RUHR
ECONOMIC PAPERS

T. Robert Fetter
Faraz Usmani

Fracking, Farmers, and Rural Electrification in India

#864

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

Editorial Office

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

Ruhr Economic Papers #864

Responsible Editor: Manuel Frondel

ISSN 1864-4872 (online) – ISBN 978-3-96973-001-0

Ruhr Economic Papers #864

T. Robert Fetter and Faraz Usmani

Fracking, Farmers, and Rural Electrification in India

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/10.4419/96973001
ISSN 1864-4872 (online)
ISBN 978-3-96973-001-0

T. Robert Fetter and Faraz Usmani[1]

Fracking, Farmers, and Rural Electrification in India

Abstract

The shale gas revolution in the United States induced an unprecedented commodity boom across northwestern India. Leveraging population-based discontinuities in the contemporaneous roll-out of India's national rural electrification scheme, we show that access to electricity increased total employment and non-agricultural employment in villages affected by this exogenous economic shock, but had no impact on labor markets elsewhere. This combination of two natural experiments highlights how complementary economic conditions drive heterogeneity in the labor-market impacts of rural electrification. It also helps explain the large variation in the reported impacts of such resource-intensive infrastructure investments globally.

JEL-Code: H54, O13, O15, O18, Q40, Q56, R23

Keywords: Rural electrification; heterogeneous impacts; labor markets; productive use; economic development; regression discontinuity; India

September 2020

1 T. Robert Fetter, Nicholas Institute for Environmental Policy Solutions, Duke University, Durham, NC, USA; Faraz Usmani, International Unit, Mathematica, Washington, DC, USA. *– We thank Fiona Burlig, Nathan Chan, Peter Christensen, Esther Heesemann, Marc Jeuland, Matthew Johnson, David Kaczan, Meera Mahadevan, Robyn Meeks, Brian Murray, Jennifer Orgill-Meyer, Subhrendu Pattanayak, Louis Preonas, Seth Sanders, Steven Sexton, Smriti Sharma, Saurabh Singhal, Eswaran Somanathan, and seminar participants at Camp Resources XXIV, Duke University, North Carolina State University, UNU-WIDER, SETI 2018, WCERE 2018, ASSA 2019, MWIEDC 2019, AERE 2019, the 2019 annual conference of the German Economic Association's Research Group on Development Economics, NEUDC 2019 and the 2019 Occasional Workshop in Environmental and Resource Economics for valuable comments and suggestions. We also thank Lauren Masatsugu for her help acquiring data that made this project possible. Usmani is grateful for support from the Duke Global Health Institute, Duke University Energy Initiative, Duke University Department of Economics, the United Nations University World Institute for Development Economics Research (UNU-WIDER), and the Cornell Atkinson Center for Sustainability. Any remaining errors are our own.* *- All correspondence to: Faraz Usmani, International Unit, Mathematica, Washington, DC, USA, e-mail: fusmani@mathematica-mpr.com*

1 Introduction

Over a billion people worldwide lack access to electricity, and many more are served by unreliable systems capable of supporting little more than a light bulb. The belief that access to reliable electricity catalyzes job creation and economic growth—reflected in the inclusion of energy access targets as part of the United Nations' (2015) Sustainable Development Goals—has thrust energy to the fore of development policy. Indeed, governments and donors are mobilizing considerable resources to ensure access for all. According to the International Energy Agency (2011), over $9 billion was spent in 2009 to extend energy services to underserved populations, a figure that it estimates must rise to over $48 billion per year by 2030 to achieve universal access. Yet evidence on the impacts of such efforts remains mixed. Dinkelman (2011) and Lipscomb et al. (2013), for instance, identify large positive effects on employment from rural electrification in South Africa and Brazil, respectively. Burlig and Preonas (2016), on the other hand, find far more muted effects on labor-market outcomes in India. Others have uncovered similarly lackluster impacts in the African context (Bernard and Torero, 2015; Lenz et al., 2017).[1]

This lack of consensus surrounding the benefits of grid expansion highlights both a significant knowledge gap and a critical policy challenge. The world's poor are constrained by far more than energy poverty (Banerjee and Duflo, 2007), and large-scale investments in energy infrastructure in low- and middle-income settings may entail profound opportunity costs. India alone is home to nearly 250 million people living without electricity (International Energy Agency, 2015). If resource-intensive grid expansion is foundational in promoting better livelihoods for unconnected populations, it represents a necessary first step for development policy. But if the benefits are highly uncertain—or, worse, illusory—scarce public resources are better targeted elsewhere, and cost-effective approaches that enhance access to only rudimentary energy services such as basic lighting may be more appropriate (Burgess et al., 2019; Grimm et al., 2017, 2020).

What drives this heterogeneity in the impacts of rural electrification, and under what conditions does grid expansion deliver measurable economic benefits? We

[1] In a recent review, Bonan et al. (2017) note that the evidence on the impacts of electrification on time allocation and labor activities suggests "mild increases in employment and labor supply, particularly for women, non-agricultural activities and more formal activities" but that the magnitude of such effects "varies significantly across studies and geographical areas." Other syntheses of the literature uncover similar heterogeneity (Bos et al., 2018).

exploit two concurrent natural experiments to shed new light on this question. As the hydraulic fracturing ("fracking") boom began in the United States (US), it induced a parallel commodity boom in India in the production of an otherwise obscure crop called guar. Guar provides a key input into the fracking process, and over sixty percent of the world's supply is grown in the semi-arid tracts of northwestern India by small and marginal farmers (Rai, 2015). Between 2007 and 2012, its price increased by nearly 1,000 percent, resulting in a large exogenous shock to rural villages in the region. Almost simultaneously, India began rolling out a massive rural electrification scheme that aimed to electrify 400,000 villages across 27 states. Villages were eligible on the basis of a strict population-based threshold, giving rise to discontinuous changes in their probability of being electrified. We combine these two natural experiments within a regression discontinuity (RD) design to evaluate how the causal effect of electrification on labor-market outcomes varies with exogenous changes in economic context.

Using data on the sectoral composition of the rural labor force from the 2001 and 2011 rounds of India's Population Census, we first show that electrification led to a six percentage-point (seventy percent) increase in non-agricultural employment in the short term in villages exposed to the exogenous commodity boom. In these same villages, agricultural employment fell by a corresponding amount, representing a reduction of about twenty percent. Using data on total employment from four consecutive rounds of the Economic Census of India, which covers all non-farm establishments in the country, we next demonstrate that these effects persisted over the longer term. Specifically, around eight years after the start of rural electrification, total employment was over 1.5 times higher in electrified villages that experienced the exogenous commodity shock. Importantly, we find no discernible evidence of either short- or longer-term labor-market effects of electrification in villages located in the rest of India, suggesting that complementary economic conditions play a crucial role in driving the impacts of large-scale electrification infrastructure.

In so doing, we build on work by Burlig and Preonas (2016), who conduct the first large-scale impact evaluation of India's rural electrification scheme. They show that the program increased electrification rates, but demonstrate that its impacts on a wide range of socioeconomic outcomes (including those related to the rural labor market) are precisely estimated null results.[2] Our results from non-boom regions of India—using an empirical strategy that follows their own—are consistent with these earlier findings. Using the exogenous shock to economic activity generated by the

[2]Results from a randomized controlled trial in Kenya by Lee et al. (2020b) echo these findings.

guar boom, however, allows us to address questions that stem from this prior body of work, and respond to calls for research that rigorously sheds light on important drivers of heterogeneity in the impacts of electrification globally (Lee et al., 2020a).

In particular, we highlight potential mechanisms for these heterogeneous effects by documenting that electrification-related labor-market dynamics are driven by the rise of complementary non-agricultural opportunities. Increased demand for guar gum spurred a shift in the labor force toward industrial-scale guar processing, which benefits from upgrades to local electricity infrastructure. Consistent with this shift, we use a separate "quadruple-differences" empirical strategy to uncover a large increase in both the number and size of non-farm establishments related to the industrial (electricity-intensive) parts of the guar production chain, such as guar processing, in regions where investments in electrification infrastructure happened alongside the exogenous increase in economic opportunity.

Our study makes three key contributions. First, our results highlight how grid-scale electrification can support potentially welfare-enhancing structural change in the rural economy. Access to electricity alone may not deliver economic and social benefits, as has been demonstrated a number of times in the literature. That electrification significantly enhances non-agricultural employment in boom areas suggests, however, that it can enable individuals, households and firms to better exploit the opportunities presented by rapidly changing economic contexts.

Second, we show that the impacts of large-scale investments in grid electrification are crucially tied to local economic conditions. For instance, grid electricity may enable local industrial production of certain goods, yet this may make little difference in the short run if complementary factors (such as demand for those goods, a trained labor force to scale up production to meet that demand, and rural roads that enable access to markets) are not also in place. If they are, however, grid-scale electricity may considerably expand how local actors take advantage of economic opportunities to generate income and enhance welfare. Prior research, which typically estimates the average treatment effect of such investments as part of national rural electrification programs, implicitly neglects these context-specific factors.[3] While the particular boom we study is clearly unique to our setting, it gives us an opportunity to investigate how electrified villages in boom and non-boom regions performed relative to unelectrified villages in the same areas. Insofar as the economic potential of certain areas can be accurately assessed *ex ante*—or if

[3]This, we contend, is one reason we observe mixed evidence from settings as diverse as Bhutan, Brazil and Vietnam (Khandker et al., 2013; Lipscomb et al., 2013; Litzow et al., 2019).

governments can complement electrification initiatives with other investments—the insights we generate can be used to inform spatial targeting of resource-intensive infrastructure by allowing policymakers to better gauge cost-benefit trade-offs, and choose appropriate grid-based and off-grid energy solutions for different contexts.

Finally, from a methodological perspective, our study is part of a growing body of work that adopts a rigorous approach to understanding treatment-effect heterogeneity in the real world.[4] That the same intervention can have different impacts in superficially similar settings points to the importance of context-dependence; learning about these contextual factors is crucial to learning from impact evaluations (Usmani et al., 2018; Vivalt, 2015). Where many studies have been conducted, rigorous meta-analyses can shed light on underlying drivers of effectiveness (e.g., Meager, 2019). In most other cases, however, such efforts are typically restricted to relatively crude subgroup analyses involving interactions of endogenous binary variables representing populations of interest with the main treatment-effect parameter. Our setting—the combination of an exogenous shock to economic activity with quasi-experimental variation in access to electricity within an RD design—provides the first opportunity to study the heterogeneous effects of access to electricity over large spatial scales in a real-world setting.

The rest of this paper is organized as follows. In Section 2, we provide background on our two natural experiments. Section 3 highlights our conceptual framework and identification strategies. Section 4 describes our data. Section 5 reports short-term impacts of electrification on the size and composition of the rural labor force. Section 6 presents longer-term impacts on total employment. Section 7 documents additional analyses to uncover mechanisms related to the growth of firms. Section 8 summarizes results, and discusses policy implications and avenues for future research.

2 Background

In this section we describe India's rural electrification scheme, provide a basic overview of fracking, and discuss how guar production in India responded to the fracking boom in the US.

[4]In its use of multiple sources of exogenous variation in real-world settings, for instance, our empirical approach is related to Duque et al. (2018), who examine how early-life exposure to adverse weather shocks (that reduce children's initial skills) in Colombia interacts with the introduction of conditional cash transfers to influence long-term outcomes.

2.1 Rural electrification

Newly independent India had only 1,500 electrified villages in 1947, and progress on rural electrification remained slow well into the late 1960s (Banerjee et al., 2014, p. 35). Severe droughts and food shortages in the early 1960s brought rural electrification into the spotlight, and since that time a number of schemes have emerged.[5] The Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY), launched in 2005, subsumed all existing grid-related rural electrification initiatives. RGGVY was charged with enhancing access to electricity in over 100,000 unelectrified and 300,000 "partially electrified" villages across 27 states. It aimed to do so primarily by installing and upgrading electricity infrastructure to support productive activities in rural economies. These included electric irrigation pumps, education and health-care facilities, and small and medium enterprises. RGGVY also extended free grid connections to rural households below the poverty line; households above the poverty line could purchase connections. Both groups remained responsible for their own power use as RGGVY did not subsidize electricity consumption.

Although largely funded by the federal government, RGGVY was implemented through district-level projects overseen by local implementing agencies, such as the State Electricity Board. The scheme proceeded in two steps. First, to qualify for RGGVY funds, the local implementing agency prepared a Detailed Project Report (DPR) for the district in question. The DPR outlined in detail the electrification-related infrastructure needs of the district, the number of households expected to be connected to the grid, and expected project costs. It also identified the set of villages eligible for electrification under RGGVY. These DPRs were reviewed and approved by India's Rural Electrification Corporation and its Ministry of Power before disbursement of funds. Once approved, district-level implementation commenced in line with the village-by-village plan outlined in the DPR.

Districts were allocated to India's Tenth (2002–2007) and Eleventh (2007–2012) Five-Year Plans for funding based on the order in which DPRs were submitted and approved. We refer to these as "RGGVY Phase I" and "RGGVY Phase II" districts, respectively, and identify these districts using state-level five-year-plan

[5]For instance, the Kutir Jyoti Yojana was launched in the late 1980s to increase access to electric lighting for households below the poverty line; the Pradhan Mantri Gramodaya Yojana, launched in 2001, extended financing to states to enhance access to public services, including electrification, in rural areas; the Remote Village Electrification program, launched in 2002, aimed to provide lighting to remote villages using solar photovoltaics and other off-grid energy technologies; and the country's Minimum Needs Program was updated in 2002 to extend financing for rural electrification to states that were seen to be performing especially poorly (Banerjee et al., 2014, p. 37–38).

progress reports. During Phase I, villages containing at least one habitation (a geographically distinct sub-village cluster of households) with a population of 300 or more as reported in the 2001 Population Census were eligible to be electrified. Approximately 178,000 villages across 235 Phase I districts in 25 states (as per 2011 administrative boundaries) fit this criterion. Funds associated with Phase I districts were disbursed between 2005 and 2008, while funds for Phase II districts—for which the RGGVY eligibility threshold was reduced to 100—were disbursed between 2008 and 2011. Selection of districts into RGGVY Phase I and Phase II is clearly endogenous. For this reason, in this paper, we restrict our analyses to Phase I districts (shown in Figure 1). This approach also accounts for the fact that village-level electrification in Phase I districts had largely been completed in advance of the 2011 Population Census, one of our main data sources, whereas villages electrified as part of RGGVY Phase II would have been captured inconsistently during Census survey enumeration, which began in April 2010.

2.2 Fracking, guar, and guar gum

Hydraulic fracturing ("fracking") involves injecting a slurry of water, sand and chemicals underground at high pressure to create or widen small fractures in the underlying rock formation. While not an entirely new approach, recent technological refinements—in particular, fracking in combination with horizontal drilling—have considerably increased the effectiveness of the process and transformed the global energy landscape (Orr, 2016).

Figure 2 illustrates the production of natural gas (panel *a*) and oil (panel *b*) from fracked and "conventional" wells in the US. In 2000, fracked wells produced less than seven percent of natural gas in the US; that share grew to 67 percent by 2015. Oil production from fracked sources underwent a similarly momentous shift from two percent in 2000 to fifty percent in 2015.

While fracking fluid consists almost entirely of water and sand or similar materials, the remaining chemical ingredients serve many purposes and add substantial value to the frac job (Fetter, 2019; Fetter et al., 2018). One of the most common additives is guar gum, which is used as a gelling agent to increase fluid viscosity. This reduces "leak-off" into the surrounding rock formation (thus reducing the amount of fluid needed for a given job) and carries sand deeper into the wellbore (thus increasing hydrocarbon return flow). Between 25–50 percent of fracking operations rely on guar gum, making it "at least two to three times preferred over synthetic [alternatives]" (Elsner and Hoelzer, 2016).

Guar gum is derived from seeds of the guar plant through a combination of heat and scouring to remove the hard outer shell, followed by milling into "splits" that are ground to powder. Guar is a drought-resistant legume primarily cultivated in the semi-arid northwestern tracts of the Indian subcontinent (Kuravadi et al., 2013). The plant tolerates relatively high temperatures and requires only sparse but regular rainfall, which makes the rain patterns associated with the monsoon in this region ideal for cultivation (Mudgil et al., 2011). Guar—whose name means "cow food"—has traditionally been cultivated as both fodder and a vegetable crop. It grows well in many different types of soil, and its nitrogen-fixing potential combined with its relatively short time to harvest make it an excellent soil-improving crop that fits conveniently into crop-rotation cycles.

India accounts for approximately eighty percent of global production (National Rainfed Area Authority, 2014).[6] The country aso occupies a dominant role in the global trade of guar gum, nearly all of which is processed domestically. Within India, guar is produced almost exclusively in the northwest, chiefly in parts of Rajasthan, as well as in Haryana and Gujarat (Figure 1). Cultivation is relatively decentralized, and the crop is grown by thousands of small and marginal farmers. While precise data on growing practices are unavailable, industry experts believe most guar is rainfed, and farmers typically plant it as a secondary crop on small subsistence plots (Beckwith, 2012).

Although guar gum has long been used in a variety of industries, the dramatic growth of fracking in the US resulted in an unprecedented expansion of guar production in India.[7] Figure 3 shows the mean annual wholesale price of guar in India between 2001 and 2016. Starting in 2009, the price of guar increased by nearly 1,000 percent relative to its level at the start of the decade. Figure 4 highlights trends in India's exports of guar gum over a similar period. Panel *a* of this figure shows that India's exports comprised about one-third of the global trade in guar gum at the start of the decade; near the height of the US fracking boom in 2012, nearly ninety percent of this trade originated in India. Panel *b* shows that the weight of India's guar gum exports more than tripled over the same period.

[6] Neighboring Pakistan is responsible for approximately fifteen percent of global production.

[7] As we show in Appendix A using village-level nighttime luminosity data, the start of the fracking boom led to large increases in economic activity across guar-growing regions.

3 Conceptual framework and empirical strategy

In this section, we highlight three main hypotheses that connect access to electricity with household-level labor supply. We then describe our empirical strategies and comment on the identifying assumptions implicit in each.

3.1 Electrification and labor supply

There are three main pathways through which electrification can affect households' labor-supply decisions. One popular argument relates to the time burden imposed by home production activities, such as collecting and preparing traditional fuels for cooking and heating. If electricity can be used for these purposes instead, it frees up household members' time for engaging in market activities. In practice, exclusive reliance on electricity for cooking is uncommon in low- and middle-income countries, and use of traditional fuels such as firewood is widespread, including among electrified households (e.g., Barron and Torero, 2017; Pattanayak et al., 2019). In such settings, electricity access is unlikely to significantly influence households' time allocation in this way.

A second prominent argument relates to the provision of lighting and its effect on total working hours. If electric lighting can enable households to allocate domestic activities that require good lighting to evening hours, daylight time can be allocated to activities that generate income. Yet households in many rural areas have already transitioned away from low-quality kerosene lighting to relatively high-quality electric lamps powered by small-scale batteries (Bensch et al., 2017). The additional benefits of electric lighting delivered by the grid in such settings are unlikely to be large.

A third channel—and one that is the focus of our paper—relates to the productive potential of income-generating activities that the household can conduct. Specifically, electrification may increase the productivity of activities that do not necessarily require electricity, such as water collection or sewing. It may also enable new opportunities to engage in activities that were previously not possible, such as soldering, metalworking or industrial production. Together, these can influence the effective wage the household faces, which changes the opportunity cost of leisure. Households already engaged in income generation may also reallocate hours to new types of work.

However, if household productivity is determined jointly by access to electricity and complementary community- and household-level factors (such as local weather conditions, access to markets, and households' stock of education and health), any improvement in electrification alone may have very little impact on households' labor allocation.[8] Prior experimental and quasi-experimental evaluations of the impact of electrification have typically ignored these context-specific complementarities. Our setting allows a unique opportunity to shed new light on this question.

3.2 Regression discontinuity design

A before–after comparison of labor-market outcomes in electrified villages located in booming guar-growing districts is unlikely to yield a causal estimate of the impact of electrification in the presence of complementary economic conditions for three reasons. First, this approach lacks a suitable "non-boom" control. Second, it neglects heterogeneity within the set of electrified villages. Larger electrified villages, for instance, are also likely to have better access to schools and health facilities, both of which can directly influence labor-force productivity. Finally, this approach fails to account for changes in other factors over the course of the decade that can act as confounders. A cross-sectional comparison of electrified boom villages with electrified non-boom villages would yield similarly unreliable estimates. Indeed, most guar-growing districts are located in Rajasthan, which, despite the recent boom, remains one of India's poorest states. A simple *ex post* comparison would likely provide an underestimate of the impact of electrification.

Our empirical strategy exploits a population-based threshold that guided the roll-out of India's rural electrification scheme as part of a village-level RD design. Villages in districts approved under RGGVY Phase I were eligible for electrification if they contained a habitation with at least 300 people. Indian villages, however, can contain multiple habitations—typically between one and three—which complicates identification. For instance, a village with a relatively large population may have been ineligible under RGGVY if its population was spread out over multiple habitations; a less populous (but more concentrated) village may have been electrified. A village's overall population can, thus, be a poor measure of its RGGVY eligibility. Without additional information on sub-village habitation characteristics, comparing villages with overall populations above the RGGVY threshold to villages with populations just below it is unlikely to yield an accurate estimate of the impact of electrification.

[8]We illustrate this using a simple model of electrification and household time allocation in Appendix B.

To address this concern, we restrict our nationwide sample of villages to single-habitation villages, following the empirical approach developed by Burlig and Preonas (2016). This allows us to estimate the local average treatment effect (LATE) of electrification on labor-market outcomes for villages with overall populations close to RGGVY's 300-person eligibility threshold. Specifically, we focus only on single-habitation villages in RGGVY Phase I districts with a population within a suitable bandwidth of 300. This allows us to account for endogeneity of district selection into the two phases of RGGVY and ensure that electrified villages (with populations just above the threshold) are comparable to unelectrified ones (with populations just below it). To measure the importance of complementary economic factors within this sample, we compare the impacts of rural electrification in villages located in boom districts to the impacts of electrification in villages located in non-boom districts.

More formally, we rely on an RD design to estimate

$$\begin{aligned} y_{vd}^{post} = & \beta_1 T_v + \beta_2 T_v G_d \\ & + \beta_3 \tilde{P}_v^{2001} + \beta_4 T_v \tilde{P}_v^{2001} + \beta_5 G_d \tilde{P}_v^{2001} + \beta_6 T_v G_d \tilde{P}_v^{2001} \\ & + \beta_7 y_{vd}^{pre} + \gamma_d + \varepsilon_{vd} \end{aligned} \tag{1}$$

for $-b \leqslant \tilde{P}_v^{2001} \leqslant b$. y_{vd}^{post} represents a post-electrification outcome of interest for village v located in district d; $\tilde{P}_v^{2001} \equiv P_v^{2001} - 300$ (where P_v^{2001} is its population in the 2001 Census round); and b denotes a suitable population bandwidth around RGGVY's 300-person eligibility threshold. Our preferred specification uses a bandwidth of fifty people on either side of this cutoff. T_v is a binary variable that equals one if $P_v^{2001} > 300$, i.e., the population of village in v in 2001 is above RGGVY's eligibility threshold. G_d is a binary variable that equals one if village v is located in a guar-growing boom district. y_{vd}^{pre} is the pre-electrification value of the outcome variable. γ_d represents a district fixed-effect, which allows us to control for all time-invariant district-specific characteristics that can independently induce variation in the level of the outcome of interest. ε_{vd} is a village-specific error term. We cluster standard errors by district to allow for correlated unobservables between villages that are located nearby and, in line with RGGVY's implementation structure, served by the same district-level electrification agency.

In Equation (1), β_1 represents the LATE of electrification in villages in non-boom regions. Our parameter of interest is β_2, which represents the additional effect of electrification in villages affected by the guar boom. If $\hat{\beta}_2$ is statistically different from zero, we conclude that the LATE for electrification in boom regions is different from that in the rest of India. Conditional on the inclusion of district fixed-effects, this highlights the degree to which complementary economic conditions—in this case, generated by the exogenous guar boom—augmented the impact of electrification.

Identification relies on continuity of potential outcomes in village population (our running variable) at the RGGVY eligibility threshold. This assumption is plausible if (i) villages are not able to manipulate their population levels—either in actuality or in administrative reporting—to influence RGGVY eligibility; and (ii) all observable and unobservable village-level covariates that may be correlated with our outcomes of interest change smoothly at the threshold. The former is unlikely to be a concern as RGGVY used figures from the 2001 round of the Population Census—which predated the announcement of RGGVY by at least four years—to gauge eligibility (Burlig and Preonas, 2016). Nevertheless, we use the RD manipulation testing procedure developed by Cattaneo et al. (2018, 2019) to empirically check for bunching at the cutoff and find no evidence to suggest that this is the case (Figure 5).

The latter component of this assumption, that all village-level covariates change smoothly at the threshold, is fundamentally untestable. That said, we provide evidence in support of it by examining the pre-RGGVY distribution of key village-level characteristics around the cutoff. We find no evidence of discontinuous changes at the 300-person mark prior to RGGVY implementation (Table E2). We are also aware of no other social program in India that uses RGGVY's 300-person habitation-level eligibility criterion.[9]

3.3 "Quadruple-differences" estimator

To highlight potential mechanisms, we use repeated cross-sectional data covering all non-farm establishments across India. Specifically, we exploit variation between establishments in (i) boom and non-boom districts; (ii) districts selected and bypassed for electrification under RGGVY Phase I; (iii) boom and non-boom industries (as indicated by an industrial classification code); and (iv) the pre- and post-electrification periods to estimate a quadruple difference-in-differences ("quadruple-differences") specification.

[9]To the best of our knowledge, the only other program that uses habitation-level population data to decide eligibility is the Pradhan Mantri Gram Sadak Yojana (PMGSY), a rural roads program that connected villages containing a habitation with at least 500 people to India's road network. Given our fifty-person bandwidth around RGGVY's 300-person threshold, all villages in our sample would have been ineligible for PMGSY.

Consider the following regression:

$$
\begin{aligned}
y_{ijdt} = &\beta_1 \left(Boom_d \times Ind_j\right) + \beta_2 \left(RGGVY_d \times Ind_j\right) \\
&+ \beta_3 \left(Boom_d \times RGGVY_d \times Ind_j\right) + \beta_4 \left(Ind_j \times Post_t\right) \\
&+ \beta_5 \left(Boom_d \times Ind_j \times Post_t\right) + \beta_6 \left(RGGVY_d \times Ind_j \times Post_t\right) \\
&+ \beta_7 \left(Boom_d \times RGGVY_d \times Ind_j \times Post_t\right) + \gamma_j + \gamma_{dt} + \varepsilon_{ijdt},
\end{aligned}
\tag{2}
$$

where y_{ijdt} represents an outcome of interest for establishment i in industry j in district d in year t. $Boom_d$ is a binary variable that equals one if district d is a guar-growing boom district; $RGGVY_d$ is a binary variable that equals one if district d was selected for electrification under RGGVY Phase I; Ind_j is a binary variable that equals one if establishment i operates in an industry related to the boom; and $Post_t$ is a binary variable that equals one if year t is in the post-electrification period, and zero otherwise. γ_j is an industry fixed-effect that controls for time-invariant differences between establishments in different industries, γ_{dt} is a district–year fixed-effect that controls for district-specific time trends, and ε_{ijdt} is an establishment–year-specific error term. We cluster standard errors at the industry level to account for intra-industry correlation in establishment-level characteristics nationally.

Our parameter of interest is β_7, the quadruple-differences estimand that indicates the impact on boom-industry establishments located in boom districts selected for electrification under RGGVY Phase I. One might be concerned that changes in these industries or districts occur at the expense of establishments in other areas. To evaluate this possibility, we formally compare estimates for our key parameter with other estimated coefficients, which highlight changes in other industry–district groups over time.

It is worth noting that the quadruple-differences specification in Equation (2) entails considerably weaker identifying assumptions than the conventional difference-in-differences design. Identification would be threatened only by a district–industry-specific time effect (i.e., a shock that alters establishment-level outcomes over time only for boom-industry establishments located in boom districts that were selected for RGGVY Phase I). While this is possible, we contend that it is unlikely in practice. We are aware of no change in industrial policy over this period, for instance, that specifically targeted establishments on the basis of their respective industrial classification codes, which we use to identify and delineate boom industries, much less one that did so for establishments only in guar-growing boom districts included in RGGVY Phase I.

4 Data

We rely on four main sources of data. First, we use technical reports published by the governments of India and the US to identify India's main guar-growing districts. We complement these with information on the roll-out of rural electrification in India to identify districts approved for electrification under RGGVY Phase I. Next, we obtain data on the composition of the village-level labor force from multiple rounds of India's Population Census. Finally, we use multiple rounds of the Economic Census to obtain establishment-level data on size, sectoral composition and total employment.

4.1 Guar production

We review three separate technical reports on guar production to identify guar-producing districts that were affected by the start of the US fracking boom. Two of these—prepared by the Agricultural and Processed Food Products Export Development Authority (2011) and the National Rainfed Area Authority (2014)—represent efforts by the Indian government to systematically quantify and summarize the nationwide production and trade of guar.[10] The third—prepared by the US Department of Agriculture—signals the growing interest the agency took in guar production as the crop grew to become India's main agricultural export to the US (Singh, 2014).

For each of these technical reports, we create a list of districts that are characterized as key producers of guar on the basis of overall production and area under cultivation (Table E1). We designate a district as a guar-growing district for the purposes of our analyses if it appears on at least two of these lists. Based on district boundaries at the time of the 2011 Population Census, we identify 23 districts in this way: thirteen in Rajasthan, six in Gujarat, and four in Haryana (Figure 1). We refer to these districts as India's "boom districts" or "boom regions." In 2011, these 23 districts were home to nearly sixty million people living over an estimated area of 300,000 km^2—roughly equal to Italy in terms of population and size.

To validate our selection of these districts, we also estimate their share in total reported guar production and area under cultivation using national data from the

[10]The Agricultural and Processed Food Products Export Development Authority (APEDA), housed within the Ministry of Commerce and Industry, supports the development of industries related to products with export potential. The National Rainfed Area Authority (NRAA) is housed within the Ministry of Agriculture and Farmers' Welfare, and provides technical advice and monitoring for government schemes in rural areas with significant levels of rainfed agriculture.

Ministry of Agriculture on annual district-wise production of the crop.[11] While these district-level production statistics provide a useful check, their quality is insufficient to serve as a principal data source. For instance, districts in the state of Haryana—consistently referred to in technical reports as one of the most important guar-producing states in India after Rajasthan—have non-missing data on guar production only for 2012. At the same time, other districts in regions of India not known for guar production consistently report trivial amounts of production for multiple years in the sample. Nevertheless, we find that the guar-growing districts we identify accounted for 84 percent of guar production and 91 percent of area under guar cultivation in 2005 (the year RGGVY was launched).

4.2 Rural electrification

As mentioned previously, we identify Phase I districts for which DPRs were successfully submitted and approved using state-level five-year-plan progress reports for RGGVY.[12] To identify villages within these districts that met RGGVY's habitation-level eligibility threshold, we obtain habitation-level population data from the census of habitations conducted by the National Rural Drinking Water Program (NRDWP) in 2009.[13] Because the NRDWP data indicate only the name—and not the unique Census code—for each habitation's corresponding village, we use a name-based matching algorithm to match it with a list of Census-designated villages. We successfully match 94 percent of the approximately 560,000 villages listed in the 2001 Population Census to their constituent habitations using a combination of exact (75 percent of matched names) and fuzzy matches that incorporate a fuzzy-matching algorithm originally developed by Asher and Novosad (2020).[14]

To further validate the quality of these matches, we calculate the discrepancy between the Census 2011 population for each village and the NRDWP 2009 population estimate that we obtain from summing the population over all habitations in a village. We drop all villages with a Census–NRDWP population discrepancy of

[11] These data are available via the Ministry of Agriculture's Crop Production Statistics Information System at https://aps.dac.gov.in/APY/Index.htm.

[12] For each state, these reports—entitled "Report C-Physical & Financial Progress of RGGVY Projects Under Implementation (Plan-wise)"—list the district name and DPR code, the name of the district-level local implementing agency, and details about the financial scope of the project and progress towards meeting village- and household-level electrification targets. They are available via the website of the Deendayal Updhayaya Gram Jyoti Yojana (DDUGJY)—into which RGGVY was ultimately subsumed—at http://www.ddugjy.gov.in/.

[13] The NRDWP census of habitations was first conducted in 2003 and again in 2009. The 2003 data are no longer publicly available; the 2009 data are available at https://ejalshakti.gov.in/.

[14] We describe our habitation–village matching procedure in detail in Appendix C.

greater than twenty percent; these, we assume, are incorrect matches. This leaves us with approximately 370,000 villages.

Our name-matched dataset consists of village-level identifiers (i.e., state, district, subdistrict and village names, and their corresponding Census codes), village-level count of habitations, village population, and population of the largest habitation. The average village in this sample contains three habitations; about 47 percent of villages contain exactly one habitation.

To obtain the analytical sample with which to estimate Equation (1), we impose three key restrictions. First, we restrict our sample to villages in only RGGVY Phase I districts. Recall that village-level electrification in these districts had largely been completed in advance of the 2011 Population Census, which serves as one of our main data sources. Because district selection into the two phases of RGGVY is endogenous, villages in Phase II districts would not serve as suitable controls for those in Phase I districts. Second, we omit all villages with more than one habitation, which allows us to more precisely gauge village-level electrification under RGGVY's 300-person population-based eligibility criterion. Finally, we look only at villages with a Census 2001 population within a fifty-person bandwidth of the RGGVY Phase I threshold, which ensures that electrified villages (with populations just above the threshold) are comparable to unelectrified ones (with populations just below it). This yields 7,655 villages located across 22 Indian states; 148 of these villages are located in boom districts.

4.3 Rural labor-market outcomes

Our data on the make-up of the rural labor force come from two sources. First, we use the 2001 and 2011 rounds of the Population Census. Specifically, the Primary Census Abstract (PCA) data tables in the Census report information by gender on three distinct village-level subgroups: (i) "main workers," who engage in any economically productive activity for at least six months per year; (ii) "marginal workers," who do so for less than six months per year; and (iii) "non-workers," who do not engage in any economically productive activity. Within the first two subgroups, workers are further categorized as cultivators, agricultural laborers, household-industry workers, or "other." A person is classified as a cultivator if they are engaged in cultivation of land that they own or lease, implying that they bear the risks associated with cultivation. In contrast, a person is classified as an agricultural laborer if they work on another person's land for payment. In rural areas, a household industry is defined as “production, processing, servicing, repairing, or making and selling (but not merely selling) of goods” that is done by one or more members of a household within the confines of the village. Finally, “other” workers include professions such as government employees, teachers and traders.[15]

For each village–year in our Census panel, we combine cultivators and agricultural laborers (both main and marginal) to calculate the population of agricultural workers, overall and by gender. We similarly combine household-industry and other workers to obtain corresponding figures for the village-level population of non-agricultural workers. These data—together with information on village population as well as the breakdown of that population into workers and non-workers—allow us to evaluate impacts along two dimensions: (i) the extensive margin, i.e., the net change in the overall labor force as a percentage of the village population; and (ii) the sectoral composition of the labor force, i.e., the relative shares of agricultural and non-agricultural workers.

We complement these data with figures on total village-level employment derived by Asher et al. (2019) from multiple rounds of the Economic Census (EC) of India.[16] The EC is an enumeration of all non-farm establishments (including informal firms, service-sector firms, and publicly-owned firms) conducted in 1990, 1998, 2005 and 2013. While not perfectly analogous to the employment statistics that we derive from the Population Census, these figures allow us to evaluate labor-market impacts and trends over a longer time period.

4.4 Establishment-level data

We turn to the EC again for establishment-level data. Specifically, we use the 2005 (fifth) and 2013–14 (sixth) rounds of the EC to obtain information on total number of employees and industrial classification—as indicated by a National Industrial Classification (NIC) code—for all establishments in India.[17] Together, the two EC rounds list approximately 100 million establishments nationwide. We first use geographic concordance tables developed by Asher et al. (2019) to match around 96 percent of these establishments to 2011 district boundaries. As the 2005 EC used a prior version of the NIC system, we then use concordance tables developed by India’s Central Statistical Organization (2008, p. 129) to link the industrial classification

[15] Additional information about these definitions is available at http://www.censusindia.gov.in/2011census/HLO/Metadata_Census_2011.pdf.

[16] These are available at http://devdatalab.org.

[17] These data are available from the Ministry of Statistics and Program Implementation at http://www.mospi.gov.in/economic-census-3.

systems used in the two EC rounds. These steps yield a repeated cross-section of over 96 million establishments (operating in 196 different industries) with which to study changes in the nature and composition of establishments in response to the roll-out of rural electrification and the start of the guar boom. The mean (median) establishment in this sample had 2.7 (2) employees.

To delineate industries that are likely to have been directly affected by the guar boom, we use the 2013 EC "Directory of Establishments" for Rajasthan, which lists the names of all establishments in the state with ten or more employees. We first identify the NIC codes corresponding to establishments that are easily recognized as guar-processing units because their names contain variants of the word "guar." We then conduct a detailed review of the breakdown of NIC codes prepared by the Central Statistical Organization (2008) to identify additional codes that can contain guar-processing units. Ultimately, we identify five three-digit NIC codes that can contain industrial units most directly related to guar processing.[18] Together, these represent approximately four percent of all establishments—and six percent of establishments with ten or more employees—in India.

5 Short-term impacts of electrification

In this section, we estimate how rural electrification affected the size and composition of the labor force across boom and non-boom regions of India in the short run. We measure these effects using data on population and employment at the village level from the Population Census. We find no evidence to suggest that electrification had a net effect on total employment in villages located within India's booming guar-growing regions. At the same time, electrification led to a large shift from agricultural to non-agricultural employment in these villages. In contrast, we find no evidence that electrification had any discernible short-run impact on labor-market outcomes in villages located in the rest of the country.

[18] As per the 2008 NIC system, these are: (i) Support activities to agriculture and post-harvest crop activities (016); (ii) Manufacture of basic chemicals, fertilizer and nitrogen compounds, plastics and synthetic rubber in primary forms (201); (iii) Manufacture of prepared animal feeds (108); (iv) Manufacture of non-metallic mineral products (239); and (v) Wholesale of agricultural raw materials and live animals (462).

5.1 Total employment

We begin by studying the short-run impacts of electrification on the total number of workers (agricultural and non-agricultural workers together) as a share of village population. Panel (*a*) of Figure 6 plots the predicted share of total workers from the 2011 round of the Population Census just above and below the RGGVY threshold separately for villages located in boom and non-boom regions. This figure graphically depicts the results from the RD specification outlined in Equation (1). It suggests that electrification had no discernible short-run effect on the size of the labor force in villages in either boom or non-boom regions.

The regression results presented in Table 1 support these findings and attach a magnitude to the effects. The estimate in the first row of this table represents the effect of electrification in non-boom districts; as the indicator variable suggests, these villages are located just above RGGVY's eligibility threshold. The estimate in the second row, the interaction of the preceding parameter with the indicator variable for being located in a boom district, represents the degree to which the impact of electrification is augmented by complementary economic factors in India's booming guar regions. Column (1) reports the main RD estimates for these two parameters for the overall working population in 2011. The magnitude of the estimates is small. In non-boom villages, for instance, the results point to a reduction in the overall size of the workforce by 0.8 percentage points (s.e. 0.6), an imprecisely estimated decrease of less than two percent. The estimated coefficient for the additional effect in electrified boom villages in the second row is similarly small. Importantly, neither of these results is statistically significant at conventional levels, and we are unable to reject the hypothesis that access to electricity had no short-run effect on total employment. Columns (2) and (3) of Table 1 report the same specification estimated separately for the share of male and female workers in 2011, respectively. The estimates are similar: electrification had no discernible short-run effect on male or female employment in both boom and non-boom villages.

Taken together, these results suggest that, on net, households did not respond to electrification by adjusting their labor choices along the extensive margin in the short run.[19] Although we cannot rule out that large-scale entry and exit of workers

[19] We also examine the extent to which electrified villages experience large-scale in-migration. We do this by testing for discontinuous changes in the 2011 population of these villages. We find that at the RGGVY threshold, electrified villages exhibit a discontinuous increase in population, driven entirely by an increase in the male population (Table E3). However, the magnitude of this change is small (an average increase of approximately three people, or less than one percent relative to the sample mean). Thus, we rule out that electrified villages are on the receiving end of large-scale

in response to electrification may have taken place, these findings stand in contrast to those from earlier work (e.g., Dinkelman, 2011) that suggests that access to electricity can increase net labor-force participation (especially for women).

5.2 Sectoral composition of employment

We next study the short-run impacts of electrification on the sectoral composition of the rural labor force (agricultural and non-agricultural workers separately). Specifically, we use data on the village-level population of agricultural and non-agricultural workers from the 2011 Population Census to study how the relative sizes of these two sectors changed in response to rural electrification.[20]

We find that electrification substantially reduced the share of the agricultural labor force and increased the share of the non-agricultural labor force in boom villages in the short run. Table 2 provides numerical results from estimating Equation (1) separately for these subgroups. Electrification reduced the share of agricultural workers in the population of non-boom villages by 1.2 percentage points (s.e. 0.6) relative to a sample mean of approximately 36 percent (column 1). Boom villages, in contrast, experienced an additional reduction in this share of over six percentage points (s.e. 1.7). In other words, complementary economic factors led to an approximately fivefold augmentation in the impact of electrification on the share of the agricultural labor force. Comparing the estimates in the third row for male (column 2) and female (column 3) agricultural workers suggests that the magnitude of this effect was especially large for women. The economic boom augmented the reduction in the share of male agricultural workers due to electrification by 2.9 percentage points (thirteen percent) and that of female agricultural workers by 3.3 percentage points (24 percent).

Columns (4)–(6) of Table 2 report corresponding estimates for the non-agricultural labor force. The first row shows that electrification appears to have had no discernible short-run impact on the share of the non-agricultural workforce in villages in non-boom districts. In contrast, column (4) shows that electrification in boom villages increased the size of the non-agricultural labor force by an additional 5.5 percentage points (s.e. 1.2), representing a seventy percent increase relative to the sample mean. This increase is nearly identical to the reduction in the share of agricultural workers in column (1). The second row of columns (5) and (6) shows that this effect is, once again, driven especially by the female workforce. As shown in column

in-migration due to electrification.

[20] We describe how we derive these figures from the Population Census in Section 4.3.

(6), electrification in boom villages increased the share of female non-agricultural workers by over three percentage points (s.e. 1.2), more than doubling the share relative to the sample mean. For male agricultural workers, the 2.3 percentage point (s.e. 1.1) additional increase represents an increase of just under 45 percent.

More broadly, comparing the results for electrified boom and non-boom villages in Table 2 shows that while the estimated coefficients for the impact of electrification in non-boom villages generally have the same signs as those for boom villages, the former are considerably smaller in magnitude and largely indistinguishable from zero. In contrast, electrification appears to have resulted in a large structural shift away from agricultural work and into non-agricultural employment in boom villages, where complementary economic conditions were present. Panels (*b*) and (*c*) of Figure 6 graphically represent the results from our RD specification for agricultural and non-agricultural workers, respectively, and visually highlight the large differences in impact across the two settings for these subgroups.[21]

5.3 Complementary economic conditions or regional effects?

Boom districts, which are home to guar production, could differ from the rest of India along a variety of metrics that induce variation in labor-market outcomes independent of rural electrification (e.g., literacy rates). The inclusion of district fixed-effects in the RD specification outlined in Equation (1) controls for all such unobserved spatial differences. We also find that villages in boom and non-boom districts were statistically indistinguishable in 2001—before the guar boom or rural electrification—along a host of key socioeconomic indicators (Table E4).

However, these district-specific characteristics might also have interacted with rural electrification in regionally distinct ways. In particular, recall that all of the booming guar-growing districts are located in either Rajasthan, Haryana or Gujarat (as shown in Figure 1). Among other things, these three neighboring states share a distinct hot, semi-arid climate, which could have differentially affected the roll-out of RGGVY Phase I in the region and driven the results reported in Table 2.[22] Additionally, institutional capacity in these states, which can also influence how effectively RGGVY was implemented, may have been regionally distinct as well. If this is the case, our analyses might simply reflect a "northwestern-India effect."

[21] Figure D1 shows results from regressions employing an alternative, non-linear functional form.

[22] Specifically, Rajasthan, Gujarat and Haryana lie nearly entirely within the hot desert (BWh) and hot semi-arid (BSh) climatic zones, as per the Köppen–Geiger climate classification system (Beck et al., 2018).

To test this, we turn to a randomization-based inference procedure (Athey and Imbens, 2017). We randomly select "placebo" boom districts from all (boom and non-boom) RGGVY Phase I districts in Rajasthan, Haryana and Gujarat, then re-estimate Equation (1) and collect the placebo estimate for $\hat{\beta}_2$, our coefficient of interest. We repeat this process 1,000 times for each of the main outcomes reported in Table 2 to obtain a distribution of placebo estimates for $\hat{\beta}_2$ for each dependent variable.

Figure 7 shows these distributions and highlights their 90 and 95 percent confidence intervals. If the differential effect of electrification on the share of agricultural and non-agricultural workers that we observe in boom districts was due to a regional effect common to all RGGVY Phase I districts in Rajasthan, Haryana and Gujarat, we would expect to find our actual estimated values for this parameter from Table 2—indicated by the dashed vertical lines in Figure 7—near the middle of these distributions. Instead, the actual estimates of $\hat{\beta}_2$ are extreme values outside the 95 percent confidence intervals of these distributions in all cases. That is, any other configuration of RGGVY Phase I districts in these three northwestern states is highly unlikely to yield estimates that are as large in magnitude. This strongly suggests that it is indeed the advent of the guar boom and its interaction with the roll-out of rural electrification as part of RGGVY Phase I that drives the results we observe.[23]

5.4 Robustness checks

We check the robustness of our main results in four ways. Note that in constructing the sample of single-habitation villages for our main RD analyses, we made two key choices: (i) during our village–habitation name-matching procedure, we discarded any village with a discrepancy of greater than twenty percent between its total Census 2011 population and its total NRDWP 2009 population (calculated by combining the population in each of its matched habitations); and (ii) we restricted our sample to villages within a fifty-person bandwidth of RGGVY's 300-person eligibility threshold. We thus start by testing the sensitivity of our main results to each of these choices.

[23] If there was considerable national-level heterogeneity in the impacts of rural electrification, it is also possible that any random subset of RGGVY Phase I districts from across India could potentially exhibit the differential impacts that we identify in Table 2. In other words, it could be the case that we observe the results that we do simply by chance. To test this, we run a similar randomization inference test in which we draw placebo boom districts from all RGGVY Phase I districts across the entire country. Figure D2, shows that our actual estimates of $\hat{\beta}_2$ are extreme values outside the 90 or 95 percent confidence intervals of these distributions in all cases, reinforcing our conclusion that the results in Table 2 are indeed driven by the interaction of the economic boom and rural electrification.

First, we estimate Equation (1) allowing for increasingly greater levels of population discrepancy in our sample but keeping fixed our preferred fifty-person RD bandwidth. Figure D3 shows how $\hat{\beta}_2$, our coefficient of interest, evolves as we relax our definition of what we consider a successful match, thereby increasing the size of the underlying analytical sample. As the sample expands to contain an increasing number of villages that are unlikely to have been good matches, the magnitude of $\hat{\beta}_2$ generally attenuates gradually as expected. In particular, we do not observe erratic changes in the magnitude of this coefficient.

We then fix the sample population discrepancy rate at our preferred level of twenty percent and vary the size of the RD bandwidth around RGVVY's 300-person eligibility threshold. Figure D4 shows how $\hat{\beta}_2$ evolves as the RD bandwidth widens. Once again, as the analytical sample expands to contain an increasingly dissimilar number of villages on either side of the RGGVY eligibility threshold, the magnitude of $\hat{\beta}_2$ attenuates smoothly.

Next, we use RGGVY Phase II districts in India to conduct a placebo test. Specifically, using only those districts of India that were not approved for rural electrification as part of RGGVY Phase I, we estimate Equation (1) for the overall share of all workers, agricultural workers and non-agricultural workers in the village population.[24] As large-scale roll out of rural electrification did not occur in these districts over the period covered by our data, we should not expect to see an impact of a village's 2001 population being above RGGVY's eligibility threshold in either boom or non-boom districts. Table E5 confirms this intuition.

Finally, we adjust our inference to account for multiple hypothesis testing using the free step-down resampling methodology of Westfall and Young (1993). This bootstrap-based procedure controls the family-wise error rate (the probability of a type I error when testing a "family" of hypotheses).[25] We combine all regressions reported in Tables 1 and 2 into a family of hypotheses and use this approach to control the family-wise error rate associated with $\hat{\beta}_2$. Table E6 reports that our main result—that electrified villages in guar-growing districts see a large reduction in the share of agricultural workers and a corresponding increase in the share of non-agricultural workers relative to electrified villages in non-guar districts—is robust to this adjustment.

[24] As shown in Figure 1, twelve non-RGGVY Phase I districts are also guar-growing districts.

[25] See Jones et al. (2019) for a detailed description of how this is implemented.

6 Longer-term impacts of electrification

We turn next to data on total village-level employment derived by Asher et al. (2019) from the 1990, 1998, 2005 and 2013 rounds of the Economic Census (EC) of India. The EC is an enumeration of all non-farm establishments (including informal firms, service-sector firms, and publicly-owned firms) in the country. While not perfectly analogous to the total employment figures that we derive from the Population Census, these data allow us to evaluate the labor-market impacts of electrification beyond the time period covered by the 2011 Population Census.

Using a compound annual growth rate approach applied to data from the 2001 and 2011 rounds of the Population Census, we first impute village-level population in 1990, 1998, 2005 and 2013 for each village in our sample. We then calculate the share of workers in each village's population in each year, which we use to estimate a modified version of the RD specification outlined in Equation (1) separately for each EC round.[26]

Table 3 presents our results. As shown in the first row of column (1), we find that the impact of electrification on employment in non-boom villages in 2013 (approximately eight years after the launch of RGGVY) is statistically indistinguishable from zero, consistent with the results presented in Table 1. In contrast, the second row of this column shows that electrification had increased the share of workers in villages in boom districts by an additional nine percentage points (s.e. 4.3) during the same period, over 1.5 times the sample mean.

We next estimate the same specification separately using data from the 1990, 1998 and 2005 rounds of the EC as part of a series of falsification tests. As these EC rounds cover the pre-boom/pre-electrification period, we should not expect to observe discontinuous changes in the share of workers in either boom or non-boom villages located above and below RGGVY's 300-person eligibility threshold in these years. The results presented in columns (2)–(4) of Table 3—covering the 1990, 1998 and 2005 EC rounds, respectively—confirm this intuition.

[26] Village-level matches between the Population and Economic Censuses are imperfect (Asher et al., 2019). We are able to obtain information on employment for seventy to ninety percent of the 7,655 villages in our main sample, depending on the EC round. For this reason, we omit controls for the pre-period level of the outcome variable in these regressions since missing observations across one or more EC rounds considerably reduce the size of our analytical sample.

7 Mechanism: Growth of complementary firms

Spatial heterogeneity in the availability of infrastructure (such as electricity from the grid) can give rise to differences in comparative advantage and guide firm-level investment and location decisions (e.g., Martin and Rogers, 1995). Variation in the intensity of an unforeseen demand shock (such as the one generated by the guar boom) can have similar effects (Adhvaryu et al., 2013). These differences can interact and drive firm-level decisions along both the extensive and intensive margins: firm entry/exit and firm shrinkage/growth, respectively. In this section, we focus on how the number and size of non-farm establishments in India evolved across boom and non-boom districts, across districts selected and excluded for electrification as part of RGGVY Phase I, and across boom and non-boom industries within these districts. We demonstrate that the agricultural boom led to increased entry of small establishments—and growth in the size of large establishments—in the guar-processing industry in electrified boom districts. This expansion in the industrial (electricity-intensive) parts of the guar production chains helps uncover potential mechanisms for the labor-market effects reported previously.

7.1 Proliferation of establishments

We look first at effects along the extensive margin by investigating differential trends in the proliferation of establishments. Using data from two consecutive rounds of the EC, we construct an industry–district-level panel by calculating the total number of establishments in each industry in each district as a percentage of all establishments in each district–year. We use this to estimate the specification outlined in Equation (2) at the industry–district level. Column (1) of Table 4 reports our results. The coefficient for the quadruple-interaction term in the last row points to a 0.2 percentage-point (s.e. 0.13) increase in the share of boom-industry establishments in electrified boom districts, a statistically significant increase of around 45 percent relative to the share of establishments in the average industry–district pair. We find no evidence of such an increase in the share of establishments in any other type of industry or in any other type of district, as shown by the coefficients reported in the fourth, fifth and sixth rows of column (1).

Separately estimating Equation (2) for the share of small (fewer than ten employees) and large (ten or more employees) establishments in each industry–district reveals that this effect was entirely driven by high rates of entry of the former.[27] Specifically, as shown in the last row of column (2), we find that the share of small boom-industry establishments increased by approximately 0.2 percentage points (s.e. 0.13) in electrified boom districts. In contrast, we find no evidence to suggest a similar change in the share of large boom-industry establishments in these districts, as shown in the last row of column (3). Once again, in both cases, we find no evidence of an impact—of either rural electrification or of the boom—on establishments elsewhere.

7.2 Size of establishments

We turn next to the relative sizes of establishments to shed light on establishment-level responses along the intensive margin. Using the total number of employees listed for each establishment in the EC as the outcome variable, we estimate the quadruple-differences specification outlined in Equation (2) on our establishment-level repeated cross-section. Columns (4)–(6) of Table 4 report our findings. Column (4) shows results from estimating Equation (2) on the full sample of establishments. We find no evidence to suggest that the size of the average establishment—either within or outside of the boom industry—changed between 2005 and 2013.

However, columns (5) and (6), which report results from separately estimating Equation (2) on the sample of small and large establishments, respectively, reveal considerable underlying heterogeneity. As shown in the the last row of column (5), we find that the size of the average small establishment in electrified boom districts fell about approximately 0.3 employees (s.e. 0.15) between 2005 and 2013. This is consistent with the high rates of entry of small establishments into this industry in these districts shown in column (2), and further suggests that this prior effect is driven by the entry of very small establishments. In contrast, the last row of column (6) shows that large boom-industry establishments in electrified boom districts grew on average by 19 employees (s.e. 6.5). This represents a 61 percent increase in the size of such establishments relative to the sample mean.

Examining the other coefficients in column (6) of Table 4 provides additional suggestive evidence of the presence of complementarities between rural electrification and the boom. As shown in the fifth row of column (6), for instance, we find that the size of the average large boom-industry establishment in an unelectrified

[27] Our definition of "small" and "large" establishments follows the convention established by the 2013 EC "Directory of Establishments," which is available at http://www.mospi.gov.in/sites/default/files/6ec_dirEst/ec6_contant_page.html.

boom district fell by approximately 9 employees (s.e. 4.7). At the same time, rural electrification alone appears to have had no discernible impact on the size of large boom-industry establishments located in districts unaffected by the boom, as shown in the sixth row of this column.

Taken together, our results point to establishment-level responses to the exogenous shock along both the extensive and intensive margins. Increased demand for guar gum spurred a shift in the labor force toward industrial-scale guar processing, which benefits from upgrades to local electricity infrastructure. Consistent with this, we find a large increase in both the number and size of non-farm establishments related to the industrial (electricity-intensive) parts of the guar production chain, such as guar processing, in electrified boom regions. Specifically, we uncover evidence of entry of small establishments into boom industries. At the same time, larger industrial units responded by increasing the scale of their operation in regions where economic opportunity complemented the wider availability of electricity. These trends help shed more light on the heterogeneous impacts of electrification on employment that we present in Sections 5 and 6 by demonstrating how interactions between infrastructure and economic contexts also guide firm- and establishment-level decisions.

8 Conclusion

In this paper, we combine two natural experiments—an exogenous fracking-induced commodity boom in northwestern India, and population-based discontinuities in the contemporaneous roll-out of a massive rural electrification scheme—within a regression discontinuity design to evaluate how the causal effect of rural electrification on labor-market outcomes changes with exogenous variation in complementary economic conditions. We assemble a variety of evidence from multiple large administrative datasets to reach three main conclusions. Our first finding is that, in villages within India's boom-affected regions, access to electricity led to a large short-run increase in non-agricultural employment relative to agricultural employment. We also show that this structural shift translated into an increase in total employment over the longer term. Second, we find that these labor-market dynamics appear to have been driven by an increase in employment by electricity-intensive industrial units that complement guar production (such as guar-processing establishments) near these communities. Third, we demonstrate that, on average, access to electricity appears to have had no discernible impact on labor-market outcomes in villages located in the rest of India.

The main implication of these findings is that complementary economic conditions and contexts are crucial for the ultimate impacts of large-scale electrification. Proponents have long claimed that reliable electricity delivered by the grid is fundamental for the structural transformation of rural economies. Its potential to drive job creation and employment growth is often central to this argument, yet the evidence base on this point remains thin. In particular, impact evaluations are typically unable to rigorously shed light on drivers of spatial and temporal heterogeneity. We show that access to electricity from the grid led to large-scale structural transformation of the rural economy in large swathes of northwestern India, which saw the simultaneous rise of exogenous but complementary economic opportunities. In the rest of India, where these complementary conditions were generally lacking, grid-scale electrification had largely negligible impacts on rural labor-market outcomes.

These results highlight the role electrification—and large-scale infrastructure, more broadly—can play in low- and middle-income countries. Alone, such investments may be insufficient, yet built in anticipation of (and to support) other policies and changes, large-scale infrastructure can provide a foundation for sustained economic growth and development. In our setting, access to grid-scale electricity allowed individuals, households, and firms to respond to rapidly changing economic contexts in ways that potentially deliver economic benefits and improve welfare. We believe that rigorously identifying other potential drivers of the success of large-scale infrastructure is a promising avenue for future research.

References

Adhvaryu, Achyuta, A. V. Chari, and Siddharth Sharma, "Firing Costs and Flexibility: Evidence from Firms' Employment Responses to Shocks in India," *Review of Economics and Statistics*, July 2013, *95* (3), 725–740.

Agricultural and Processed Food Products Export Development Authority, "APEDA Agri Exchange Ready Reckoner Series: Guargum," Technical Report, Ministry of Commerce & Industry 2011. Retrieved from http://apeda.gov.in/apedawebsite/six_head_product/Guargum_final_Profile.pdf on April 16, 2017.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, March 2020, *110* (3), 797–823.

_ , Tobias Lunt, Ryu Matsuura, and Paul Novosad, "The Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG)," 2019. Working paper.

Athey, S. and G.W. Imbens, "The Econometrics of Randomized Experiments," in "Handbook of Field Experiments," Elsevier, 2017, pp. 73–140.

Banerjee, Abhijit V and Esther Duflo, "The Economic Lives of the Poor," *Journal of Economic Perspectives*, January 2007, *21* (1), 141–167.

Banerjee, Sudeshna Ghosh, Douglas Barnes, Kristy Singh Bipuland Mayer, and Hussain Samad, *Power for All: Electricity Access Challenge in India*, World Bank, 2014.

Barron, Manuel and Maximo Torero, "Household electrification and indoor air pollution," *Journal of Environmental Economics and Management*, November 2017, *86*, 81–92.

Beck, Hylke E., Niklaus E. Zimmermann, Tim R. McVicar, Noemi Vergopolan, Alexis Berg, and Eric F. Wood, "Present and future Köppen-Geiger climate classification maps at 1-km resolution," *Scientific Data*, October 2018, *5* (1).

Beckwith, Robin, "Depending On Guar For Shale Oil And Gas Development," *Journal of Petroleum Technology*, December 2012, *64* (12), 44–55.

Bensch, Gunther, Jörg Peters, and Maximiliane Sievert, "The lighting transition in rural Africa — From kerosene to battery-powered LED and the emerging disposal problem," *Energy for Sustainable Development*, August 2017, *39*, 13–20.

Bernard, Tanguy and Maximo Torero, "Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia," *Economic Development and Cultural Change*, April 2015, *63* (3), 459–484.

Bonan, Jacopo, Stefano Pareglio, and Massimo Tavoni, "Access to modern energy: a review of barriers, drivers and impacts," *Environment and Development Economics*, July 2017, *22* (05), 491–516.

Bos, Kristine, Duncan Chaplin, and Arif Mamun, "Benefits and challenges of expanding grid electricity in Africa: A review of rigorous evidence on household impacts in developing countries," *Energy for Sustainable Development*, June 2018, *44*, 64–77.

Burgess, Robin, Michael Greenstone, Nicholas Ryan, and Anant Sudarshan, "Demand for Electricity in a Poor Economy," 2019. Working paper.

Burlig, Fiona and Louis Preonas, "Out of the Darkness and Into the Light? Development Effects of Rural Electrification in India," 2016. Energy Institute at Haas Working Paper № WP-268R.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Manipulation Testing Based on Density Discontinuity," *The Stata Journal: Promoting communications on statistics and Stata*, March 2018, *18* (1), 234–261.

_ , _ , and _ , "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, July 2019, pp. 1–7.

Central Statistical Organization, "National Industrial Classification (All Economic Activities)," Technical Report, Ministry of Statistics and Program Implementation 2008.

Correia, Sergio, "Singletons, cluster-robust standard errors and fixed effects: A bad mix," 2015. Technical note.

Dinkelman, Taryn, "The Effects of Rural Electrification on Employment: New Evidence from South Africa," *American Economic Review*, December 2011, *101* (7), 3078–3108.

Duque, Valentina, Maria Rosales-Rueda, and Fabio Sanchez, "How do early-life shocks interact with subsequent human-capital investments? Evidence from administrative data," 2018. Working paper.

Elsner, Martin and Kathrin Hoelzer, "Quantitative Survey and Structural Classification of Hydraulic Fracturing Chemicals Reported in Unconventional Gas Production," *Environmental Science & Technology*, March 2016, *50* (7), 3290–3314.

Fetter, T Robert, "Fracking, Toxics, and Disclosure," 2019. Working paper.

_ , Andrew L Steck, Christopher Timmins, and Douglas H Wrenn, "Learning by Viewing? Social Learning, Regulatory Disclosure and Firm Productivity in Shale Gas," 2018. National Bureau of Economic Research (NBER) Working Paper № 25401.

Grimm, Michael, Anicet Munyehirwe, Jörg Peters, and Maximiliane Sievert, "A First Step up the Energy Ladder? Low Cost Solar Kits and Household's Welfare in Rural Rwanda," *The World Bank Economic Review*, October 2017, *31*, 631649.

__ , Luciane Lenz, Jörg Peters, and Maximiliane Sievert, "Demand for Off-Grid Solar Electricity: Experimental Evidence from Rwanda," *Journal of the Association of Environmental and Resource Economists*, May 2020, *7* (3), 417–454.

International Energy Agency, "World Energy Outlook 2011: Energy for All," Technical Report, OECD/IEA 2011.

__ , "WEO-2015 Special Report: India Energy Outlook," Technical Report, OECD/IEA 2015.

Jones, Damon, David Molitor, and Julian Reif, "What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study," *The Quarterly Journal of Economics*, August 2019, *134* (4), 1747–1791.

Khandker, Shahidur R., Douglas F. Barnes, and Hussain A. Samad, "Welfare Impacts of Rural Electrification: A Panel Data Analysis from Vietnam," *Economic Development and Cultural Change*, April 2013, *61* (3), 659–692.

Kuravadi, N A, S Verma, S Pareek, P Gahlot, S Kumari, U K Tanwar, P Bhatele, M Choudhary, K S Gill, V Pruthi, S K Tripathi, K S Dhugga, and G S Randhawa, "Guar," in "Agricultural Sustainability: Progress and Prospects in Crop Research," Elsevier, 2013, pp. 47–60.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram, "Does Household Electrification Supercharge Economic Development?," *Journal of Economic Perspectives*, February 2020, *34* (1), 122–144.

__ , __ , and __ , "Experimental Evidence on the Economics of Rural Electrification," *Journal of Political Economy*, April 2020, *128* (4), 1523–1565.

Lenz, Luciane, Anicet Munyehirwe, Jörg Peters, and Maximiliane Sievert, "Does Large-Scale Infrastructure Investment Alleviate Poverty? Impacts of Rwanda's Electricity Access Roll-Out Program," *World Development*, January 2017, *89*, 88–110.

Lipscomb, Molly, A. Mushfiq Mobarak, and Tania Barham, "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil," *American Economic Journal: Applied Economics*, January 2013, *5* (2), 200–231.

Litzow, Erin L., Subhrendu K. Pattanayak, and Tshering Thinley, "Returns to rural electrification: Evidence from Bhutan," *World Development*, September 2019, *121*, 75–96.

Martin, Philippe and Carol Ann Rogers, "Industrial location and public infrastructure," *Journal of International Economics*, November 1995, *39* (3-4), 335–351.

Meager, Rachael, "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments," *American Economic Journal: Applied Economics*, January 2019, *11* (1), 57–91.

Mudgil, Deepak, Sheweta Barak, and Bhupendar Singh Khatkar, "Guar gum: processing, properties and food applications—A Review," *Journal of Food Science and Technology*, October 2011, *51* (3), 409–418.

National Rainfed Area Authority, "Potential of Rainfed Guar (Cluster beans) Cultivation, Processing and Export in India," Technical Report, Ministry of Agriculture & Farmers Welfare 2014. Retrieved from http://nraa.gov.in/pdf/Rainfed-guar-final-pdf.pdf on April 16, 2017.

Orr, Isaac, "Hydraulic Fracturing: A Game-Changer for Energy and Economies," in "Alternative Energy and Shale Gas Encyclopedia," John Wiley & Sons, Inc., April 2016, pp. 700–719.

Pattanayak, S. K., M. Jeuland, J. J. Lewis, F. Usmani, N. Brooks, V. Bhojvaid, A. Kar, L. Lipinski, L. Morrison, O. Patange, N. Ramanathan, I. H. Rehman, R. Thadani, M. Vora, and V. Ramanathan, "Experimental evidence on promotion of electric and improved biomass cookstoves," *Proceedings of the National Academy of Sciences*, May 2019, *116* (27), 13282–13287.

Rai, Durgesh K, "Trends and Economic Dynamics of Guar in India," 2015. Indian Council for Research on International Economic Relations Working Paper № 311.

Singh, Santosh K., "An Analysis of Guar Crop in India," Technical Report, United States Department of Agriculture 2014.

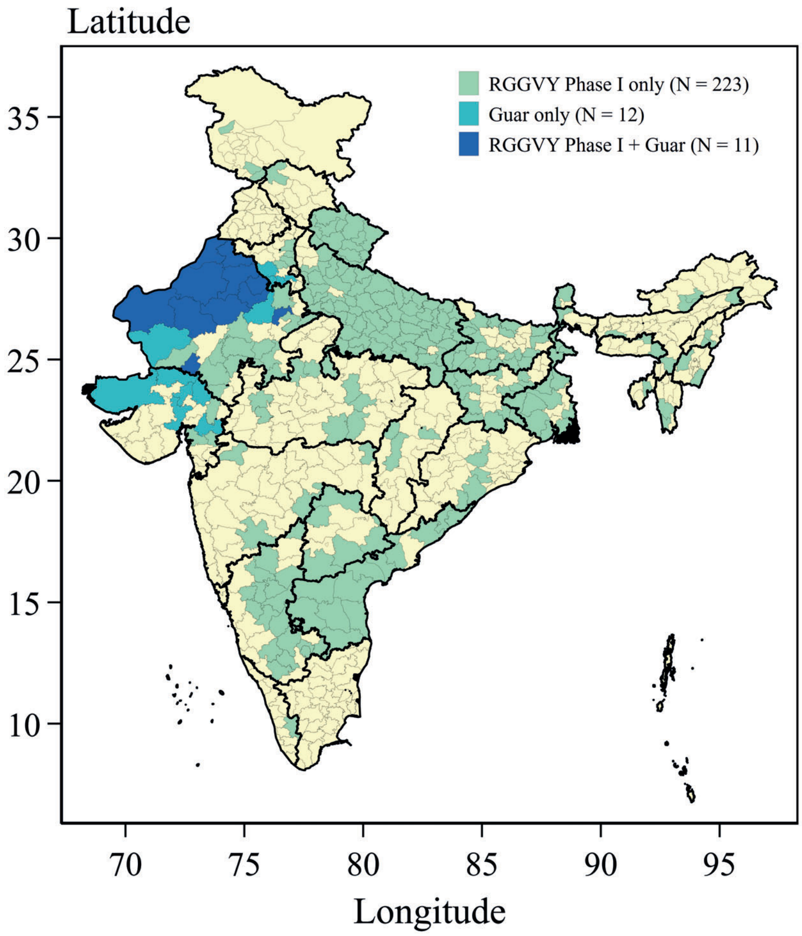
United Nations, "Transforming our world: the 2030 Agenda for Sustainable Development," Technical Report A/RES/70/1, United Nations General Assembly 2015.

Usmani, Faraz, Marc Jeuland, and Subhrendu K. Pattanayak, "NGOs and the effectiveness of interventions," May 2018. UNU-WIDER Working Paper № 59/2018.

Vivalt, Eva, "Heterogeneous Treatment Effects in Impact Evaluation," *American Economic Review*, May 2015, *105* (5), 467–470.

Westfall, Peter H. and S. Stanley Young, *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, Wiley-Interscience, 1993.

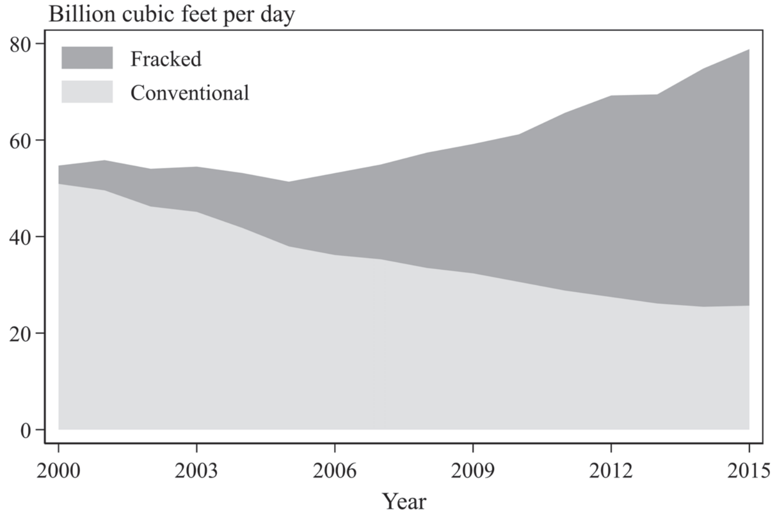
Figure 1: Districts of India, by guar-production and electrification status



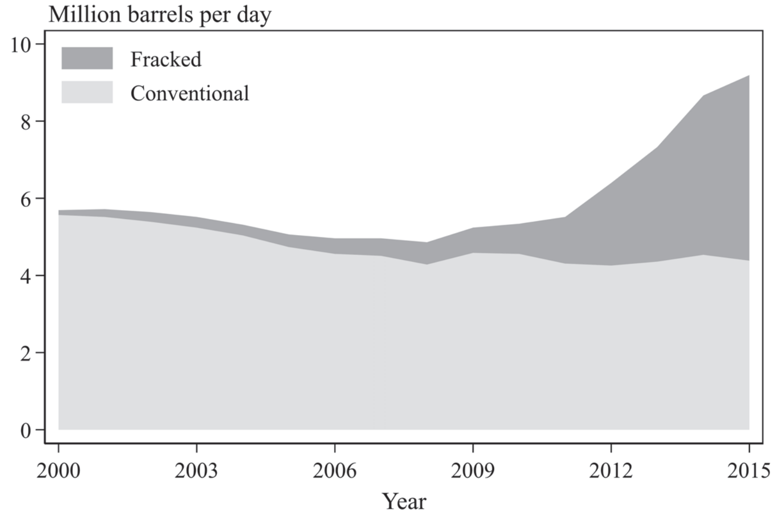
Notes. This map shows India's 2011 state (thick lines) and district (thin lines) boundaries. Districts are shaded by their electrification and guar-production status. Unshaded districts were neither approved for the roll-out of electrifcation as part of RGGVY Phase I nor contribute appreciably to guar production in India.

Figure 2: Natural gas and oil production in the United States, by source

(a) Marketed natural gas production

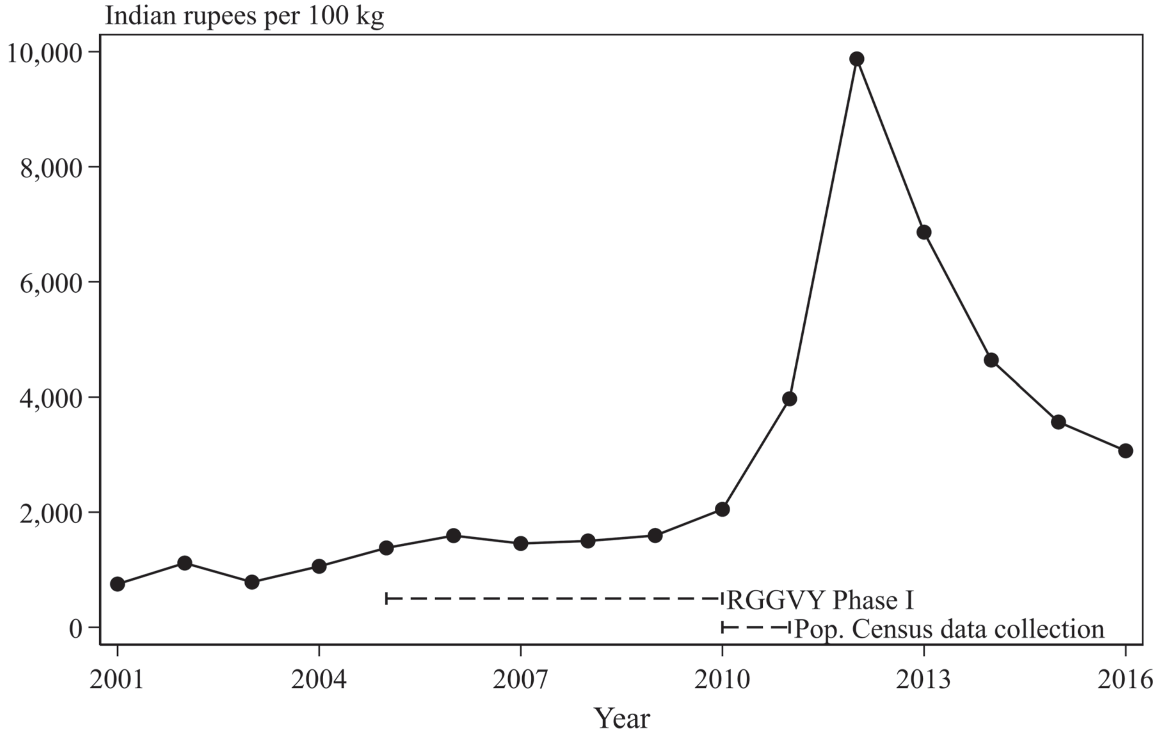


(b) Oil production



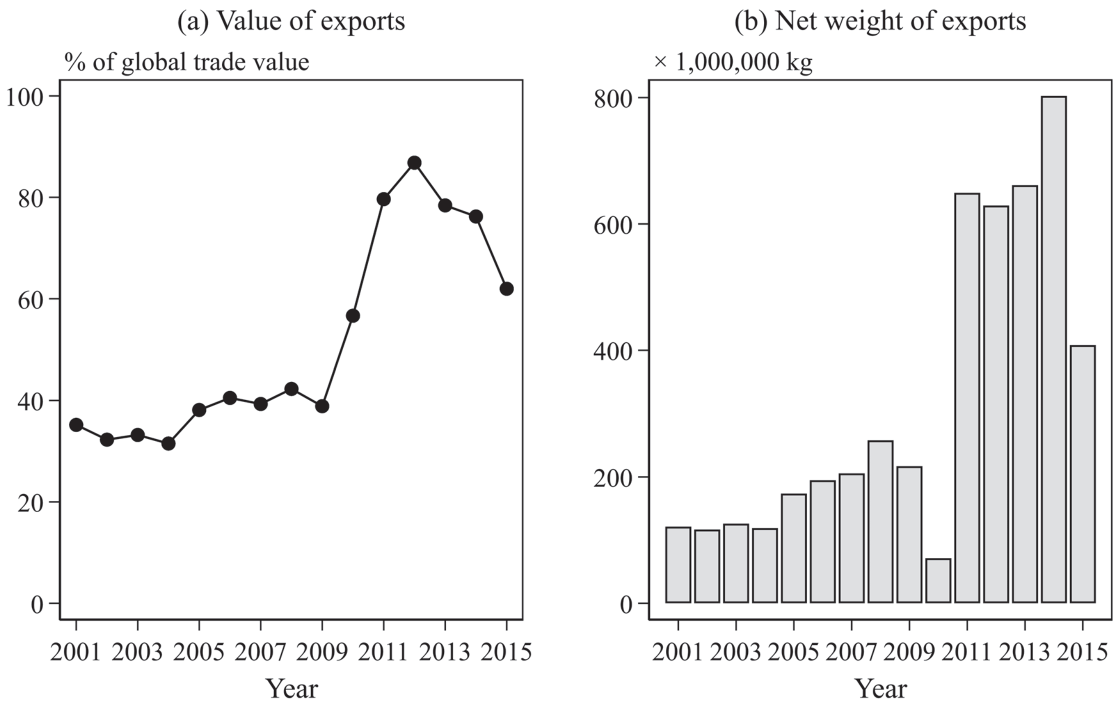
Notes. This figure shows marketed natural gas (panel *a*) and crude oil (panel *b*) produced from fracked and "conventional" wells in the United States between 2000 and 2015. Marketed natural gas production excludes natural gas used for repressuring the well, vented and flared gas, and any nonhydrocarbon gases. Source: United States Energy Information Administration, IHS Global Insight, and DrillingInfo, Inc, as outlined at https://www.eia.gov/todayinenergy/detail.php?id=26112 and https://www.eia.gov/todayinenergy/detail.php?id=25372.

Figure 3: Mean annual wholesale price of guar in India



Notes. This figure plots weighted annual means of the modal wholesale trade price per 100 kg of guar using price and quantity data from daily guar trades occurring at agricultural wholesale markets across India. These data are available from the Indian Ministry of Agriculture's Agricultural Marketing Information Network (AGMARKNET) portal (`https://agmarknet.gov.in`), which covers around 3,000 agricultural wholesale markets in total. Dashed lines indicate timelines for electrification under RGGVY Phase I and for data collection for the 2011 round of the Population Census of India.

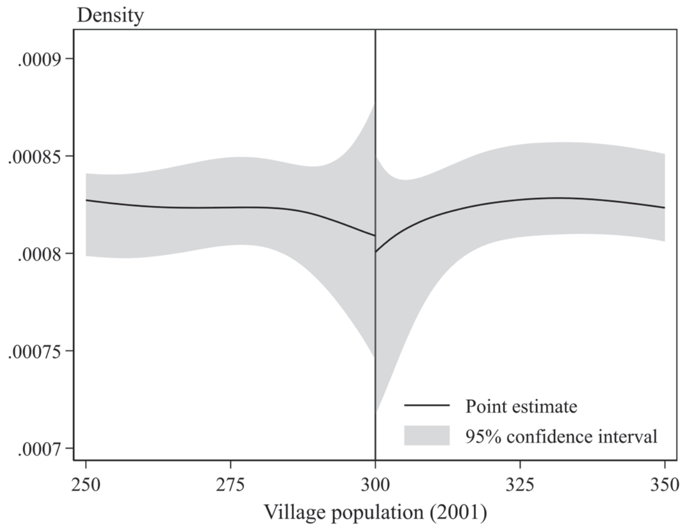
Figure 4: Value and net weight of India’s guar gum exports



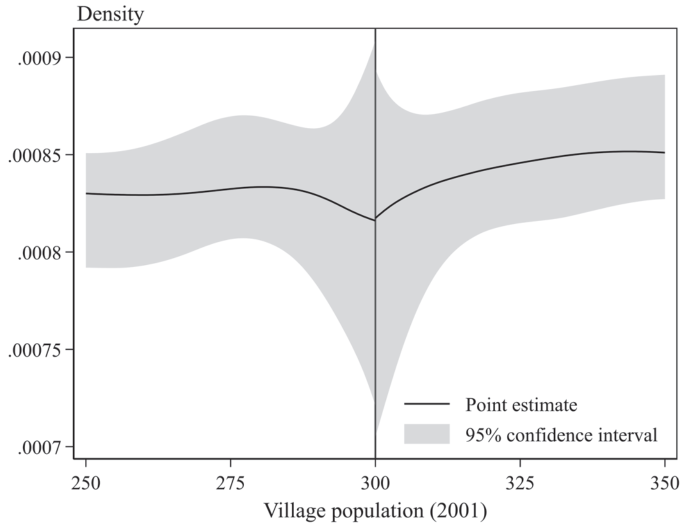
Notes. This figure shows the share of total global trade value (panel *a*) and net weight (panel *b*) of India’s exports of guar gum between 2001 and 2015 based on data for guar gum (product code HS 130232) from the United Nations Comtrade Database (`https://comtrade.un.org`). Guar cultivation in India exhibited a reduction in 2009–10 on account of drought conditions, resulting in a reduction in the weight of its guar gum exports (Rai, 2015).

Figure 5: Village population changes smoothly at RGGVY Phase I eligibility threshold

(a) All single-habitation villages

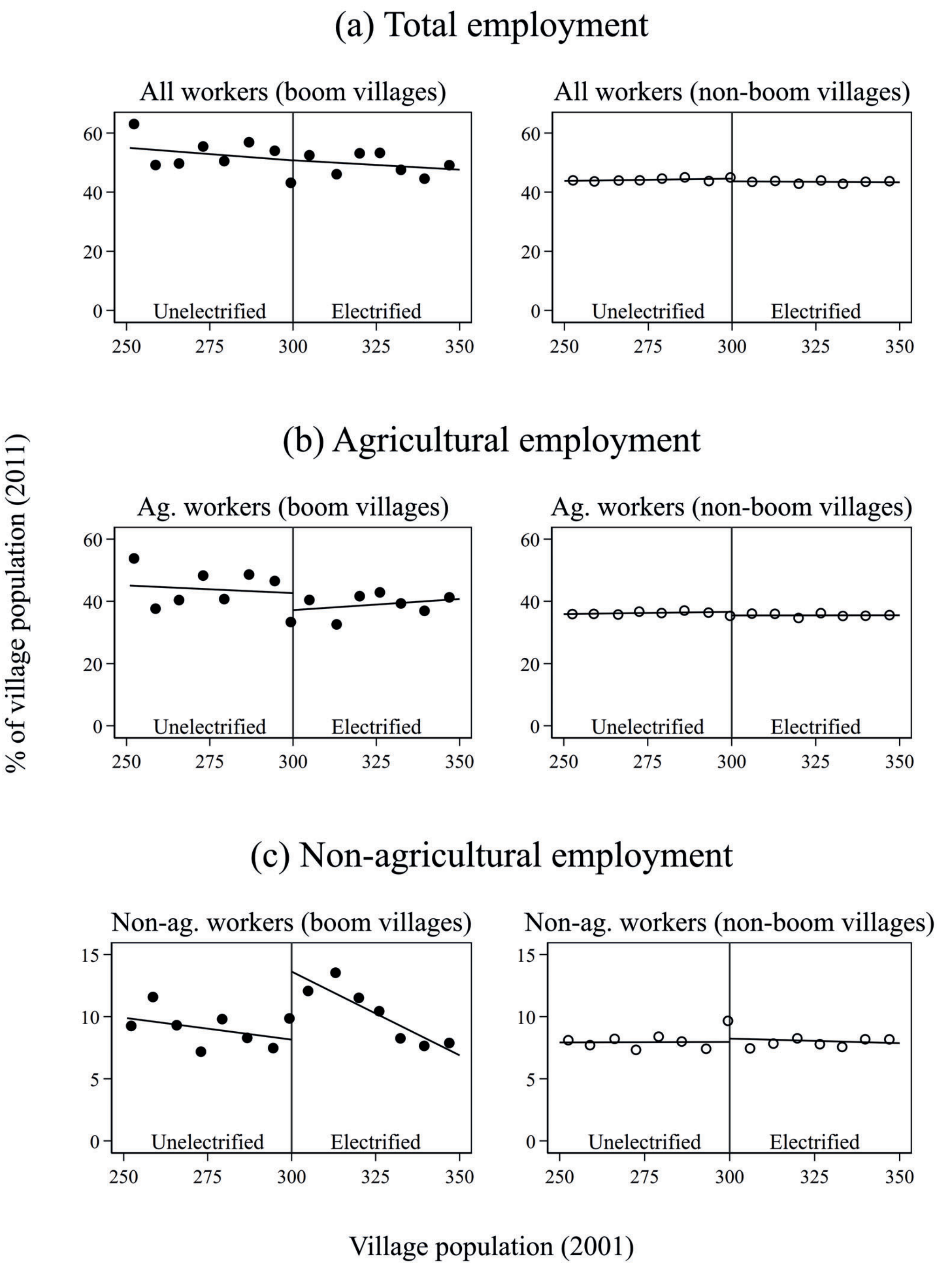


(b) RGGVY Phase I single-habitation villages



Notes. This figure shows the local polynomial density estimate (solid line) along with robust bias-corrected 95 percent confidence intervals (shaded area) for our RD running variable (2001 village population) following the RD manipulation testing procedure developed by Cattaneo et al. (2018, 2019). Panel *a* shows these results for all fuzzy-matched single-habitation villages within a fifty-person bandwidth of RGGVY's 300-person eligibility cutoff ($N = 14{,}668$); panel *b* is restricted to single-habitation villages located within districts approved for the roll-out of electrification as part of RGGVY Phase I ($N = 7{,}655$).

Figure 6: RD results of impact of electrification on short-term labor-market outcomes



Notes. This figure shows the results from estimating the regression discontinuity specification outlined in Equation (1). The outcome variables presented in this figure come from the Primary Census Abstract tables of the 2011 round of the Population Census. Column (1) of Table 1 reports corresponding numerical estimates for the sub-figures shown in panel (*a*); columns (1) and (4) of Table 2 report corresponding numerical estimates for the sub-figures shown in panels (*b*) and (*c*), respectively. Best-fit lines are constructed using predicted values from the regressions. Each solid (hollow) dot represents the mean value of the relevant outcome variable for approximately 10 (500) villages in fifteen-person bins.

Figure 7: Evaluating differential impact of electrification using randomization inference (placebo boom districts from RJ, HY and GJ)

Notes. Each panel of this figure plots the distribution of 1,000 estimated values of $\hat{\beta}_2$ from a randomization-based inferential procedure (Athey and Imbens, 2017). In each iteration, we randomly assign eleven RGGVY Phase I districts from Gujarat, Haryana, and Rajasthan—the three northwestern states in which all of the booming guar districts are located—to a placebo boom group, then re-estimate the regression discontinuity specification outlined in Equation (1) to obtain a $\hat{\beta}_2$ placebo value for the degree to which the guar boom augments the impact of electrification. The dashed vertical line indicates the corresponding estimated value of $\hat{\beta}_2$ reported in Table 2. Dark (light) shading represents the 90 (95) percent confidence interval of each distribution. Gujarat, Rajasthan and Haryana contain 31 RGGVY Phase I districts.

Table 1: RD estimates of short-term impact of electrification on total employment

		(1)	(2)	(3)
		All workers (% of 2011 population)		
		All	Male	Female
$\hat{\beta}_1$	$\mathbb{1}$ (Village pop. (2001) > 300)	−0.78	−0.13	−0.62
		(0.55)	(0.20)	(0.46)
$\hat{\beta}_2$	$\mathbb{1}$ (Village pop. (2001) > 300) × $\mathbb{1}$ (Village in boom district)	0.14	−0.13	0.07
		(2.57)	(1.37)	(1.43)
District FEs		Yes	Yes	Yes
Census (2001) controls		Yes	Yes	Yes
N		7649	7649	7649
Adjusted R^2		0.39	0.38	0.39
Mean of outcome		43.98	27.51	16.47

Notes. This table shows results from estimating Equation (1). The results in column (1) correspond to those presented graphically in panel (*a*) of Figure (6). The outcome variable for each regression comes from the Primary Census Abstract tables of the 2011 round of the Indian Census. Each regression includes all single-habitation villages in RGGVY Phase I districts with a 2001 population within a fifty-person bandwidth of RGGVY's 300-person eligibility threshold. Estimates associated with the population running variable $\left(\tilde{P}^{2001}_{vds}\right)$ are omitted. Following Correia (2015), six singleton observations are excluded. Standard errors—in parentheses—are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: RD estimates of short-term impact of electrification on agricultural and non-agricultural employment

		(1)	(2)	(3)	(4)	(5)	(6)
		Agricultural workers (% of 2011 population)			Non-agricultural workers (% of 2011 population)		
		All	Male	Female	All	Male	Female
$\hat{\beta}_1$	$\mathbb{1}$ (Village pop. (2001) $>$ 300)	−1.17**	−0.27	−0.91**	0.53	0.17	0.27
		(0.59)	(0.29)	(0.39)	(0.40)	(0.24)	(0.24)
$\hat{\beta}_2$	$\mathbb{1}$ (Village pop. (2001) $>$ 300) $\times$ $\mathbb{1}$ (Village in boom district)	−6.39***	−2.85***	−3.25**	5.60***	2.30**	3.22***
		(1.71)	(0.97)	(1.34)	(1.19)	(1.12)	(1.23)
District FEs		Yes	Yes	Yes	Yes	Yes	Yes
Pre-electrification controls		Yes	Yes	Yes	Yes	Yes	Yes
N		7649	7649	7649	7649	7649	7649
Adjusted R^2		0.37	0.36	0.38	0.16	0.25	0.07
Mean of outcome		35.96	22.27	13.68	8.02	5.23	2.79

Notes. This table shows results from estimating the regression discontinuity specification outlined in Equation (1). The results in columns (1) and (4) correspond to those presented graphically in panels (b) and (c), respectively, of Figure 6. The outcome variable for each regressions is constructed using data from the Primary Census Abstract tables of the 2011 round of the Population Census. Specifically, "agricultural workers" represents a village-level sum of main and marginal cultivators and agricultural laborers, while "non-agricultural workers" represents a village-level sum of main and marginal household-industry and workers. All regressions control for the 2001 value of the outcome variable. Each regression includes all single-habitation villages in RGGVY Phase I districts with a 2001 population within a fifty-person bandwidth of RGGVY's 300-person eligibility threshold with non-missing data for the respective outcome variable. Estimates associated with the population running variable $\left(\tilde{P}_{vds}^{2001}\right)$ are omitted. Following Correia (2015), six singleton observations are excluded. Standard errors—in parentheses—are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: RD estimates of longer-term impact of electrification on total employment

		(1)	(2)	(3)	(4)
		All workers (% of village population)			
		Post-boom/electrification	Falsification tests		
		2013	1990	1998	2005
$\hat{\beta}_1$	$\mathbb{1}$ (Village pop. (2001) $>$ 300)	-2.16	-2.59	-3.16	-2.25
		(2.95)	(2.77)	(3.80)	(2.72)
$\hat{\beta}_2$	$\mathbb{1}$ (Village pop. (2001) $>$ 300) $\times$	8.72**	2.61	1.81	3.41
	$\mathbb{1}$ (Village in boom district)	(4.25)	(3.00)	(4.11)	(3.13)
District FEs		Yes	Yes	Yes	Yes
Pre-period controls		No	No	No	No
N		6938	6920	5563	6561
Adjusted R^2		0.0054	0.0017	-0.0055	0.0015
Mean of outcome		5.67	3.55	5.78	4.78

Notes. This table shows results from estimating a modified version of Equation (1), as follows: $y^t_{vd} = \beta_1 T_v + \beta_2 T_v G_d + \beta_3 \tilde{P}_v^{2001} + \beta_4 T_v \tilde{P}_v^{2001} + \beta_5 G_d \tilde{P}_v^{2001} + \beta_6 T_v G_d \tilde{P}_v^{2001} + \gamma_d + \varepsilon_{vd}$, where y^t_{vd} represents the number of workers (as a percentage of population) in year t in village v in district d. The total number of workers is derived by Asher et al. (2019) from the corresponding round of the Economic Census of India; village population for each corresponding year is imputed from the 2001 and 2011 figures for village population in the Population Census of India using a compound annual growth rate approach. Each regression includes all single-habitation villages in RGGVY Phase I districts with (i) a 2001 population within a fifty-person bandwidth of RGGVY's 300-person eligibility threshold; and (ii) non-missing data on village-level population of workers. Estimates associated with the population running variable $\left(\tilde{P}_{vds}^{2001}\right)$ are omitted. Following Correia (2015), 7, 8, 13 and 10 singleton observations are excluded from the regressions in columns (1), (2), (3) and (4), respectively. Standard errors—in parentheses—are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Impact of electrification on count and size of establishments

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of establishments (% of district total)			Number of employees		
	All	< 10 employees	⩾ 10 employees	All	< 10 employees	⩾ 10 employees
$\mathbb{1}$(Boom district) × $\mathbb{1}$(Boom industry)	−0.40*	−0.40*	−0.0071	0.70	−0.12	3.98
	(0.22)	(0.21)	(0.0099)	(1.22)	(0.18)	(7.92)
$\mathbb{1}$(RGGVY Phase I district) × $\mathbb{1}$(Boom industry)	−0.23	−0.23	−0.0028	−0.30	−0.0095	−1.28
	(0.22)	(0.22)	(0.0062)	(0.42)	(0.091)	(4.18)
$\mathbb{1}$(Boom district) × $\mathbb{1}$(RGGVY Phase I district) × $\mathbb{1}$(Boom industry)	0.31*	0.30*	0.0084**	−0.042	0.094	3.04
	(0.16)	(0.16)	(0.0033)	(0.73)	(0.084)	(7.46)
$\mathbb{1}$(Boom industry) × $\mathbb{1}$(Post)	0.15	0.16	−0.012	0.075	0.011	9.81***
	(0.19)	(0.19)	(0.0073)	(0.23)	(0.13)	(1.55)
$\mathbb{1}$(Boom district) × $\mathbb{1}$(Boom industry) × $\mathbb{1}$(Post)	−0.015	−0.019	0.0048	−0.93	−0.018	−9.66**
	(0.10)	(0.11)	(0.0081)	(1.23)	(0.15)	(4.68)
$\mathbb{1}$(RGGVY Phase 1 district) × $\mathbb{1}$(Boom industry) × $\mathbb{1}$(Post)	0.12	0.12	0.0048	0.077	0.024	−4.08
	(0.13)	(0.13)	(0.0029)	(0.20)	(0.087)	(4.29)
$\mathbb{1}$(Boom district) × $\mathbb{1}$(RGGVY Phase 1 district) × $\mathbb{1}$(Boom industry) × $\mathbb{1}$(Post)	0.23*	0.23*	0.0059	0.57	−0.26*	19.3***
	(0.13)	(0.13)	(0.013)	(1.49)	(0.15)	(6.54)
Industry FEs	Yes	Yes	Yes	Yes	Yes	Yes
District ×Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Unit of analysis	Industry–District	Industry–District	Industry–District	Establishment	Establishment	Establishment
N	251,860	251,860	251,860	96,134,462	94,360,948	1,773,512
Adjusted R^2	0.42	0.42	0.12	0.0072	0.23	0.015
Mean of outcome	0.51	0.50	0.010	2.75	2.21	31.6

Notes. This table shows results from estimating Equation (2). Columns (1)–(3) present results from estimating the specification using an industry–district-level panel. Columns (4)–(6) present results from estimating the specification using a district-level repeated cross-section of establishments. An establishment is assumed to belong to a "boom industry" if its 2008 National Industrial Classification (NIC) code is one of the following: (i) Support activities to agriculture and post-harvest crop activities (016); (ii) Manufacture of basic chemicals, fertilizer and nitrogen compounds, plastics and synthetic rubber in primary forms (201); (iii) Manufacture of prepared animal feeds (108); (iv) Manufacture of non-metallic mineral products (239); and (v) Wholesale of agricultural raw materials and live animals (462). Standard errors—in parentheses—are clustered at the industry level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix to "Fracking, farmers, and rural electrification in India"

T. Robert Fetter†
Duke University

Faraz Usmani‡
Mathematica

†Nicholas Institute for Environmental Policy Solutions, Duke University, Durham, NC, USA. rob.fetter@duke.edu.
‡International Unit, Mathematica, Washington, DC, USA. fusmani@mathematica-mpr.com.

Appendix A Using nighttime luminosity to evaluate the impact of the fracking-induced guar boom on economic activity

Did the fracking-induced guar boom in northwestern India have a meaningful impact on economic activity? To answer this question, we rely on the synthetic control methodology (SCM) applied to two decades of nighttime luminosity data covering nearly all of India's approximately 600,000 villages. We find that guar-growing districts shine brighter at night as a result of the start of the guar boom than a synthetic "counterfactual." As nighttime luminosity is a widely accepted proxy for regional economic activity, these results point to a large increase in economic activity in India's guar-growing regions due to the start of the United States' fracking boom.

A.1 Synthetic control methodology

Like the conventional difference-in-differences estimator, the SCM relies on differences between "treated" and "untreated" units before and after an event of interest (Abadie and Gardeazabal, 2003; Abadie et al., 2010). However, SCM does not give equal weight to all untreated units. Instead, it hinges on using a linear combination of untreated units to generate a weighted average whose pre-treatment outcome trends closely match those of the treated unit. This synthetic "counterfactual" unit is then projected into the post-treatment period and compared with the treated unit to gauge the direction and magnitude of impacts.

This feature makes it particularly attractive for estimating treatment effects in small-sample settings such as our own, in which only 23 mostly contiguous districts in nothwestern India are assumed to be "treated" by the fracking boom. Indeed, many applications have featured only one treated unit that is compared with multiple untreated units over time (e.g., Coffman and Noy, 2011; Singhal and Nilakantan, 2016).

Formally, let T_0 represent the number of pre-treatment periods (out of T total periods) and J represent the number of untreated units. Let $\mathbf{W} = (w_1, \ldots, w_J)$ be a $(J \times 1)$ vector of non-negative weights such that $\sum_{j=1}^{J} w_j = 1$. Each $w_j \in \mathbf{W}$ represents the weight of the j^{th} untreated unit. Let $\mathbf{Y}_1$ be a $(T_0 \times 1)$ vector of outcome measures in the treated unit for each pre-treatment period t. Similarly, let $\mathbf{Y}_0$ be a $(T_0 \times J)$ matrix that contains the same outcome measures for each untreated unit j in pre-treatment period t. Broadly, the aim of the SCM is to pick $\mathbf{W}^*$ such that:

$$\mathbf{Y}_1 = \mathbf{Y}_0 \mathbf{W}^*. \tag{A.1}$$

Applications of the SCM typically specify a set of k pre-treatment characteristics $\mathbf{X}$ as predictors, where $\mathbf{X}$ includes observed covariates $\mathbf{Z}$ that are unaffected by the treatment as well as linear combinations of the pre-treatment outcomes $\mathbf{Y}$. Given $\mathbf{Y}$ and $\mathbf{X}$, $\mathbf{W}$ is picked so as to minimize the root-mean-squared prediction error (RMSPE) of the predictors:

$$\mathbf{W}^* = \underset{\mathbf{W}}{\arg\min} \left\{ \sqrt{(\mathbf{X}_1 - \mathbf{X}_0\mathbf{W})' \mathbf{V} (\mathbf{X}_1 - \mathbf{X}_0\mathbf{W})} \right\}, \tag{A.2}$$

where the subscripts denote treated and untreated units as in Equation (A.1), and $\mathbf{V}$ represents a

$(k \times k)$ matrix that specifies the relative importance of the predictors.[28] Placebo tests determine the statistical significance of the effects observed in the post-treatment period. Specifically, the treated unit is excluded from the sample, and the analysis is repeated for each untreated unit, which is now assumed to have been treated instead. The presence of many large effects in the resulting distribution of post-treatment placebo effects suggests that the original estimated effect may have been the result of chance.[29]

A.2 Nighttime luminosity

Nighttime luminosity measures are increasingly used by economists to investigate changes in regional economic activity over time (Doll et al., 2006; Henderson et al., 2012). Recent applications also demonstrate that they serve as useful proxies for information on socioeconomic outcomes in low-income settings, where high-quality statistical data are often missing (Chen and Nordhaus, 2011; Pinkovskiy and Sala-i-Martin, 2016). This work typically uses data generated as part of the Defense Meteorological Satellite Program (DMSP) led by the National Oceanic and Atmospheric Administration (NOAA). DMSP satellites take pictures of the Earth every night. NOAA processes and cleans these nightly images to remove irregularities (such as cloud cover or solar glare), averages them across years, and makes the annual composite images publicly available.[30] Each pixel of these annual images—representing 30 arc seconds or approximately 1 km^2 at the equator—is assigned a number on a relative brightness scale ranging from 0 to 63.

Most prior research has relied on these annual composites. While annual averages certainly provide useful information, they smooth away substantial variation in brightness over the calendar year and are, therefore, less precise (Min et al., 2017). We use a considerably richer dataset of monthly village-level nighttime luminosity measures developed by Gaba et al. (2016), who revisit the complete archive of raw visible band (VIS) imagery captured during every night in India between 1993 and 2013 to generate each observation. Because the DMSP includes multiple satellites, this archive consists of approximately 30,000 high-resolution image strips. Brightness values are extracted from these images for each date from each pixel corresponding to the latitude and longitude of each of India's approximately 600,000 villages. These values are processed in line with NOAA recommendations to remove irregularities, and the resulting 4.4 billion observations are aggregated to the village-month level by taking the median measurement for each village over the course of a month. In addition, because the 0–63 relative brightness levels in the raw data are not directly comparable over time, additional image processing and background noise reduction procedures are applied to generate statistically recalibrated visible band (SR-VIS) measures, which enable more reliable comparisons both cross-sectionally and across time.[31]

We use these data to evaluate differential impacts of the fracking-induced guar boom on nighttime luminosity—and, by proxy, economic activity—across boom and non-boom regions of

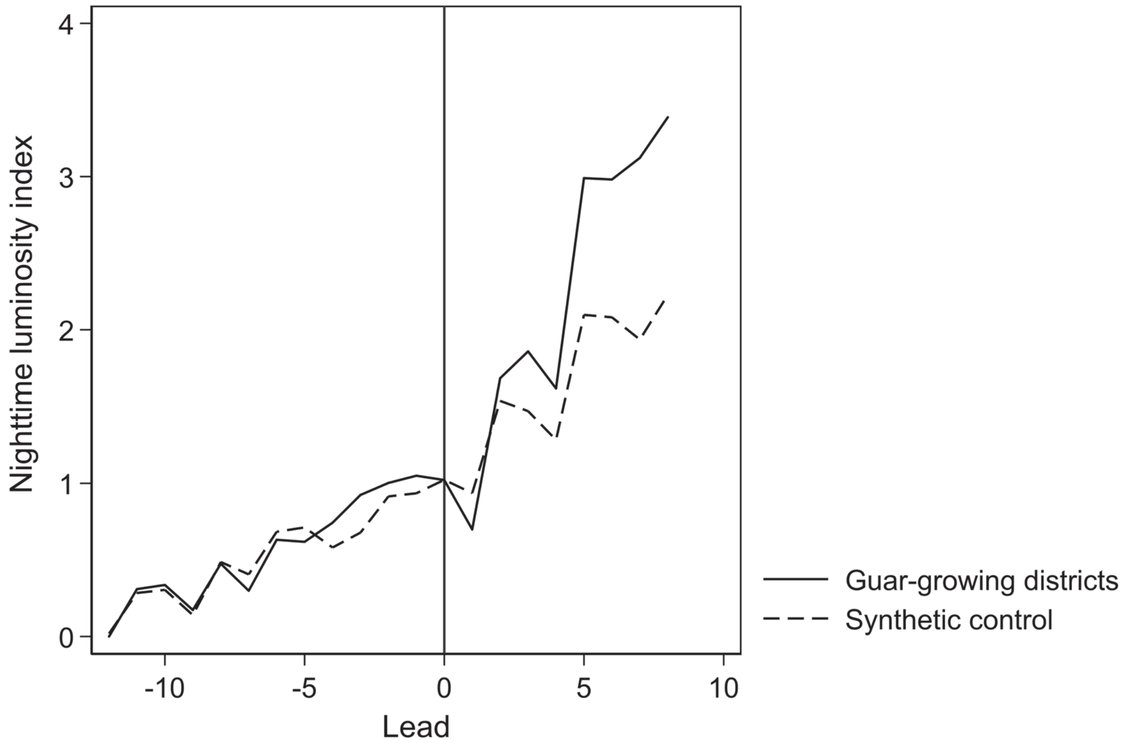
[28] Abadie and Gardeazabal (2003) choose **V** so as to minimize the RMSPE of the outcome variable in the pre-treatment period.

[29] Given the geographical spread of the guar shock across many districts in northwestern India, our analysis relies on an extension to this basic approach developed by Cavallo et al. (2013), who generalize the application of SCM to multiple treated units possibly at different time periods.

[30] NOAA's annual composite images are available at https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html.

[31] Min et al. (2017)—who use SR-VIS data to study power-supply irregularity across rural India—describe these image-processing procedures in more detail. The data are available at http://api.nightlights.io/.

Figure A1: Pre-/post-guar-boom trends in nighttime luminosity in guar-growing districts



Notes. This figure presents results from a synthetic control approach to evaluate the impact of the start of the fracking-induced guar boom in India on nighttime luminosity in India's guar-growing districts (as shown in Figure 1). The outcome variable is an index of nighttime luminosity, aggregated to the district-year level from the village-month level. The fracking-induced guar boom is assumed to begin in 2006, indicated by the vertical line. Other years (covering the period 1993-2013) are presented as leads and lags relative to 2006.

India. Because we identify guar-growing regions of India at the district level, in our analysis we rely on district-month measures of nighttime brightness.[32]

A.3 Results

We specify a parsimonious predictive model of nighttime luminosity, namely, one in which nighttime luminosity in district d in year t is a function of nighttime luminosity in year $t-1$.[33] Figure A1 presents our main results. The solid line highlights the trend in mean monthly nighttime brightness for India's guar-growing districts. The vertical line represents the start of the fracking boom in the United States (assumed to be 2006). The dashed line represents mean monthly nighttime brightness

[32]Gaba et al. (2016) determine these by identifying the median village light output within each district boundary for each month.

[33]Prior applications of the SCM have often used contemporaneous or lagged values of the outcome variable for all units j' as the sole predictor in estimation of treatment effects for unit j (e.g., Acemoglu et al., 2016). The justification for this approach is that the outcome variable fully characterizes all observed and unobserved determinants.

for a “counterfactual” set of guar-growing districts (unaffected by the fracking-induced guar boom). As described earlier, this is generated by estimating a set of weights for monthly nighttime brightness data for all other Indian districts over the pre-fracking-boom period (1993-2005) that are used to most closely track pre-boom—and predict post-boom—nighttime brightness trends in the guar-growing areas. The divergence in the two lines in the post-boom period is stark, and suggests that the start of fracking-induced boom resulted in sizable increases in nighttime brightness—and, by extension, economic activity in India’s guar-growing regions. Indeed, p-values estimated year-by-year using placebo tests for each post-boom year indicate that by 2011, the probability of this increased economic activity being detectable from space in this way by chance is extremely low (Table A1).

Table A1: Impact of fracking-induced guar boom on nighttime luminosity in Rajasthan

(1) Year	(2) Estimated coefficient	(3) p-value
2007	-0.24***	0.0006
2008	0.15	0.58
2009	0.39**	0.04
2010	0.33**	0.01
2011	0.89	0.14
2012	0.90**	0.03
2013	1.19***	0.004

Notes. This table presents the estimate effect of the fracking-induced guar boom on nighttime luminosity in India’s guar-growing districts (relative to a synthetically generated set of guar-growing districts) for each post-boom year (column 2). Column (3) presents p-values associated with each estimated coefficient, obtained by adjusting the observed effect sized by the pre-treatment match quality as outlined by Cavallo et al. (2013). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix B Electrification and household time allocation

More formally, changes in individuals' productive potential can be captured in an application of the basic home-production and household time-allocation model (Gronau, 1977). In this framework, the representative individual in household i obtains utility from consumption (c_i) and leisure $\left(t_i^l\right)$. Consumption is generated through a home-production function:

$$c_i = c\left(t_i^h, x_i, v_i; \psi_i\right), \tag{B.1}$$

where t_i^h is the time allocated to home-based work; x_i is a numeraire input to home production that is purchased in the market; and v_i is non-labor income. The production productivity parameter ψ_i is determined by

$$\psi_i = f\left(\eta_i, \varepsilon_i, \gamma\right), \tag{B.2}$$

where η_i represents the household's electrification status on a continuous scale, thus capturing both access and quality. Productivity is also determined by household- and community-level factors represented by ε_i and γ, respectively. For instance, households' stock of education and health can drive the labor productivity of its members. Community-level characteristics—such as weather, institutions, and differences in local or regional economic conditions—can play a similar role.

The problem of the household's representative individual is then given by

$$\max_{c_i, t_i^l} u_i = u\left(c_i, t_i^l; \psi_i\right), \tag{B.3}$$

subject to time and budget constraints, given by

$$t_i^m + t_i^h + t_i^l \leqslant T \tag{B.4}$$

$$x_i \leqslant w_i t_i^m + v_i, \tag{B.5}$$

where t_i^m is the time allocated to market-based work; T is the total time endowment; and w_i is the market wage. Equations (B.4) and (B.5) together yield the household's full-income constraint:

$$w_i T + v_i = x_i + w_i\left(t_i^h + t_i^l\right). \tag{B.6}$$

The Lagrangian associated with the household's problem described in Section 3.1 is as follows:

$$\max_{c_i, t_i^l} \mathscr{L} = u\left(c\left(t_i^h, x_i, v_i; \psi_i\right), t_i^l\right) + \lambda\left(w_i T + v_i - x_i - w_i\left(t_i^h + t_i^l\right)\right). \tag{B.7}$$

Ignoring the i subscripts, this yields the following first-order conditions for an interior solution:

$$\mathscr{L}_{t^l} = u_{t^l} - \lambda w = 0 \tag{B.8}$$

$$\mathscr{L}_{t^h} = u_c c_{t^h} - \lambda w = 0 \tag{B.9}$$

$$\mathscr{L}_x = u_c c_x - \lambda = 0 \tag{B.10}$$

$$\mathscr{L}_\lambda = wT + v - x - w\left(t^h + t^l\right) = 0. \tag{B.11}$$

These first-order conditions indicate that household's time allocations are chosen to equate the marginal rate of substitution between leisure and consumption with (i) the shadow value of home production; and (ii) the shadow value of market-based activities. Specifically, from Equations (B.8), (B.9) and (B.10):

$$\frac{u_{t^l}}{u_c} = c_{t^h} = c_x w. \tag{B.12}$$

From this, the general form of the household's optimal time allocation to home production is obtained:

$$t^{h*} = f_{t^h}(w, v; \psi). \tag{B.13}$$

Equations (B.8), (B.10) and (B.11) can be solved jointly to obtain the household's optimal time allocation to leisure and its demand for the market-purchased home-production input:

$$t^{l*} = f_{t^h}(w, v; \psi) \tag{B.14}$$

$$x^* = f_x(w, v; \psi). \tag{B.15}$$

Equation (B.15) and Equation (B.13) combined with the household's consumption production function yield the household's optimal consumption:

$$c^* = c\left(t^{h*}, x^*, v; \psi_i\right). \tag{B.16}$$

Finally, combining the household's time constraint with Equations (B.13) and (B.14) yields the household's time allocation to market-based activities:

$$t^{m*} = T - t^{h*} - t^{l*} = f_{t^m}(w, v; \psi). \tag{B.17}$$

We look to investigate how changes in the household's access to electricity (η_i) interact with community-level factors (γ) to influence the household's productive potential (ψ_i) and determine the time it allocates to home production, leisure, and market-based activities. Specifically, we exploit exogenous variation in levels of economic activity across boom and non-boom regions to shed light on how and why differences in the impacts of electricity access can emerge.

There are at least two reasons our model does not offer a clear answer to this question. First, even if we assume that an improvement in the household's access to electricity increases its productive potential, additional assumptions are necessary about the shape of the home-production function in Equation (B.1) to predict how changes in productivity due to simultaneous changes in electrification and community-level characteristics influence time allocation. Second, even with such assumptions in place, variation in household-level preferences over labor and leisure—the shape of the household utility function—may give rise to counteracting income and substitution effects. Indeed, an increase in its productive potential may ultimately induce a household to allocate *less* time to income-generating activities.

This ambiguity is compounded by the role of household-level characteristics (ε_i). The household's opportunity cost of leisure is determined by factors such as its stock of education and health, liquidity or credit constraints, or its "entrepreneurial spirit." Thus, how the impacts of electrification on labor-market outcomes vary with economic conditions is ultimately a question best answered with data. Our setting allows a unique opportunity to address this question.

Appendix C Habitation–village matching procedure

We use a multi-step matching procedure to identify villages eligible for electrification under RGGVY Phase I based on the populations of their constituent habitations, and identify corresponding village names from the 2001 and 2011 Census to those in the 2009 census of habitations conducted by the National Rural Drinking Water Program (NRDWP). The NRDWP habitation census covers 1.65 million habitations in 574,259 villages.[34] Because the NRDWP survey indicates only the name of each village (and not its unique Census code), matching on names is necessary; however, not all village names match exactly between the names used in NRDWP and those used in the Census, even conditional on an exact match for state and district. Accordingly, our matching process incorporates a combination of exact and fuzzy name matches, prioritizing exact matches where possible.

We treat the 2001 Primary Census Abstract (PCA) villages as the master dataset. As a first step for matching village names with the 2009 NRDWP habitations data, we standardize state, district, block, and village names to correct minor differences in spelling between the names in use by the NRDWP and the Census. We also account for districts that were renamed between 2001 and 2009. Our procedure for standardizing state and district names is sufficiently comprehensive to achieve a 100 percent match among state and district names between the NRDWP and Census, except for a handful of cases where districts are split or combined (not just renamed) between 2001 and 2009.[35]

We use information from the state, district, block, and village level, and prioritize exact matches. Where exact name matches are not possible, we employ a fuzzy match, using the "masalafied Levenshtein" distance and "Masala merge" code in Stata and Python (Asher and Novosad, 2020). This is a modification of the standard Levenshtein string distance metric, one that lowers the cost of certain substitutions that are common in Indian languages.[36] We thus create a five-tier matching hierarchy:

1. Exact match on state, district, block, and village name;
2. Exact match on state, district, and village name, with a fuzzy match on block name;
3. Exact match on state and district name, with a fuzzy match on block and village name;
4. Exact match on state, district, and village name (without regard to block name); and
5. Exact match on state and district name, with a fuzzy match on village name (without regard to block name).

[34] This includes five of the seven Union Territories—Chandigarh, Dadra and Nagar Haveli, Daman and Diu, Lakshadweep, and Puducherry—and Goa. However, we exclude these from the merge process because Goa and all seven Union Territories were fully electrified prior to 2005, so were excluded in RGGVY (Ministry of Power, 2012). Excluding the seven Union Territories and Goa, the 2009 survey covers 1.65 million habitations in 573,702 villages.

[35] One approach to match villages in split or combined districts would be to geolocate all villages from the old district(s) into the new district(s). We take a somewhat less intensive approach and look for name-based village matches in a proper subset of the old or new district area—specifically, an area of known overlap between old and new. For instance, Tiruppur district in Tamil Nadu was formed in 2009 from parts of Coimbatore and Erode. Among villages in the NRDWP belonging to Tiruppur district, we look for matching Census village names within Erode district, but not within Coimbatore district. We also flag any matches associated with split or combined districts. We have run our matching algorithm excluding these flagged matches and, after completing all five steps of the multi-step procedure, achieved virtually identical results.

[36] Additional information about Masala merge (including its underlying code) is available at https://github.com/paulnov/masala-merge.

Of the 563,338 villages in the 2001 Census, we match 531,325 to villages in the NRDWP habitation census (94.3 percent). Of these village matches, about 75 percent (400,457) are exact matches (i.e., the first and fourth tiers of our five-tier matching hierarchy), and 48 percent (271,774) are exact matches on state, district, block, and village name.[37] Further, our algorithm achieves a 90 percent or greater match rate across every state with the exception of Tripura (36 percent), Tamil Nadu (76 percent), Jammu and Kashmir (78 percent), Nagaland (82 percent), and Assam (83 percent). We also match at least 96 percent of villages in each of the three northwestern states where guar is produced (98 percent in Rajasthan and Gujarat, and 96 percent in Haryana).

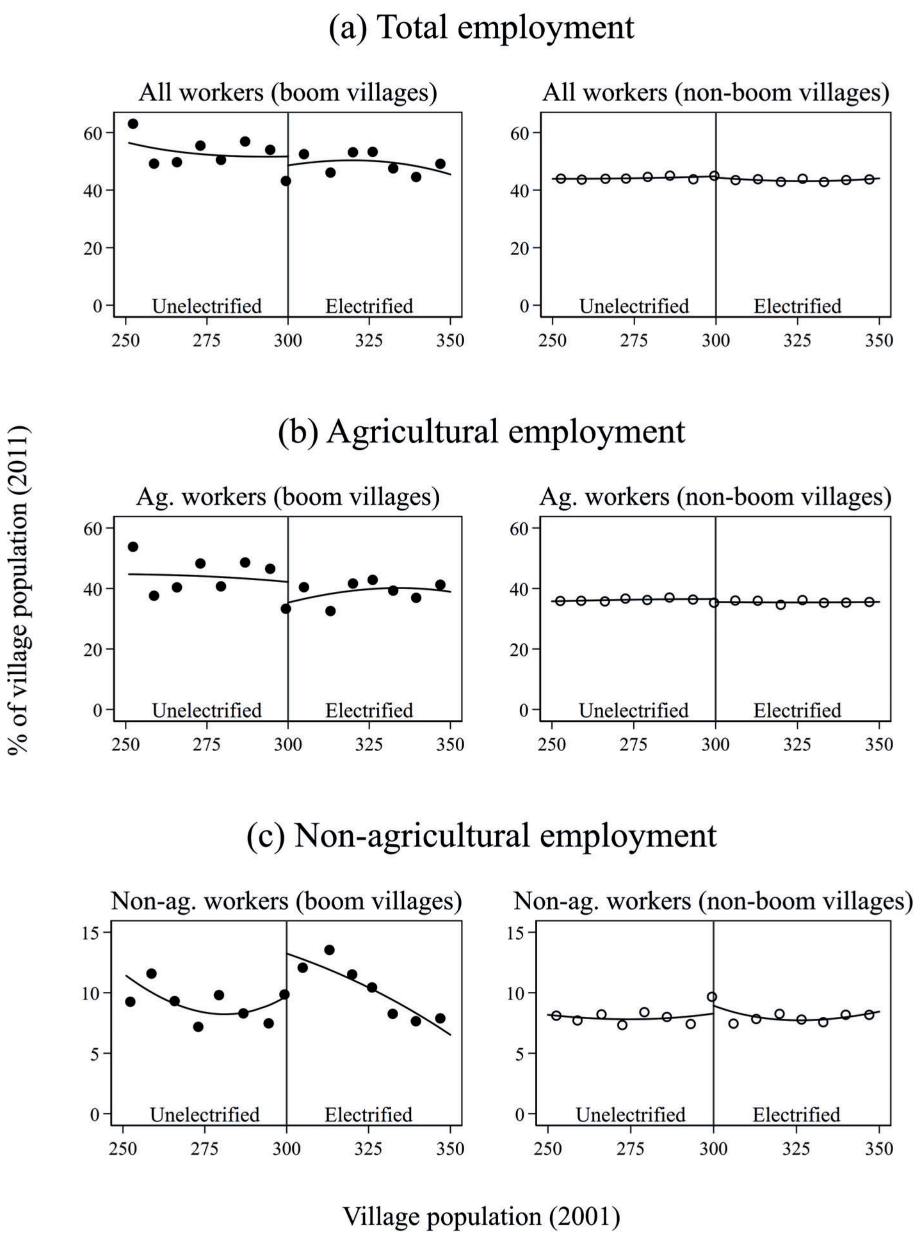
As a further verification step, we compare the village population recorded by the NRDWP in 2009 to the village population recorded by the 2011 PCA. For any village name match in which these figures diverge by more than 20 percent, we exclude the village from the matched set.[38] Using this matched sample, we identify single-habitation villages, and use the population of each of these in the 2001 round of the Census to gauge its eligibility for electrification under RGGVY Phase I.

[37] Our match rate is comparable to others in the literature. For instance, Burlig and Preonas (2016) report matching 86 percent of villages from the 2003 and 2009 NRDWP habitation surveys to corresponding Census villages. While Asher and Novosad (2020) do not report a village-level match rate, they do indicate they matched over 85 percent of habitations listed in the PMGSY to corresponding Census villages. Aggarwal (2018) and Kaczan (2020), who also evaluate the impact of India's rural roads program, report match rates of 80 and 83 percent, respectively.

[38] We have also run our analysis using thresholds other than 20 percent and find substantially similar results (Figure D3).

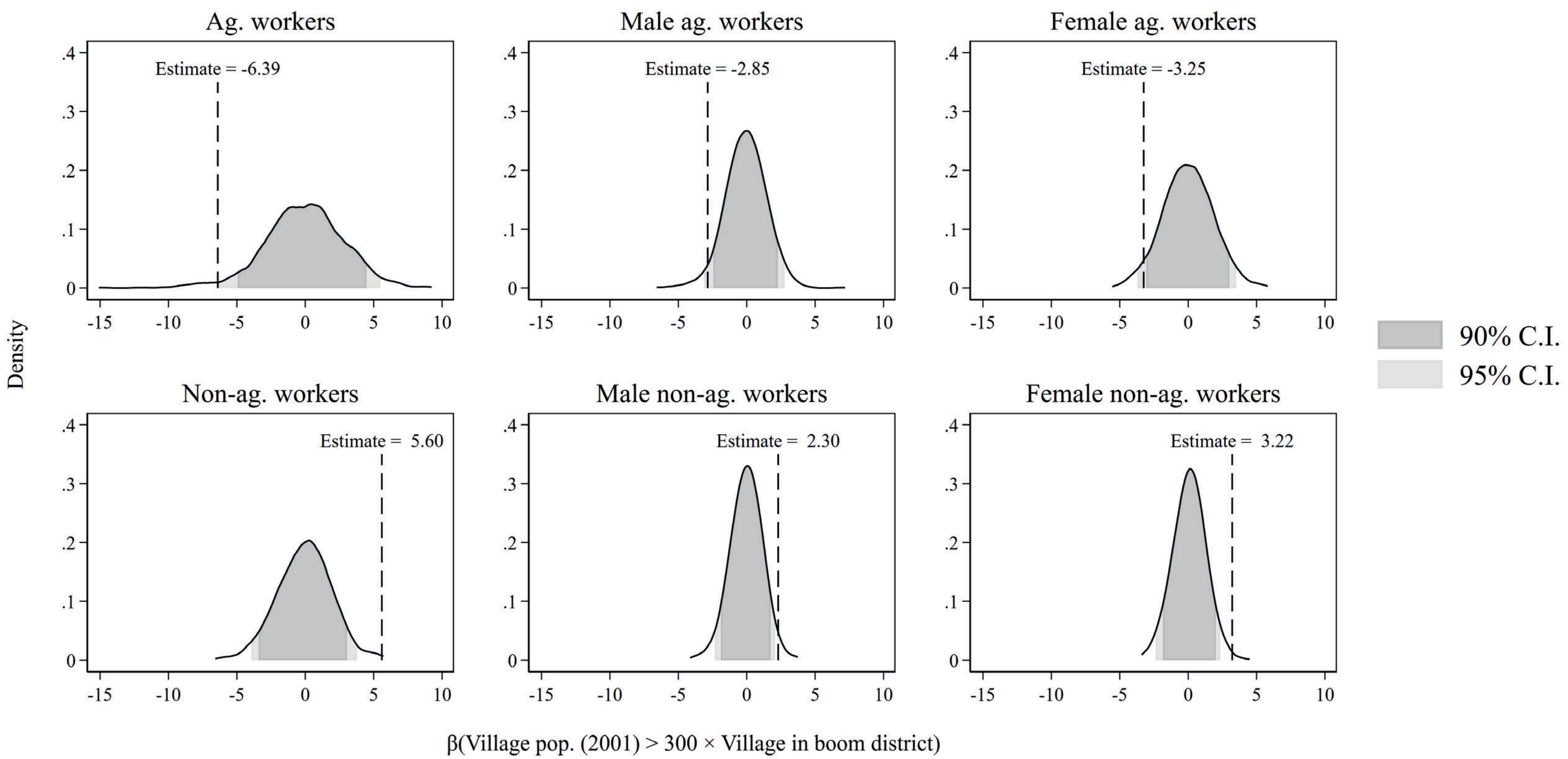
Appendix D Additional Figures

Figure D1: RD results (quadratic interaction)



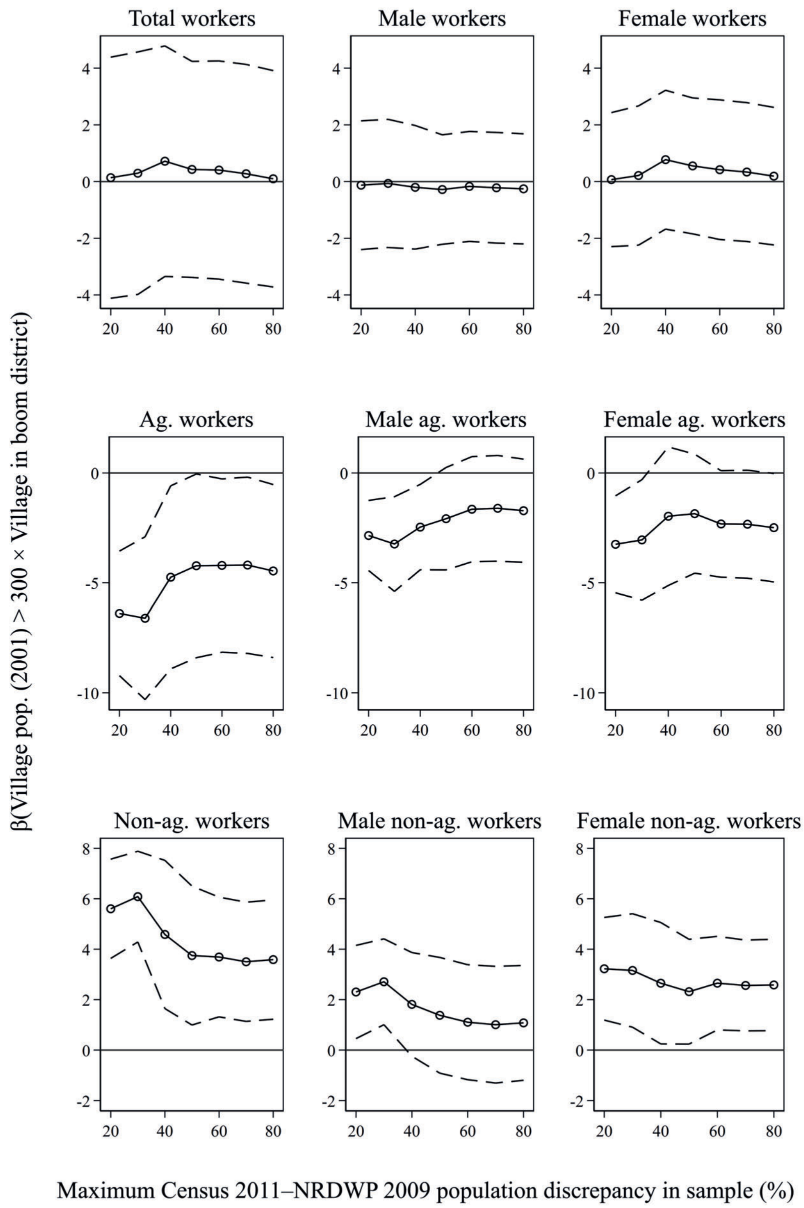
Notes. This figure shows results from estimating a modified version of the regression discontinuity specification outlined in Equation (1) that includes a fully-interacted term for the square of the population running variable. The outcome variables presented in this figure come from the Primary Census Abstract tables of the 2011 round of the Population Census. Best-fit curves are constructed using predicted values from the regressions. Each solid (hollow) dot represents the mean value of the relevant outcome variable for approximately 10 (500) villages in fifteen-person bins.

Figure D2: Evaluating differential impact of electrification in boom and non-boom districts in 2011 using randomization inference



Notes. Each panel of this figure plots the distribution of 1,000 estimated values of $\hat{\beta}_2$ from a randomization-based inferential procedure (Athey and Imbens, 2017). In each iteration, we randomly assign eleven RGGVY Phase I districts to a placebo boom group, and re-estimate the regression discontinuity specification outlined in Equation (1) to obtain a $\hat{\beta}_2$ placebo value for the degree to which the guar boom augments the impact of electrification. The dashed vertical line indicates the corresponding estimated value of $\hat{\beta}_2$ reported in Table 2. Dark (light) shading represents the 90 (95) percent confidence interval of each distribution.

Figure D3: Sensitivity of results to varying Census–NRDWP population discrepancy rates



Notes. This figure shows how the results reported in Table 1 (column 1) and Table 2 (columns 1 and 4) for the estimated value of $\hat{\beta}_2$ evolve as we relax the Census 2011–NRDWP 2009 population discrepancy threshold we impose during our fuzzy matching procedure to validate matches (see Appendix C). Markers represent point estimates from regressions; dashed lines indicate corresponding 90 percent confidence intervals.

Figure D4: Sensitivity of results to varying RD bandwidths

Notes. This figure shows how the results reported in Table 1 (column 1) and Table 2 (columns 1 and 4) for the estimated value of $\hat{\beta}_2$ evolve as we vary the population bandwidth around RGGVY's 300-person eligibility threshold to identify our analytical sample. Markers represent point estimates from regressions; dashed lines indicate corresponding 90 percent confidence intervals.

Appendix E Additional tables

Table E1: Designating guar-growing districts based on technical reports

State	(1)	(2) Technical reports	(3)	(4) Analytical sample
	APEDA	NRAA	USDA	Guar-growing district
Rajasthan		Alwar		
	Barmer	Barmer	Barmer	✓
			Bhilwara	
	Bikaner	Bikaner	Bikaner	✓
	Churu	Churu	Churu	✓
	Dausa	Dausa		✓
	Hanumangarh	Hanumangarh	Hanumangarh	✓
	Jaipur		Jaipur	✓
	Jaisalmer		Jaisalmer	✓
			Jalore	
	Jhunjhunu	Jhunjhunu	Jhunjhunu	✓
	Jodhpur	Jodhpur	Jodhpur	✓
	Nagaur	Nagaur	Nagaur	✓
			Pali	
	Sikar	Sikar	Sikar	✓
	Sirohi	Sirohi		✓
	Sri Ganganagar	Sri Ganganagar	Sri Ganganagar	✓
Haryana	Bhiwani	Bhiwani		✓
	Gurgaon	Gurgaon		✓
	Mahendragarh	Mahendragarh		✓
	Rewari	Rewari		✓
Gujarat	Ahmedabad	Ahmedabad		✓
	Banaskantha	Banaskantha		✓
	Kutch	Kutch		✓
	Mehsana	Mehsana		✓
	Sabarkantha	Sabarkantha		✓
	Vadodara	Vadodara		✓
Punjab		Bathinda		
		Firozpur		
		Mansa		
		Sri Muktsar Sahib		

Notes. Columns (1), (2) and (3) of this table list districts that are characterized as India's top guar producers in technical reports released by the Agricultural and Processed Food Products Export Development Authority (2011, p. 13), the National Rainfed Area Authority (2014, p. 3) and the US Department of Agriculture (Singh, 2014, p. 23), respectively. Column (4) indicates those that are mentioned by at least two of these technical reports, and which we thus designate as guar-growing districts for the purposes of our analyses (as described in Section 4.1).

Table E2: Testing for discontinuous changes at RGGVY Phase I threshold in 2001

Outcome variable (2001)	(1) $\mathbb{1}$(Village pop. (2001) > 300) Coef.	(2) Std. Err.	(3) N	(4) Adj. R^2	(5) Mean of outcome
Number of households	-0.08	(61.96)	7649	0.64	53.97
Females (% of population)	-0.01	(16.43)	7649	0.28	48.73
Ages 0–6 (% of population)	0.04	(35.86)	7649	0.36	17.78
Scheduled Caste/Tribe (% of population)	-0.57	(338.47)	7649	0.28	36.02
Literate (% of population)	-0.01	(6.49)	7649	0.36	45.01
All workers (% of population)	-1.00	(1.93)	7649	0.38	43.98
Agricultural workers (% of population)	-0.38	(228.06)	7649	0.32	37.22
Non-agricultural workers (% of population)	-0.62	(2.87)	7649	0.15	6.76
Area (Hectares)	-14.02	(101.07)	7649	0.36	158.07
Irrigated area (% of total area)	-0.65	(387.19)	7324	0.40	35.67
Primary schools (per 1,000 people)	-0.10	(0.40)	7649	0.27	1.97
Community health workers (per 1,000 people)	0.05	(0.22)	7649	0.10	0.20
$\mathbb{1}$(Bus facilities)	0.01	(4.77)	7649	0.22	0.17
$\mathbb{1}$(Postal facilities)	0.02	(0.13)	7649	0.15	0.18
$\mathbb{1}$(Approach: Paved road)	0.00	(4.37)	7649	0.10	0.37
$\mathbb{1}$(Power supply)	0.03	(0.08)	7649	0.35	0.66

Notes. Column (1) reports the value of $\hat{\beta}_1$ obtained from estimating the following regression specification on our main analytical sample of single-habitation villages located in RGGVY Phase I districts: $y^{2001}_{vds} = \beta_0 + \beta_1 T_{vds} + \beta_2 \tilde{P}^{2001}_{vds} + \beta_3 T_{vds} \tilde{P}^{2001}_{vds} + \gamma_d + \varepsilon_{vds}$, where y^{2001}_{vds} represents an outcome variable for village v in district d in state s in 2001, T_{vds} is a binary variable that equals one if the population of village v in 2001 is greater than 300, $\tilde{P}^{2001}_{vds}$ is the population running variable, and γ_d represents a district fixed-effect. Standard errors—in column (2)—are clustered at the district level and inferred from p-values obtained using the free step-down resampling methodology of Westfall and Young (1993). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E3: RD estimates of impact of electrification on total population in 2011

		(1)	(2)	(3)
		Total population (2011)		
		All	Male	Female
$\hat{\beta}_1$	$\mathbb{1}$ (Village pop. (2001) $>$ 300)	3.39**	2.22**	1.16
		(1.70)	(0.94)	(0.91)
$\hat{\beta}_2$	$\mathbb{1}$ (Village pop. (2001) $>$ 300) $\times$	2.27	6.53*	-4.45
	$\mathbb{1}$ (Village in boom district)	(6.29)	(3.77)	(2.91)
District FEs		Yes	Yes	Yes
Census (2001) controls		Yes	Yes	Yes
N		7649	7649	7649
Adjusted R^2		0.54	0.54	0.49
Mean of outcome		349.24	178.92	170.32

Notes. This table shows results from estimating Equation (1). Each regression includes all single-habitation villages in RGGVY Phase I districts with a 2001 population within a fifty-person bandwidth of RGGVY's 300-person eligibility threshold. Estimates associated with the population running variable $\left(\tilde{P}_{vds}^{2001}\right)$ are omitted. Following Correia (2015), six singleton observations are excluded. Standard errors—in parentheses—are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E4: Differences between villages in guar and non-guar-growing districts in 2001

Outcome variable (2001)	(1)	(2)	(3)	(4)	(5)	(6)
	All RGGVY Phase I villages			RD sample villages		
	Non-boom	Boom	p-value of difference	Non-boom	Boom	p-value of difference
Total population	1390.86	1502.61	0.056*	299.99	306.22	0.284
	(1654.30)	(1455.01)		(29.18)	(28.44)	
Number of households	247.03	231.91	0.142	54.07	48.87	0.385
	(316.25)	(225.51)		(11.68)	(8.78)	
Females (% of population)	48.62	48.26	0.940	48.74	47.89	0.484
	(2.91)	(2.40)		(3.04)	(2.85)	
Age 0–6 (% of population)	18.07	19.87	0.940	17.75	19.09	0.972
	(4.17)	(3.47)		(4.57)	(4.16)	
Scheduled Caste/Tribe (% of population)	31.65	27.24	0.599	36.33	20.77	0.285
	(27.65)	(22.97)		(34.78)	(26.04)	
Literate (% of population)	44.74	47.00	0.462	44.94	48.81	0.434
	(14.48)	(13.35)		(16.33)	(13.87)	
Total workers (% of population)	41.61	46.14	0.219	43.91	47.12	0.434
	(12.86)	(10.32)		(14.14)	(12.05)	
Agricultural workers (% of population)	33.79	37.82	0.847	37.17	39.94	0.742
	(14.10)	(13.32)		(15.27)	(13.96)	
Non-agricultural workers (% of population)	7.81	8.31	0.729	6.75	7.18	0.972
	(7.57)	(7.59)		(7.89)	(9.32)	
Area (Hectares)	358.69	1428.10	0.219	148.41	648.16	0.486
	(756.26)	(2316.45)		(224.87)	(1161.81)	
Irrigated area (% of total area)	38.36	21.09	0.940	35.97	21.21	0.972
	(33.84)	(25.04)		(33.69)	(27.24)	
Primary schools (per 1,000 people)	1.27	1.45	0.628	1.95	3.01	0.678
	(2.24)	(6.64)		(1.81)	(1.06)	
Community health workers (per 1,000 people)	0.15	0.10	0.940	0.20	0.11	0.910
	(1.10)	(0.60)		(0.83)	(0.59)	
$\mathbb{1}$(Bus facilities)	0.27	0.60	0.092*	0.17	0.32	0.393
	(0.44)	(0.49)		(0.37)	(0.47)	
$\mathbb{1}$(Postal facilities)	0.40	0.64	0.219	0.18	0.36	0.678
	(0.49)	(0.48)		(0.38)	(0.48)	
$\mathbb{1}$(Approach: Paved road)	0.53	0.60	0.219	0.37	0.34	0.434
	(0.50)	(0.49)		(0.48)	(0.48)	
$\mathbb{1}$(Power supply)	0.73	0.89	0.940	0.65	0.84	0.972
	(0.45)	(0.31)		(0.48)	(0.36)	
N	182051	6232		7507	148	

Notes. This table reports mean and standard deviations (in parentheses) for villages located in boom and non-boom districts of India. Columns (1) and (2) report these values for our full sample of habitation-matched villages in RGGVY Phase I districts; column (4) and (5) report these values for our main analytical sample of single-habitation villages. Columns (3) and (6) report the p-value for $\hat{\beta}_1$ obtained from estimating the following regression specification on the relevant sample: $y_{vds}^{2001} = \beta_0 + \beta_1 G_{ds} + \gamma_s + \varepsilon_{vds}$, where y_{vds}^{2001} represents an outcome variable for village v in district d in state s in 2001, G_{ds} is a binary variable that equals one if village v is located in a guar-growing district, and γ_s represents a state fixed-effect. Standard errors (not shown) are clustered at the district level; p-values are obtained using the free step-down resampling methodology of Westfall and Young (1993). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E5: Placebo RD estimates of short-term impact of electrification on labor-market outcomes

		(1) All workers	(2) Ag. workers	(3) Non-ag. workers
		(% of 2011 population)		
$\hat{\beta}_1$	$\mathbb{1}$ (Village pop. (2001) $>$ 300)	0.29	−0.18	0.55
		(0.56)	(0.71)	(0.43)
$\hat{\beta}_2$	$\mathbb{1}$ (Village pop. (2001) $>$ 300) $\times$ $\mathbb{1}$ (Village in boom district)	−0.63	1.74	−2.39
		(1.51)	(2.21)	(1.80)
District FEs		Yes	Yes	Yes
Census (2001) controls		Yes	Yes	Yes
N		6992	6992	6992
Adjusted R^2		0.38	0.45	0.32
Mean of outcome		48.23	39.94	8.28

Notes. This table shows results from estimating Equation (1) on a sample of single-habitation villages located in RGGVY Phase II districts (where large-scale electrification did not occur during the period covered by our data) with a Census 2001 population within a fifty-person bandwidth around RGGVY's 300-person eligibility threshold. Outcome variables for regressions reported in columns (1)–(3) are constructed using data from the Primary Census Abstract tables of the 2011 round of the Indian Census. Specifically, "agricultural workers" represents a village-level sum of main and marginal cultivators and agricultural laborers, while "non-agricultural workers" represents a village-level sum of main and marginal household-industry and "other" workers. Estimates associated with the population running variable $\left(\tilde{P}_{vds}^{2001}\right)$ are omitted. Following Correia (2015), 21 singleton observations are excluded. Standard errors—in parentheses—are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table E6: RD estimates with multiple hypothesis test adjustment

Outcome variable	(1) $\hat{\beta}_2$	(2) Adj. p-value
All workers (% of population)	0.14	0.997
Male	-0.13	0.996
Female	0.07	0.997
Agricultural workers (% of population)	-6.39*	0.095
Male	-2.85	0.203
Female	-3.25	0.265
Non-agricultural workers (% of population)	5.60**	0.043
Male	2.30	0.296
Female	3.22	0.265

Notes. Column (1) reports the estimated $\hat{\beta}_2$ coefficients from Tables 1 and 2. Column (2) reports corresponding p-values for this "family" of regressions, adjusted for multiple hypothesis testing using the free step-down resampling methodology of Westfall and Young (1993). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix references

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, jun 2010, *105* (490), 493–505.

_ **and Javier Gardeazabal**, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, feb 2003, *93* (1), 113–132.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton, "The value of connections in turbulent times: Evidence from the United States," *Journal of Financial Economics*, aug 2016, *121* (2), 368–391.

Aggarwal, Shilpa, "Do rural roads create pathways out of poverty? Evidence from India," *Journal of Development Economics*, jul 2018, *133*, 375–395.

Agricultural and Processed Food Products Export Development Authority, "APEDA Agri Exchange Ready Reckoner Series: Guargum," Technical Report, Ministry of Commerce & Industry 2011. Retrieved from `http://apeda.gov.in/apedawebsite/six_head_product/Guargum_final_Profile.pdf` on April 16, 2017.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, March 2020, *110* (3), 797–823.

Burlig, Fiona and Louis Preonas, "Out of the Darkness and Into the Light? Development Effects of Rural Electrification in India," 2016. Energy Institute at Haas Working Paper № WP-268R.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano, "Catastrophic Natural Disasters and Economic Growth," *Review of Economics and Statistics*, dec 2013, *95* (5), 1549–1561.

Chen, X. and W. D. Nordhaus, "Using luminosity data as a proxy for economic statistics," *Proceedings of the National Academy of Sciences*, may 2011, *108* (21), 8589–8594.

Coffman, Makena and Ilan Noy, "Hurricane Iniki: measuring the long-term economic impact of a natural disaster using synthetic control," *Environment and Development Economics*, nov 2011, *17* (02), 187–205.

Correia, Sergio, "Singletons, cluster-robust standard errors and fixed effects: A bad mix," 2015. Technical note.

Doll, Christopher N.H., Jan-Peter Muller, and Jeremy G. Morley, "Mapping regional economic activity from night-time light satellite imagery," *Ecological Economics*, apr 2006, *57* (1), 75–92.

Gaba, Kwawu Mensan, Brian Min, Anan Thakker, and Chris Elvidge, "`nightlights.io`: Twenty Years of India Lights," 2016. Accessed on September 9, 2016.

Gronau, Reuben, "Leisure, Home Production, and Work–the Theory of the Allocation of Time Revisited," *Journal of Political Economy*, 1977, *85* (6), 1099–1123.

Henderson, J. Vernon, Adam Storeygard, and David N Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, apr 2012, *102* (2), 994–1028.

Kaczan, David J., "Can roads contribute to forest transitions?," *World Development*, May 2020, *129*, 104898.

Min, Brian, Zachary O'Keeffe, and Fan Zhang, "Whose Power Gets Cut? Using High-Frequency Satellite Images to Measure Power Supply Irregularity," 2017. World Bank Policy Research Working Paper № 8131.

Ministry of Power, "Rajiv Gandhi Grameen Vidyutikaran Yojana," 2012. http://pib.nic.in/newsite/PrintRelease.aspx?relid=83765.

National Rainfed Area Authority, "Potential of Rainfed Guar (Cluster beans) Cultivation, Processing and Export in India," Technical Report, Ministry of Agriculture & Farmers Welfare 2014. Retrieved from http://nraa.gov.in/pdf/Rainfed-guar-final-pdf.pdf on April 16, 2017.

Pinkovskiy, Maxim and Xavier Sala-i-Martin, "Lights, Camera … Income! Illuminating the National Accounts-Household Surveys Debate," *The Quarterly Journal of Economics*, feb 2016, *131* (2), 579–631.

Singh, Santosh K., "An Analysis of Guar Crop in India," Technical Report, United States Department of Agriculture 2014.

Singhal, Saurabh and Rahul Nilakantan, "The economic effects of a counterinsurgency policy in India: A synthetic control analysis," *European Journal of Political Economy*, dec 2016, *45*, 1–17.

Westfall, Peter H. and S. Stanley Young, *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, Wiley-Interscience, 1993.