RUHR

ECONOMIC PAPERS

Gunther Bensch
Gunnar Gotz
Jörg Peters

Effects of Rural Electrification on Employment: A Comment on Dinkelman (2011)

rwi

#840

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel, Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #840

Responsible Editor: Manuel Frondel

ISSN 1864-4872 (online) – ISBN 978-3-86788-973-5

Ruhr Economic Papers #840

Gunther Bensch, Gunnar Gotz, and Jörg Peters

Effects of Rural Electrification on Employment: A Comment on Dinkelman (2011)

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/86788973
ISSN 1864-4872 (online)
ISBN 978-3-86788-973-5

Gunther Bensch, Gunnar Gotz, and Jörg Peters[1]

Effects of Rural Electrification on Employment: A Comment on Dinkelman (2011)

Abstract

This paper replicates and extends the seminal paper by Dinkelman (2011) on the impacts of electrification on female employment. We revisit the validity of the identification strategy that uses the land gradient as an instrumental variable (IV). Our robustness checks cast doubt on the exclusion restriction as the IV drives the outcome variable in non-electrified regions. We also demonstrate that it is more difficult to disentangle the effects of electricity and road infrastructure than the original paper claims, because the IV affects both. We additionally highlight that the IV is weak, consequently preventing interpretation of the point estimates that are used throughout the original paper. The concomitance of a questionable exclusion restriction and a weak IV is particularly problematic. We conclude by arguing that the take-aways of the original paper for policy and the academic literature need to be reconsidered. In general terms, our comment shows the difficulties of using geographical variation as a natural experiment for infrastructure evaluation.

JEL-Code: O13, C52, H43, O18

Keywords: Replication; research transparency; energy access; infrastructure; instrumental variables

February 2020

1 Gunther Bensch, RWI; Gunnar Gotz, German Institute for Development Evaluation (DEval), Bonn; Jörg Peters, RWI and University of Passau. *- This paper has benefitted from a reply provided by Taryn Dinkelman. We also thank two anonymous reviewers at the American Economic Review (AER) as well as the editor, Esther Duflo, for very helpful comments. The AER rejected a previous version of the paper in August 2019. The authors are grateful to Thomas K. Bauer, Nathan Fiala, Marc Jeuland, E. Somanathan, and Colin Vance for valuable comments and suggestions.* *– All correspondence to: Jörg Peters, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, e-mail: joerg.peters@rwi-essen.de*

1. Introduction

Development effects of large-scale electrification remain a debated issue. The most prominent study on the topic is the paper by Taryn Dinkelman, published in the *American Economic Review* in 2011. At the center of the paper is an instrumental variables (IV) approach with the interaction of the land gradient and time as the identifying variable. The paper studies labor market outcomes of an electrification roll-out during the post-apartheid era between 1996 and 2001 in ex-homeland communities[1] within South Africa's KwaZulu-Natal province. To justify the IV, the author argues that the South African electricity provider, ESKOM, planned electrification projects primarily based on cost considerations, with the costs of building distribution lines being higher in steeper areas. The gradient is thus assumed to exogenously affect the ordering of community connections over time. The key result of the IV analysis is that rural electrification leads to an increase in employment rates of about 9 percentage points among females aged 15 to 59, relative to a baseline employment rate of about 7%, but no similar effect for men. Additional IV estimations undertaken in Dinkelman (2011) provide suggestive evidence on potential mechanisms, including that less time is spent on firewood collection and more time is spent under electric light.

The paper has had a significant impact on both policy and the academic literature, one that has grown steadily since its publication.[2] Beyond providing an estimate for the effects of electrification on female employment growth, the paper has inspired many researchers to solve the identification challenge in large-scale infrastructure evaluations through a combination of geographical variation and household or enterprise data – in a similar vein as the previous paper on dam constructions by Duflo

[1] Homelands were pockets of land designated for black South Africans during apartheid.

[2] As of January 2020, the paper has 724 citations on Google Scholar (GS), with the number of citations per year still being on the rise (see Annex A1). Among the 199 long papers that were published in the 2011 AER volume, the Dinkelman paper ranked 13th in terms of GS citations as of April 2019. The mean and median GS citation numbers of this volume were 168 and 88, respectively (with 603 citations of Dinkelman 2011 at the time).

and Pande (2007). The use of IVs based on geographical variation has recently been criticized by Lee et al. (2020a). Therefore, as much as Dinkelman (2011) has contributed to promoting empirical research in the field, it is important to critically assess the reliability and robustness of this approach and these influential results. We thus re-examine the analysis in the original paper. Using the taxonomy by Clemens (2017) and Christensen and Miguel (2018), we perform a reanalysis in which we conduct additional tests and revisit the statistical interpretation of the results. Our findings qualify the interpretation of the original results and caution against the widespread use of geographic variation as instruments.

We successfully replicate all tables from Dinkelman (2011) using the dataset provided by the *American Economic Review*. However, we show that the robustness checks provided in the original paper do not substantiate the exclusion restriction and that strong additional assumptions are needed to maintain the causal claim of the original results. First, we demonstrate that the IV drives the key outcome, female employment, not only in electrified regions, but also in non-electrified regions. This questions the exclusion restriction, unless we make specific assumptions about the selection process of communities into treatment status (see Section 2.2 for further discussion of this). Second, we provide evidence that the IV identification fails to disentangle the effects of road and electricity access. Both are (weakly) correlated with the IV and controlling for road access, we argue, is not possible because it is likely to be confounded by the same unobservables that electrification is. The identifying variation is thus only coming from the fact that electrification status varies over time and road access does not. Hence, in order to maintain a causal interpretation of the original results, we need to assume that roads only have a one-time effect on economic development and no effect on the development trajectory. Interpreting road access in the spirit of trade models (see Donaldson 2019) this assumption is similar to assuming that removing trade barriers has only level and no growth effects (Section 2.3). Moreover, we summarize the broader literature on economic geography that conclusively shows that the land gradient affects economic development through different channels, including transport costs (Section 3).

Furthermore, we highlight that the IV is weak according to different tests, consequently preventing interpretation of the point estimates used in the original paper. Dinkelman (2011) implicitly accounts for this by providing weak-IV-robust confidence intervals in the results tables (which turn out to be wide, spanning 3 to 26 percentage points) but only refers to the point estimate of 9 percentage points throughout the paper. Not least, according to the econometrics literature, the concomitance of a questionable exclusion restriction and IV strength is particularly problematic, rendering a reliable interpretation of the key results impossible, even those derived from the confidence intervals (Section 2.1).

We believe that the research question that Dinkelman (2011) tries to answer is important both for policy makers and researchers. Infrastructure expansion and electrification are expensive. According to the International Energy Agency, for Africa alone, the investment required to achieve universal electricity access by 2030 is 31 billion USD annually (IEA 2017). Ensuring that policy makers are making decisions about such expansion based on a full understanding of the quality of the evidence should be a first order goal for researchers.

2. Validity of Dinkelman's land gradient instrumental variable

2.1. Weak instrument tests

Relevance as the first validity criterion of IV-based identification strategies requires that the instrument is sufficiently correlated with the endogenous regressor, conditional on other covariates. If this condition is not fulfilled, the IV yields biased point estimates in finite samples (see Bound et al. 1995). With a first-stage F-statistic of 8 and hence below 10, the IV used in Dinkelman (2011) qualifies as weak by standards that were state-of-the-art at publication of the original paper (Staiger and Stock 1997; Baum et al. 2007). As we show in Annex A3, the IV also does not pass the more recently developed weak instruments test by Montiel Olea and Pflueger (2013). Consequently, the point estimates presented in Dinkelman (2011) are biased.

Dinkelman (2011) does not explicitly test or discuss the weakness of the IV. Instead, the author provides state-of-the-art weak IV robust Anderson-Rubin (AR) confidence intervals and mentions in a footnote that the AR intervals "address concerns about overoptimistic inference with a possibly weak instrument" (Dinkelman 2011, p. 3096). These AR confidence intervals account for the decreasing precision that weak instruments generally induce (Andrews et al. 2019; Nelson and Startz 1990; Murray 2017). From this interval we can learn that the positive effect of electrification on female employment is between 3 and 26 percentage points, so a much wider range than the "9 to 9.5 percentage points" Dinkelman refers to in her introduction, the results section and the conclusion (Dinkelman 2011, p. 3080, 3096, 3105).[3] The methodologically correct interpretation of results thus is that a negative or zero effect of electrification on female employment can be confidently rejected. It is not possible, though, to conclude whether the true effects are modest (i.e. 3 percentage points) or massive (26 percentage points). Moreover, even this interval interpretation is at stake if the exclusion restriction is violated (Andrews et al. 2019), which will be assessed in the following section.

2.2. Exclusion restriction

The second validity requirement of exogeneity imposes that the IV is uncorrelated with any other determinant of the dependent variable, conditional on the included covariates. This is not testable, but robustness tests can underpin or cast doubts on the exclusion restriction (see for example Altonji et al. 2005, Donaldson 2018, and Jaeger et al. 2018). In this section, we revisit all robustness checks conducted in the original paper.

[3] This confidence interval slightly differs from the [0.05; 0.30] interval reported in Table 4 of Dinkelman (2011, p.3095), since we determined the interval bounds with a precision of 0.01, while Dinkelman used a precision of 0.05.

Dinkelman (2011) conducts a placebo test on ex-homeland communities in KwaZulu-Natal that had been connected before the observation period. In an OLS-based regression, she finds no correlation between the outcome variable, female employment, and the IV, land gradient, and interprets this as supporting her exclusion restriction. A problem with this placebo test is that it is a priori inconclusive, because even significant differences found in this test would not have invalidated the identification assumption: Dinkelman's identification assumption also implies that regions with smoother land gradient should have been connected earlier and exposed longer to the treatment than those with a steeper gradient, and hence should be on a different development trend.

An alternative placebo test is to run the same OLS regression using untreated parts of the sample. For this test to provide support to the exclusion restriction, one would expect the IV not to be correlated with the outcome variable. We conduct such a test using the original data, where the untreated parts of the sample are ex-homeland communities in the same province that are still non-electrified in 2001, the end of the observation period. Results are presented in Columns (1) and (2) of Table 1. They indicate a significant correlation between the land gradient and female employment changes in non-electrified areas. In fact, the coefficient of the land gradient variable in this placebo sample is the same as in the main study sample used in Dinkelman's paper (replicated in Columns (3) and (4) in Table 1).

Neither of the two placebo tests can provide full clarity about the validity of the instrument, as they are based on endogenously selected sub-samples. Yet, our test reveals a strong additional assumption that is needed to maintain the exclusion restriction: there must be a confounding variable that drives communities into non-treatment (and hence into our placebo sample) and that is correlated with both the IV and female employment in these communities. Plus, this confounder must not be correlated with the IV and the treatment status in the treated communities (and hence Dinkelman's main sample). This is possible in theory but seems unlikely in practice. In consequence, this additional test suggests that the land gradient also affects labor market outcomes through other channels than electrification (see as well Section 3). For a very strong IV this might not be problematic if one assumes that electrification is still the *main* channel. Yet, as we have shown in Section 2.1., the land gradient is a weak IV.

Table 1: Placebo experiment for effect of gradient on the female employment rate

Dependent variable	Δ female employment rate			
Estimation method	OLS			
Sample	Non-electrified communities (no Eskom project)		All communities	
	(1)	(2)	(3)	(4)
Land gradient x 10	-0.007*	-0.007**	-0.007**	-0.007**
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
	[0.069]	[0.046]	[0.042]	[0.031]
Baseline controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Other Services#	No	Yes	No	Yes
R^2	0.065	0.081	0.064	0.076
N communities	1,451	1,451	1,816	1,816

Notes: In the placebo experiment in Columns (1) and (2), the same specification as in column (1) of Table 6 in Dinkelman (2011) is estimated with the only difference that the analysis is conducted in the sub-sample of non-electrified villages. Columns (3) and (4) reproduce the results from estimating the same specification when using the entire sample from Dinkelman's main analysis. Robust standard errors clustered at village level in parentheses and *p*-values in square brackets. # Other Services refer to water and sanitation access in the communities.

** significant at the 5% level

* significant at the 10% level

3. The land gradient, road access, and economic development

In this section, we argue that Dinkelman's exclusion restriction is a very strong assumption based on the role that the literature ascribes to economic geography, the land gradient, and transportation in driving the trajectory of economic development. We then provide evidence that the land gradient is likely to not only affect electricity roll-out, but also road access.

3.1 The role of roads revisited

While some of the economic geography literature refers to very long-term processes that might be accounted for in Dinkelman (2011)'s regional fixed effects[4], other processes materialize within a time horizon that affects her exogeneity assumption. Transport costs are an important determinant of economic performance (see for example Storeygard 2016). Terrain slope arguably affects transport costs directly (through fuel consumption, for example), but also indirectly through higher construction costs for roads and railways and hence also within the 1996-2001 period. By analogy with Dinkelman's intuition for electricity networks, the land gradient has been used prominently in the literature as an IV for the impacts of transportation infrastructure (see, for example, Donaldson 2018, Shrestha 2019, and Djemai 2018).

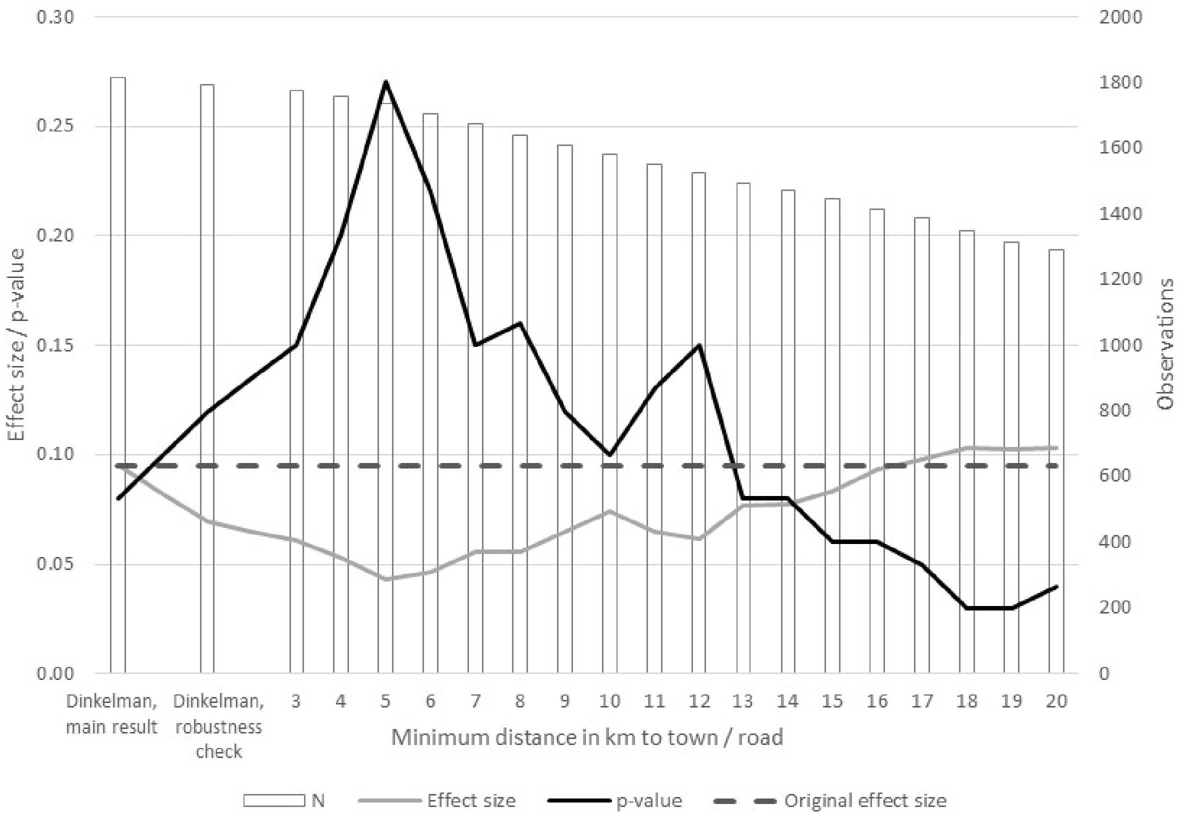
Dinkelman (2011) seeks to account for the potential confounding related to transportation infrastructure in three ways. First, she includes road access as a control variable in the estimations. However, this is not a valid approach: As Deuchert and Huber (2017) show, similar to bad controls in OLS estimations, including post-instrument covariates in a 2SLS regression leads to biased results if these covariates are at the same time correlated with unobservable confounders. Road access is a post-instrument covariate because it is affected by the land gradient and it is correlated with the suspected time-varying unobservables that originally confound the electrification and employment relationship (e.g. business potentials of or political interest in a region).

Second, Dinkelman conducts a robustness check intended to eliminate road access as a confounder by re-running her main IV regression, now excluding those communities that are cut directly by a major national road (Dinkelman 2011, p. 3091). Results of this robustness check confirm her main results to the extent that the effect size does not change, and the significance level only decreases slightly, which the author plausibly ascribes to the declining sample size. Yet, the median distance to a road in the Dinkelman (2011) data set is 33 km. The binary definition of road access adopted in Dinkelman (2011) treats communities at one- or two-kilometers distance to the major

[4] See for example Fenske (2013), Hurtt et al. (2011), and Michalopoulos and Papaioannou (2013).

national road as equally unconnected as communities at 20 or more kilometers distance. Communities near major roads, though, have arguably better access to markets than the remoter ones. We therefore extend Dinkelman's robustness check by running separate regressions where we successively exclude villages with increasing distances between the centroid of the communities and the nearest road.

Figure 1: Effect of electrification on employment depending on distance to roads



Notes: Dinkelman's main results and results from her robustness check that excludes communities cut directly by a major national road are presented on the left of the figure. Since the maximum distance between roads and the centroid of the communities for communities excluded in her robustness test is mostly below 3km, our depiction on the right of the figure starts with the exclusion of villages closer than 3km and goes up to 20km. Across this range, between 21 and 25 percent of dropped sites are treatment units.

Figure 1 shows the results on effect sizes, significance levels, and underlying sample sizes. Dinkelman's results are depicted on the very left side of the horizontal axis. Effect sizes initially decline considerably and reach their minimum when we exclude communities at a road distance of 5 km. Here, the impact estimate is clearly insignificant (*p*-value of 0.27). Interestingly, this 5 km distance coincides with what is considered to most appropriately measure access to roads in Africa (Raballand et al. 2010). The IV estimations only confirm the original results in terms of significant results and similar effect sizes when the distance to roads or towns exceeds 12 km (despite a lower sample size). This pattern suggests that roads and electricity co-determine the outcome and it is not possible to disentangle one from the other.

Third, the identification in Dinkelman (2011) is based on changes in electricity access and employment over time, while road access is time invariant. One could therefore argue that any level effect of gradient and road access is differenced away. Interpreted in this sense, the pattern in Figure 1 would indicate treatment heterogeneity, i.e. electricity affects the outcome differently according to the intensity of road access. However, this interpretation requires another strong assumption, namely, that roads only lead to a one-time level effect. This is possible, but the literature that conceives access to roads and transportation as a reduction in trade costs[5] suggests that it has general equilibrium effects leading to a different development trend in areas with transportation access, also over time. It is therefore more likely that communities closer to roads in the Dinkelman (2011) sample grow at different rates also years after the road connection.

3.2. Land gradient as an IV for road access

In this subsection, we show that the likelihood of having access to a road is as much driven by the land gradient as electricity access is and that both can similarly explain employment growth in the sample. To this end, we now use the land gradient as an IV for road access instead of electricity access (thereby following the intuition of Donaldson 2018, Shrestha 2019, and Djemai 2018). For that purpose, we take Dinkelman's model

$$\Delta y_{jdt} = \alpha_1 + \alpha_2 \Delta \mathrm{T}_{jdt} + X_{jd0}\beta + \lambda_d + \left(\delta_j + \Delta\varepsilon_{jdt}\right),$$

[5] See, for example, Donaldson (2018), Gollin and Rogerson (2014), Jedwab et al. (2017), Redding and Turner (2015), and Storeygard (2016).

which estimates the impact of an electricity project (the dummy T_{jdt}), occurring in community j in district d at time t, on the growth in labor market outcomes over time Δy_{jdt} in that community, with δ_j and $\Delta\varepsilon_{jdt}$ being part of the unobserved error term. We now only swap the roles of electrification and road access. This is, we replace the electrification treatment by baseline road access, $\mathrm{RoadAccess}_{jd0}$, remove baseline road access from the set of control variables, X_{jd0}, and add the electrification status in the respective year to the revised set of control variables, $\tilde{X}_{jdt}$. In the absence of data on over-time variation in road access, we estimate the effect of time-invariant differences in road access,[6] while the original analysis also benefits from over-time variation in electricity access. Our estimation equation thus becomes

$$\Delta y_{jdt} = \alpha_1 + \alpha_2 \mathrm{RoadAccess}_{jd0} + \tilde{X}_{jdt}\beta + \lambda_d + (\delta_j + \Delta\varepsilon_{jdt}).$$

Note that the ambition of this test is not to use land gradient as a well-identified IV for road access, but rather to show that it – technically – works similarly well as Dinkelman's application for electricity infrastructure.

Table 2 presents results for a definition of road access by the 5km distance threshold in line with the previous finding in Figure 1 as well as Raballand et al. (2010). Results are qualitatively identical for any choice of the cutoff between three and 14 km and for continuous, non-linear measures of road access[7] in that road access has an effect on female employment, statistically significant at the 10% level. Note that like in Dinkelman's original analysis the IV is weak and problematic post-instrument covariates are used. The magnitude of the coefficient as well as the width of the weak IV robust AR confidence interval are very similar to Dinkelman's electrification effect. These results underpin that electrification is not the only channel through which land

[6] District fixed effects capture this community-specific effect insufficiently given that there are on average 200 communities per district with different proximities to roads.

[7] As an alternative to the binary measure of road access, one may adopt a continuous measure. If we include road distance in a linear functional form, the effect becomes insignificant. Yet, it is more plausible to assume that benefits of infrastructure diminish over distance and level off at some cut-off distance. For such transformations of road distance that are convex downward, the effect of road distance is significant again across a range of reasonable cut-off distances (e.g. 20km and 50km).

gradient affects employment.

Table 2: IV results for road access as different treatment variable

Dependent variable	Δ female employment rate			
Estimation method	IV (gradient)			
	(1)	(2)	(3)	(4)
Treatment: access to roads within 5 km distance	0.035	0.115	0.104*	0.106*
	(0.058)	(0.075)	(0.059)	(0.058)
	[0.548]	[0.124]	[0.080]	[0.069]
AR 95% Confidence Interval			{0.03; 0.29}	{0.03; 0.28}
Baseline controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Other Services	No	No	No	Yes
"Effective F" statistic (Montiel–Pflueger)	7.03	7.00	7.11	7.77
N communities	1,816	1,816	1,816	1,816

Notes: All model specifications are equal to columns 5 to 8 of table 4 from the original study with the only difference that the instrumented variable is access to roads instead of electrification and that we excluded binary road access from the control variables while adding electrification status; robust standard errors clustered at village level in parentheses and *p*-values in square brackets.
* significant at the 10% level

4. Conclusion

The pattern of results described in Dinkelman (2011) has stimulated much useful discussion and research on the economic impacts of electricity provision. Yet, this comment has shown that a causal interpretation of the results requires much stronger assumptions than the original paper acknowledges. More specifically, we have pointed out that in order to disentangle the effect of electrification and road access, one must assume that roads only have a one-time effect on employment levels and no effect on the development trend. One could stress these assumptions, resorting to particularities of the post-Apartheid homeland sample in South Africa. This would be at the expense of an already very limited external validity. More generally, our comment amplifies concerns expressed in Lee et al. (2020a) about using geographic variation to isolate the contribution of electrification from other infrastructure and simultaneous factors.

Beyond potential additional assumptions, the weakness of the IV makes the original paper's results hard to interpret. While providing weak IV robust confidence intervals in the result tables was the best Dinkelman (2011) could do to account for a weak IV in her analysis, these wide intervals merely back an interpretation that electrification has non-zero effects on female employment. A more explicit discussion in the original paper of whether the point estimates can be used instead would have been desirable.

Our comment does not generally question the positive effect of electricity on economic development. But a growing literature based on more clearly identified program evaluations (including a study conducted by an author of this comment) suggests that effects in rural Africa are smaller than expected by governments and donor organization (see Chaplin et al. 2017, Lee et al. 2020b, and Lenz et al. 2017). According to Bos et al. (2018) and Peters and Sievert (2016), both review studies, this contrasts with more positive evidence from Latin America and Asia (see Grogan and Sadanand 2013, Grogan 2018, Kassem 2019, Khandker et al. 2012, Lipscomb et al. 2013, Rud 2012, and van de Walle et al. 2017). Next to this regional demarcation, a different perspective on the variation in the literature is that experimental studies come to more conservative conclusions compared to observational studies (Bayer et al. 2019). Lee et al. (2020a) point out that most of these latter studies are based on IVs and indeed Burlig and Preonas (2016) raise doubts about the positive effect of rural electrification in India, using a well-identified identification.

Yet another contribution of the Dinkelman (2011) study was to compile an impressive number of datasets that was unique at the time and that allowed for investigating many interesting outcomes beyond employment (a reconstruction of Dinkelman's Theory of Change and identification strategies on each level can be found in Annex A2). A broader look into this might actually reconcile the Dinkelman (2011) data with the newer evidence in that there is no increase in wages and labor demand from firms due to general electricity expansion. The income and welfare effects are thus ambiguous. Dinkelman (2011)'s findings on time use – very much driven by electric cooking, which is another particularity of South Africa – in turn reconcile with the improved cooking literature (Jeuland and Pattanayak 2012, Bensch and Peters 2015, Pattanayak et al. 2019). The focus on the IV-based female employment results might even obscure these important insights.

In sum, following Brown and Wood (2019), we believe that replications and reanalyses like ours should aim at providing additional information on the original study and thereby qualifying more carefully the original results. The degree of confidence one concedes to an IV approach is always subjective. Yet, we believe that if the original paper had been more explicit about the caveats presented in this comment, its interpretation by the academic literature and policy makers would have been different. Especially in such a policy-relevant field like development economics, the use of empirical methods requires careful and transparent communication to the reader.

References

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human Resources*, 40(4), 791-821.

Andrews, I. Stock, J. & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11, 727-753.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal*, 7, 465-506.

Bayer, P., Kennedy, R., Yang, J., & Urpelainen, J. (2020). The need for impact evaluation in electricity access research. *Energy Policy*, forthcoming.

Bensch, G., & Peters, J. (2015). The intensive margin of technology adoption–Experimental evidence on improved cooking stoves in rural Senegal. *Journal of Health Economics*, 42, 44-63.

Bos, K., Chaplin, D., & Mamun, A. (2018). Benefits and challenges of expanding grid electricity in Africa: A review of rigorous evidence on household impacts in developing countries. *Energy for Sustainable Development*, 44, 64-77.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430), 443-450.

Brown, A. N., & Wood, B. D. (2019). Replication studies of development impact evaluations. *Journal of Development Studies*, 55(5), 917-925.

Burlig, F., & Preonas, L. (2016). Out of the darkness and into the light? development effects of rural electrification. Energy Institute at Haas WP, 268.

Chaplin, D., Mamun, A., Protik, A., Schurrer, J., Vohra, D., Bos, K., Burak, H., Meyer, L., Dumitrescu, A., Ksoll, C. and Cook, T., (2017). Grid electricity expansion in Tanzania by MCC: Findings from a rigorous impact evaluation. Report submitted to the Millennium Challenge Corporation, Washington, DC: Mathematica Policy Research.

Chernozhukov, V., & Hansen, C. (2008). The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, 100(1), 68-71.

Clemens, M. A. (2017). The meaning of failed replications: A review and proposal. *Journal of Economic Surveys*, 31(1), 326-342.

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3), 920-80.

Deuchert, E., & Huber, M. (2017). A cautionary tale about control variables in IV estimation. *Oxford Bulletin of Economics and Statistics*, 79(3), 411-425.

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

Djemai, E. (2018). Roads and the spread of AIDS in Africa. *Journal of Health Economics*. 60, 118–141

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5), 899-934.

Duflo, E., & Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, *122*(2), 601-646.

Fenske, J. (2013). Does land abundance explain African institutions? *The Economic Journal*, 123(573), 1363-1390.

Gollin, D., & Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, 107, 38-48.

Grogan, L., & Sadanand, A. (2013). Rural electrification and employment in poor countries: Evidence from Nicaragua. *World Development*, 43, 252-265.

Grogan, L. (2018). Time use impacts of rural electrification: Longitudinal evidence from Guatemala. *Journal of Development Economics*, *135*, 304-317.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., ... & Jones, C. D. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, 109(1-2), 117.

IEA (2017). Energy Access Outlook 2017. The International Energy Agency (IEA), Paris.

IEA, IRENA, UNSD, WB, WHO (2019). *Tracking SDG 7: The Energy Progress Report 2019.* Washington D.C.

Jaeger, D., T.J. Joyce & R. Kaestner (2018). A cautionary tale of evaluating identifying assumptions: Did reality TV really cause a decline in teenage childbearing? *Journal of Business and Economic Statistics*, 1-10.

Jedwab, R., Kerby, E., & Moradi, A. (2017). History, path dependence and development: Evidence from colonial railways, settlers and cities in Kenya. *Economic Journal*, 127(603), 1467-1494.

Jetter, M. (2017). The effect of media attention on terrorism. *Journal of Public Economics*, 153, 32-48.

Jeuland, M. A., & Pattanayak, S. K. (2012). Benefits and costs of improved cookstoves: assessing the implications of variability in health, forest and climate impacts. *PloS One*, 7(2), e30338.

Kassem, D. (2019). Does electrification cause industrial development? Grid expansion and firm turnover in Indonesia. Mimeo.

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

Kovandzic, T. V., Schaffer, M. E., Vieraitis, L. M., Orrick, E. A., & Piquero, A. R. (2016). Police, crime and the problem of weak instruments: Revisiting the "more police, less crime" thesis. *Journal of Quantitative Criminology*, 32(1), 133-158.

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,* forthcoming.

Lenz, L., Munyehirwe, A., Peters, J., & Sievert, M. (2017). Does large scale infrastructure investment alleviate poverty? Impacts of Rwanda's electricity access roll-out program, *World Development*, 89, 88-110.

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

Michalopoulos, S., & Papaioannou, E. (2013). National institutions and subnational development in Africa. *Quarterly Journal of Economics*, 129(1), 151-213.

Montiel Olea, J.L., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3), 358-369.

Murray, M.P. (2017). Linear model IV estimation when instruments are many or weak. *Journal of Econometric Methods*, 6(1), 1-22.

Nelson C., & Startz, R. (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica*, 58, 967-976.

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, 94(1), 20-36.

Pattanayak, S. K., Jeuland, M., Lewis, J. J., Usmani, F., Brooks, N., Bhojvaid, V., ... & Ramanathan, N. (2019). Experimental evidence on promotion of electric and improved biomass cookstoves. *Proceedings of the National Academy of Sciences*, 201808827.

Peters, J., & Sievert, M. (2016). Impacts of rural electrification revisited–the African context. *Journal of Development Effectiveness*, 8(3), 327-345.

Pflueger, C. E., & Wang, S. (2015). A robust test for weak instruments in Stata. *Stata Journal*, 15(1), 216-225.

Raballand, G., Macchi, P., & Petracco, C. (2010). *Rural Road Investment Efficiency: Lessons from Burkina Faso, Cameroon, and Uganda*. Washington, DC: World Bank.

Redding, S. J., & Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. In: Handbook of Regional and Urban Economics (Vol. 5, pp. 1339-1398).

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

Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of Political Economy*, 126(2), 850-901.

Shrestha, S. A. (2019). Roads, participation in markets, and benefits to agricultural households: Evidence from the topography-based highway network in Nepal. *Economic Development and Cultural Change*. Forthcoming.

Staiger, D. O., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 557-586.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*.

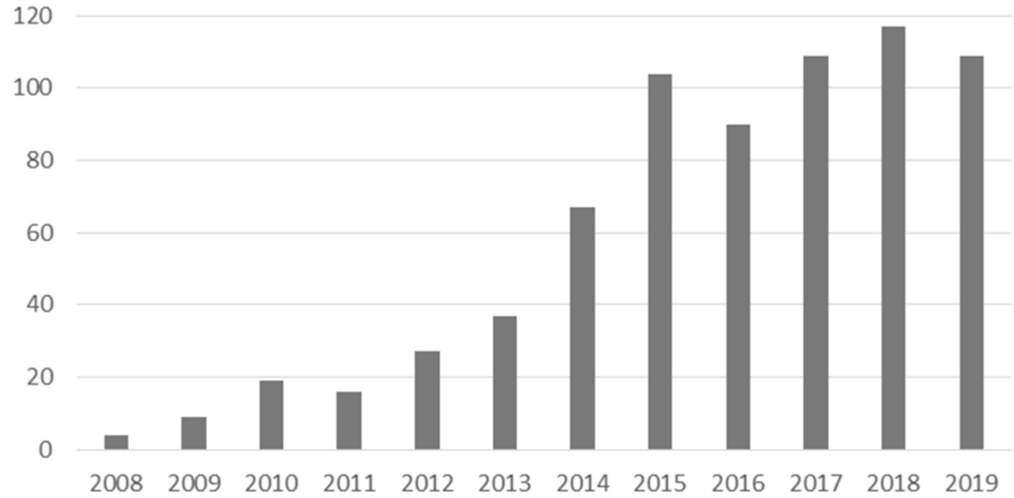
Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *Review of Economic Studies*, *83*(3), 1263-1295.

van de Walle, D., Ravallion, M., Mendiratta, V., & Koolwal, G. (2017). Long-term gains from electrification in rural India. *World Bank Economic Review*, 31(2), 385-411.

Annex

Annex A1: Citation record of Dinkelman (2011)

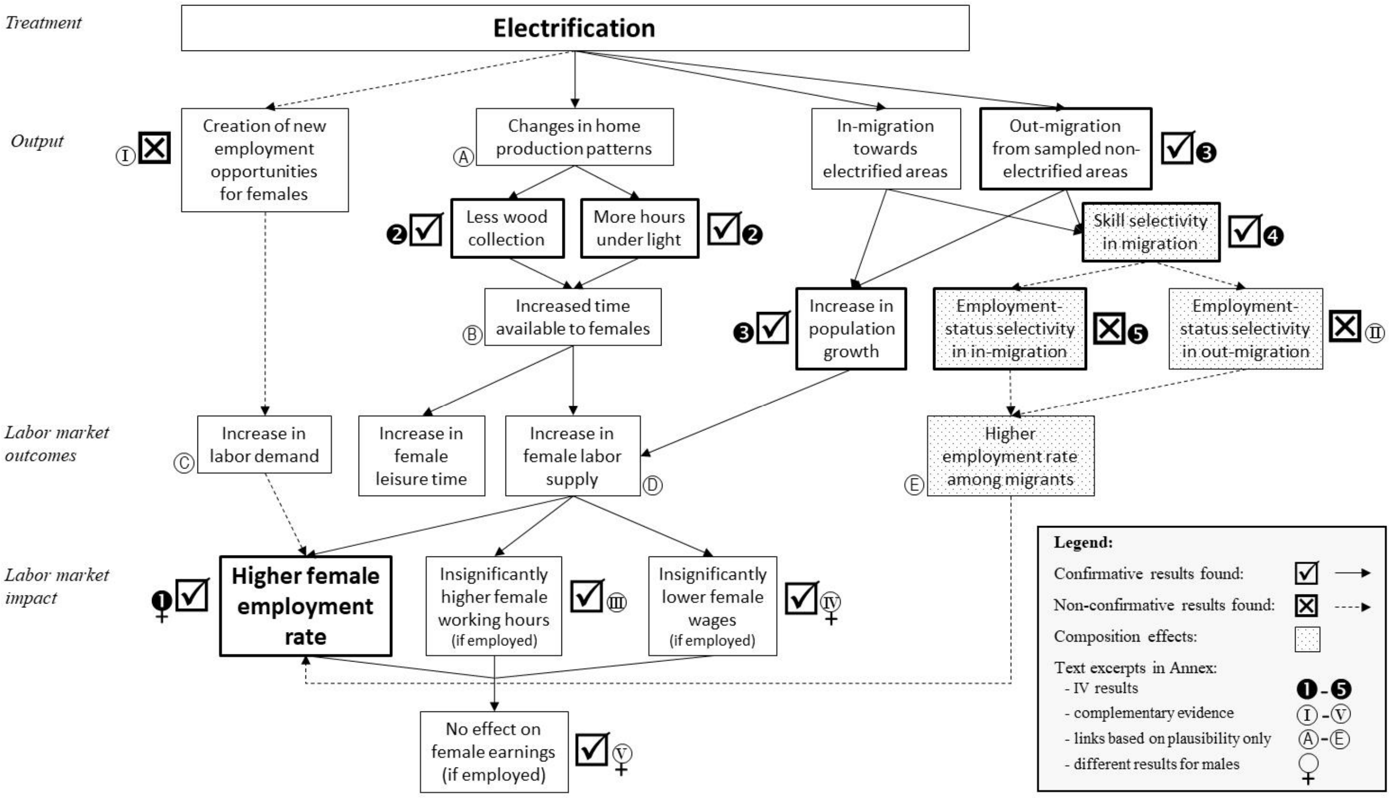
Figure A1: Google Scholar Citations of Dinkelman (2011)



Note: Including citations of the working paper version. Accessed on January 24h, 2020.

Annex A2: Theory of change analysis

Figure A2: Theory of change underlying Dinkelman (2011)



Note: Table A1 provides an extensive account of how the different elements of the Theory of Change are substantiated in Dinkelman (2011). Complementary evidence refers to results that are not based on the main IV framework using the large-sample community panel but smaller-sample fixed effects or IV estimations or cross-sectional data analysis.

Table A1: Text excerpts and quantitative results that lay the foundation for Dinkelman's Theory of Change

Element of the Theory of Change	Text excerpts from Dinkelman (2011)	Quantitative results from Dinkelman (2011)	Analyses for males
❶ Higher female employment	"Results from both analyses show that employment in rural KZN increases in the wake of electrification. Female employment measured in the census rises by a significant 9 to 9.5 percentage points […]" (Dinkelman 2011, p. 3080)	Main IV regressions with controls (N=1816): - treatment coefficient of 0.095* for female employment, with *p*=0.08 (Dinkelman 2011, Table 4, column 8; mainanalysis_communitydata.do)	Same estimations for males yield insignificant results (*p*=0.59) (Dinkelman 2011, Table 5, column 8; mainanalysis_communitydata.do)

Element of the Theory of Change	Text excerpts from Dinkelman (2011)	Quantitative results from Dinkelman (2011)	Analyses for males
❷ Less wood collection and ❷ More hours under light and Ⓐ Changes in home production patterns	"[…] results illustrate substantial shifts towards using electricity for home production […]. Average rates of electric lighting rise […], reliance on wood for cooking falls […], and cooking with electricity rises […]" (Dinkelman 2011, p. 3101)	IV regressions with controls and household energy sources as dependent variables (N=1816): - 23 pp larger increase in the share of hh that cook with electricity** - 63 pp larger increase in the share of hh that light with electricity*** - 28 pp larger decrease in the share of hh that cook with wood* (Dinkelman 2011, Table 8, rows 1 to 3)	No differentiation by gender
Ⓑ Increased time available to females	"[…] the results […] suggest that one important channel through which electricity affects the rural labor market is by "freeing up" women's time for the market." (Dinkelman 2011, p. 3101)	-	Only implicit differentiation by gender (household chores are women's tasks)
Ⓘ Creation of new employment opportunities for females and Ⓒ Increase in labor demand	"[…] any electricity project that generates new firms and new demand for labor should have spatial spillover effects into neighboring areas. […] positive spillovers in these non-electrified areas would dampen any effects of household electrification. […] IV coefficients are large, positive, and close to the main IV estimate […] Using this test, there is no evidence of large spillovers across communities." (Dinkelman 2011, p. 3100) "I rule out the possibility that household electrification stimulated large scale rural industrialization and hence a shift in labor demand by showing the absence of cross-community employment spillovers." (Dinkelman 2011, p. 3080) "The fact that […] provides additional evidence that electrification did not spark large increases in the demand for labor through rural industrialization." (Dinkelman 2011, p. 3105)	IV regressions with controls after excluding nonelectrified areas within a one- (N=-34%) and five-kilometer (N=-54%) radius of an electrified area: - female employment increases insig. (*p*=0.19 and *p*=0.37 for the one and five km definition, respectively) - these coefficients are not significantly different from the coefficient in the main IV estimations (see ①) (Dinkelman 2011, Table 9, mainanalysis_communitydata.do)	No basis for analysis of spillover effects given the insignificant results in the main IV estimations (see ①).

Element of the Theory of Change	Text excerpts from Dinkelman (2011)	Quantitative results from Dinkelman (2011)	Analyses for males
	"[…], it is implausible that household electrification created jobs by sparking the industrialization of rural KZN." (Dinkelman 2011, p. 3100).		
❸ Increase in population growth and ❸ Out-migration from sampled non-electrified areas	"A final channel through which electrification may affect measured employment growth is through migration. […] electrified areas have significantly higher population growth rates than non-electrified areas" (Dinkelman 2011, pp. 3102-3103) "growth of the incumbent population (excluding recent in-migrants […]) remains higher in areas that receive an Eskom project by virtue of gradient." (Dinkelman 2011, p. 3104) "[…] differential out-migration, while substantial, […]" (Dinkelman 2011, p. 3081)	IV regressions with controls and population growth as dependent variable (N=1816): - 390% higher overall population growth*** - 435% higher growth of non-in-migrant population*** (Dinkelman 2011, Table 10, column 2)	No heterogeneity across sexes can be found when running the same analyses for both sexes separately (own calculations)
Ⓓ Increase in labor supply	"This suggests that household electrification operates as a labor-saving technology shock to home production in rural areas, releasing female time from home to market work. […] As further evidence that electricity stimulated a net increase in labor supply to the market, […]" (Dinkelman 2011, p. 3080)	-	Only implicit differentiation by gender (household chores are women's tasks)
❹ Skill selectivity in migration	"I present some evidence that this type of compositional change is present in my sample. […] A combination of skilled migrants flowing toward flatter areas at higher rates and skilled migrants leaving steeper areas at higher rates could account for these compositional changes." (Dinkelman 2011, p. 3104)	IV regressions with controls and population growth as dependent variable (N=1816): - 13% higher share of females with high school* (Dinkelman 2011, Table 10, Panel A, column 4)	Same analysis for males yields insignificantly higher share of males with high school by 8% (*p*=0.21) (Dinkelman 2011, Table 10, Panel A, column 6; mainanalysis_communitydata.do)

Element of the Theory of Change	Text excerpts from Dinkelman (2011)	Quantitative results from Dinkelman (2011)	Analyses for males
❺ Employment-status selectivity in in-migration	"[…] differential in-migration can be ruled out as a confounder of the employment results […] electrification effects […], if anything, are larger for incumbent women, […]" (Dinkelman 2011, p. 3104)	IV regressions with controls and redefined employment growth rate as dependent variable, where recent in-migrants are excluded from both the numerator and denominator (N=1816): - treatment coefficient of 0.116* for female employment (compared to 0.095* in the main IV estimation) (Dinkelman 2011, Table 10, Panel B, column 4)	Same analysis for males yields treatment coefficient of 0.086 (p=0.13) for male employment (Dinkelman 2011, Table 10, Panel B, column 6; mainanalysis_communitydata.do)
Ⓘ Employment-status selectivity in out-migration	"[…] out-migrants from rural KZN […] are significantly *less* likely to be employed, relative to incumbents. Other researchers have also documented these facts." (Dinkelman 2011, p. 3104)	Cross-sectional data from a migration module included in the 2002 September Labour Force Survey: - out-migrants have significantly lower rates of employment (18%) than incumbents (36-38%) (Dinkelman 2011, online Appendix 3, Table 9)	No differentiation by gender
Ⓔ Higher employment rate among migrants	"[…], if in-migrants to electrifying areas already have jobs elsewhere or if out-migrants from nonelectrified areas take their jobs with them, we might mistakenly attribute employment growth to new household electrification, when the main effect of the roll-out is merely to change the composition of the community." (Dinkelman 2011, p. 3082) "[…] given the profile of out-migrants and the results for incumbent-only employment rates, we can conclude that even this type of migration in response to electrification cannot account for all of the employment effects of electrification documented in Section V." (Dinkelman 2011, p. 3105)	-	No differentiation by gender
Ⓜ Higher female working hours (if employed)	"[…] none of the electrification coefficients is precisely estimated in this small sample. […] Women work […] more […] in MDs with	FE regressions of magisterial district (N=146): - women's working hours increase insig. (p=0.19) by	Same analysis for males yields an insig. working hours

Element of the Theory of Change	Text excerpts from Dinkelman (2011)	Quantitative results from Dinkelman (2011)	Analyses for males
	higher electrification rates, compared to the same MDs in periods of lower electrification." (Dinkelman 2011, p. 3099)	about 1.3 hours per week for the average change in electrification rate (15%) (Dinkelman 2011, Table 7, Panel B, column 6; p.3099; supplanalysis_hhsurveydata.do)	increase by about 1.6 hours per week (p=0.32) (Dinkelman 2011, Table 7, Panel B, column 8; p.3099; supplanalysis_hhsurveydata.do)
Ⓘⓥ Lower female wages (if employed)	"[…] female wages fall (albeit imprecisely) in districts where electrification is expanding more rapidly." (Dinkelman 2011, p. 3080)	FE regressions of magisterial district (N=146): - women's wages fall insig. (p=0.20) by about 20% for the average change in electrification rate (15%) (Dinkelman 2011, Table 7, Panel C, column 2; p.3100; supplanalysis_hhsurveydata.do)	Same analysis for males yields no effect on wages (p=0.73) (Dinkelman 2011, Table 7, Panel C, column 4; p.3100; supplanalysis_hhsurveydata.do)
Ⓥ No effect on female earnings	"[…] there are no significant differences in female earning across electrifying and nonelectrifying areas […] or within an MD that sees growing electrification over time […]" (Dinkelman 2011, p. 3100)	FE regressions of magisterial district (N=146): - women's earnings fall insig. (p=0.54) by about 9% for the average change in electrification rate (15%) (Dinkelman 2011, Table 7, Panel D, column 6; supplanalysis_hhsurveydata.do)	Same analysis for males yields 16% higher earnings (p=0.02) (Dinkelman 2011, Table 7, Panel D, column 8; supplanalysis_hhsurveydwata.do)

Note: FE = fixed effects, hh = household, IV = instrumental variables, MD = magisterial district, N = number of observations, p = p-value.

Annex A3: Weak instrument analysis

In general, first-stage *F*-statistics are used to assess the relevance of an IV. The original paper shows the first-stage regressions and presents *F*-statistics, yet without discussing them. Dinkelman's *F*-statistics correspond to the Kleibergen-Paap rank Wald *F*-statistic. According to Baum et al. (2007) and Staiger and Stock (1997) this *F*-statistic should be higher than 10 in order to reject weak identification, as it is the case in Jetter (2017), for example. The *F*-statistics in Dinkelman's case, however, all fall below this threshold (see Table A2).

Table A2: Dinkelman's main results and first-stage statistics related to weak instrument testing

Dependent variable	Δ female employment rate			
Estimation method	IV (gradient)			
	(1)	(2)	(3)	(4)
Treatment: Electrification	0.025	0.074	0.090*	0.095*
	(0.045)	(0.060)	(0.054)	(0.055)
	[0.583]	[0.215]	[0.097]	[0.083]
AR 95% Confidence Interval‡			{0.02; 0.25}	{0.03; 0.26}
Baseline controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Other Services#	No	No	No	Yes
N communities	1,816	1,816	1,816	1,816
First-stage F-statistic (Kleibergen-Paap rank Wald)	4.20	4.87	8.33	8.25
´Effective F´-statistic (Montiel–Pflueger)	4.22	4.89	8.36	8.28
Weak IV critical value for 10% of Worst Case Bias	23.11			
Weak IV critical value for 30% of Worst Case Bias	12.04			

Notes: All model specifications are equal to columns 5 to 8 of table 4 from the original study; robust standard errors clustered at village level in parentheses and *p*-values in square brackets. # Other Services refer to water and sanitation access in the communities. ‡ Even though the same approach has been adopted (see Chernozhukov and Hansen 2008), the presented confidence intervals slightly differ from the [0.05; 0.30] interval reported in Table 4 of Dinkelman (2011, p.3095), since we determined the interval bounds with a precision of 0.01, while Dinkelman used a precision of 0.05.
*significant at the 10% level.

A statistically more substantiated weak instruments test has been developed after the publication of Dinkelman's study by Montiel Olea and Pflueger (2013). The authors propose that an IV is weak if there is a 5% chance that the bias in the IV estimator is

10% of a ‘worst case’ scenario in which the IV is assumed to be completely irrelevant (see also Ramey and Zubairy 2018, Kovandzic et al. 2015, and Andrews et al. 2019). Their so-called ´effective´ *F*-statistic and the critical value for a 10% bias are shown in Table A2. These thresholds follow the same logic as the older standard approach by Stock and Yogo (2005). As can be taken from the table, the *F*-statistics for all of Dinkelman’s IV specifications not only fall below this critical value, but even below the threshold of 30%, the highest threshold level reported by Pflueger and Wang’s (2015) Stata command *weakivtest*. In sum, by all standards, the IV used in Dinkelman (2011) qualifies as weak.