausibly exogenous to outcomes, we provide more background information on the ranking and network expansion algorithm employed for the REMP. JICA and REA identified 1,200 rural growth centres (RGCs) in Zambia that were supposed to contain the set of rural villages and small towns that were deemed both of substantial economic or political important, while at the same time still remained unelectrified at baseline in 2008.[90] Based on 2000 Census population data and back-of-the-envelope population growth estimates, RGCs are ranked by their expected total electricity demand (see an excerpt of the REMP's Table 5.11 below), which is essentially a linear function of each RGC's expected population size in 2030.

The REMP then proposes an electrification roll-out that results from executing the following algorithm: in the first step, the highest ranked (i.e. most important) RGC around each existing electric grid substation is selected. Then, this highest ranked RGC is bundled with all RGCs along the way between substation and highest ranked RGC into a 'package', irrespective of nearby RGC's ranking. Third, a hypothetical return on investment is calculated for each possible configuration of a given package while dropping nearby RGCs one by one to find the optimal package size.[91] Any un-selected RGC from

the lowest unit life time cost in net present value is selected.

[89]The methodology only assumes a constant growth rate of 2.9% for the number of customers across all types, thus an implication is that if an unelectrified RGC starts with zero customer of a certain type, electrification will not create new customers in this category.

[90]The exact selection process for the 1,207 RGCs remains unclear and is neither documented in the full, final REMP report, nor any of the many appendices to the report. We are currently in conversation with REA and the Ministry of Energy to understand better how this initial list was formed. However, since most of our analysis focuses on variation within this list of locations, we set potential concerns about endogenous or systematically biased sample selection aside for the moment.

[91]This algorithm arrives at a non-trivial optimum bundle of RGCs being combined into a package due to the non-linear economic benefits and costs of electrification: each new RGC to be connected along the way will add additional cost (a small extension branch line off of the main new distribution line that

Table 5-11 Temporary Electrification Priority of RGCs Based on Demand Criteria (1/13)

Ranking	RGC	District	Priority	Province	# of HHs (2006)	# of HHs (2030)	Daily Max Load
1	Mpulungu Central	Mpulungu	4	Northern	2,000	3,972	2,200,731
2	Mwinilunga Boma	Mwinilunga	0	North-Western	1,900	3,774	2,093,225
3	Shangombo	Shangombo	1	Western	1,100	2,185	1,277,541
4	Boma	Luangwa	1	Lusaka	580	1,152	752,118
5	Chienge	Chienge	6	Luapula	560	1,113	642,046
6	Mpongwe	Mpongwe	22	Copperbelt	441	876	499,270
7	Nsama Sub Boma	Kaputa	1	Northern	441	876	499,270
8	Talayi	Milenge	1	Luapula	202	402	241,663
9	KPG Market	Kapiri Mposhi	22	Central	7,400	14,697	8,141,484
10	Chisanga	Kasama	9	Northern	5,000	9,930	5,530,061
11	Chindenza School	Katete	3	Eastern	5,000	9,930	5,515,225
12	Mtandaza RHC	Katete	5	Eastern	5,000	9,930	5,509,327
13	Kagoro	Katete	1	Eastern	4,000	7,944	4,401,462
14	Sikapila	Mporokoso	2	Northern	3,646	7,241	4,012,551
15	Kauwe	Kazungula	14	Southern	3,511	6,973	3,927,399
16	Palace Chipepo Mukuni-Ngombe	Kapiri Mposhi	23	Central	3,500	6,951	3,847,529
17	Twapia	Ndola	2	Copperbelt	3,333	6,620	3,735,864
18	Nchembwe	Kapiri Mposhi	20	Central	3,100	6,157	3,416,570
19	Kapungwe	Petauke	1	Eastern	3,084	6,125	3,407,622
20	Nyamphinga	Petauke	10	Eastern	3,084	6,125	3,405,327
21	Chikalawa	Petauke	12	Eastern	3,084	6,125	3,402,990
22	Mwanjawanthu	Petauke	3	Eastern	3,036	6,030	3,345,943
23	Kasenengwa Rural Centre	Chipata	5	Eastern	3,000	5,958	3,321,933
24	Kamphambe	Katete	7	Eastern	3,000	5,958	3,308,596
25	Sikatengwa	Lundazi	22	Eastern	2,949	5,857	3,246,409
26	Mushili	Samfa	18	Luapula	2,751	5,464	3,032,798
27	Madimawe Rural Health Centre	Chipata	8	Eastern	2,667	5,297	2,953,674
28	Matonje	Petauke	13	Eastern	2,587	5,138	2,851,518
29	Mumbi	Petauke	5	Eastern	2,503	4,971	2,762,099
30	Kawama East	Mufulira	6	Copperbelt	2,448	4,862	2,732,408
31	Chimutende,	Katete	2	Eastern	2,472	4,910	2,728,967
32	Thendere	Isoka	4	Northern	2,400	4,767	2,698,020
33	Ntipo	Isoka	7	Northern	2,420	4,807	2,685,921
34	Chinkhombe	Katete	13	Eastern	2,385	4,737	2,634,234
35	Mushindomo	Solwezi	21	North-Western	2,380	4,727	2,628,469
36	Mulekatembo	Isoka	5	Northern	2,350	4,667	2,618,045
37	Kaula	Kabompo	14	North-Western	2,350	4,667	2,585,437
38	Lukulu Township	Lukulu	1	Western	2,012	3,996	2,512,048
39	Nande	Senanga	8	Western	2,148	4,266	2,393,483
40	George Camp	Ndola	4	Copperbelt	2,165	4,300	2,383,973

the package (because its addition added more cost than benefit), will enter the pool for bundling with any neighbouring package, or will get allocated to an off-grid solution, and therefore be removed from our sample. Finally, the overall budget available for electrification until 2030 is divided into annual construction allowances that are filled year by year starting from the electrification package with the highest overall financial return. A schematic drawing of this idea is presented in Figure 5: the sketch highlights the existing substation (orange circle), the target location of the given package, i.e. the highest ranked RGC around the substation (red circle), as well as the resulting inconsequential RGCs (blue circles) that are of lower ranking but are accidentally electrified early since they happen to lie on a path between substation and highly ranked RGC. Figure A14 provides one example of how the resulting setting of project packages may look like in practice: for expositional purposes we plot a subset of packages (numbers 6 to 10) in Zambia's Central Province, highlighting each package's highest ranked RGC (black circle around coloured dot), the inconsequential RGCs in that package (coloured dots), and un-selected RGCs that are allocated to off-grid solutions and exit our sample (triangles). Ranking numbers are displayed as labels next to each RGC, confirming that

needs to be constructed), but adds further consumers that will demand electricity, helping to spread the fixed cost of the main new distribution line more widely. See for example the red (main new distribution) line and yellow (branch extension) lines depicted in Figure 5 for a schematic example.

inconsequential RGCs are of vastly lower priority (i.e. a higher ranking number) than the highest ranked RGC, confirming that they should have only been electrified much later were it nor for their status of lying on a path to a highly ranked RGC. Interestingly, Figure A14 also highlights, how the resulting electrification roll-out that resulted from the REMP's ranking and package bundling algorithm does approximate the experimental ideal of inconsequentially electrified RGCs presented in Figure 5 rather well.[92]

A.IV.5 Construction of predicted electrification instrumental variable

The instrumental variables that we use for various regression specifications all stem from the planned year of inconsequential RGCs' electrification via distribution lines. We turn the planned electrification year into a panel dataset with a dummy for planned electrification status, and then aggregate the dummy into a planned electrification percentage at desired geographical unit.

In order to create an annual schedule of package electrification, JICA carried out cost analyses to estimate the overall cost of electrifying each of the 180 project packages. With an annual budget of 50 million USD, REA divided the 180 project packages into 22 "annual project phases" from 2009 to 2030. Generally, the number of packages scheduled for electrification per year decreases as time goes on (e.g. 15 packages are planned for electrification in 2009 while only 7 are planned for 2022).

The actual investment in the rural electrification plan consistently falls behind the planned annual budget of 50 million USD, resulting in an increasing misalignment between the planned and actual year of electrification. Based on available information on previous funding and reasonable assumptions, we make adjustments to planned year of electrification in the data. To transform the adjusted planned year of electrification for inconsequential RGCs to the instrumental variable used in ward-level regressions, we first create a panel dataset containing a dummy variable for planned electrification status that switches from 0 to 1 in the planned year of electrification. To accurately geo-identify RGCs to the best of our ability, we use the GPS coordinates of RGCs as provided by REA. These coordinates allow us to pinpoint the location, and consequently the corresponding ward information, of 1,207 RGCs, out of which 784 are inconsequential RGCs powered by distribution lines. Specifically, we overlay the RGC coordinates with the 2010 Zambian census shapefile to identify the district and ward. We then take the simple average of the planned electrification dummy across all RGCs within a given ward to compute a planned electrification percentage by year. Since the rural electrification

[92]We produce three dozen maps in the style of Figure A14 for all parts of the country, and confirm throughout that planned package electrification is remarkably close to the quasi-experimental ideal.

plan was created in 2009, we assume the planned electrification percentage/dummy to be 0 for all years prior to 2009.

First stage regression results are strong and robust. Similarly, our choice of two way fixed-effects appears to be conservative without qualitatively affecting first stage coefficients or instrument weakness. F statistics of Cragg-Donald weak instrument tests are in the acceptable range with p-values below 5%.

Thus, an alternative empirical strategy is to use the planned (or 'predicted') year of inconsequential RGCs' electrification as an instrumental variable for the actual electrification status derived either from school or health facility level (at the RGC- or ward-level, respectively). In practice, and to exploit the fact that the instrument introduces plausibly exogenous variation in the temporal dimension of the roll-out (i.e. lower ranked RGCs being electrified earlier due to happening to lie on a path to a much higher ranked RGC), instrumental variables specifications are run on top of a RGC-level difference-in-differences strategy, implemented via two-way fixed effects to take the staggered nature of the roll-out into account.

On average, planned electrification at the RGC-level does match actual electrification (derived from schools data) at the RGC-level acceptably well, although some issues are apparent, namely that the planned electrification does advance a lot faster than the actual, in aggregate across all RGCs, from at least 2010 onwards. To account for this overprediction of the speed of actual electrification (which was hindered by budgetary constraints until at least 2012), we adjust the REMP's planned electrification algorithm to use the predicted total budget under current actual expenditure until 2017 (less than USD100m) for a grand total budget of approximately USD300m to be spent by 2030, instead of the originally intended budget of USD1.1bn which construction will hopelessly fall short of. The planned expansion under this more realistic budget still underpredicts the speed of expansion in aggregate, but performs reasonably well across first stage specifications (which take both temporal and spatial variation into account, given the difference-in-differences nature of our preferred first stage specification).