The Price of Power: Costs of Political Corruption in Indian Electricity

Meera Mahadevan*

October 3, 2023

Abstract

Politicians may target public goods to benefit their constituents, at the expense of others. I study corruption in the context of Indian electricity and estimate the welfare consequences. Using new administrative billing data and close-election regression discontinuities, I show that *billed* electricity consumption is lower for constituencies of the winning party by almost 40%, while *actual* consumption, measured by nighttime lights, is higher. I document the covert way in which politicians subsidize constituents by manipulating bills. These actions have substantial welfare implications, with an efficiency loss of over $0.6 billion, leading to unreliable electricity supply and significant negative consequences for development.

JEL: O13, P16, H11, Q4, D73

Keywords: Electricity, political economy, corruption, India

*University of California, Irvine, meera.m@uci.edu, meeramahadevan.com. I am grateful to Ryan Kellogg, Dean Yang, Hoyt Bleakley, and Catherine Hausman for their guidance, and Achyuta Adhvaryu, Sam Asher, Prashant Bharadwaj, Fiona Burlig, Melissa Dell, Ray Fisman, Kelsey Jack, Gaurav Khanna, Karthik Muralidharan, Paul Novosad, Subhrendu Pattanayak, Francesco Trebbi, Faraz Usmani, Shaoda Wang, Jeffrey Weaver, Catherine Wolfram and Pinghui Wu for helpful comments. Thanks to seminar participants at the NBER EEE Spring Meeting 2021 (EEE), NBER Political Economy Fall Meeting 2019 (POL), Berkeley/VSE PE Conference, University of Michigan, University of Southern California, Vancouver School of Economics (UBC), Rutgers, University of Melbourne, University of California – Irvine, University of Virginia, University of Western Ontario, American University, National University of Singapore, University of Melbourne, and UCSD, and participants at Development Day (U Chicago), PACDEV, AERE, NEUDC, WCERE 20118, CU Environmental & Resource Economics Workshop 2018, Canadian Resource and Environmental Economics Workshop 2018, SETI Workshop and Occasional Workshop (UC Santa Barbara) for useful comments. I thank the Michigan Institute for Teaching and Research in Economics (MITRE) for research funding.

1 Introduction

A classic concern in political economy is the extent to which legislators or bureaucrats favor the interests of some groups over others for political gain (Finan and Schechter, 2012). A ruling party may provide their constituents preferential access to public goods after winning an election to deliver on campaign promises (Cruz et al., 2020), or instead target new voters in constituencies where they lost elections (Callen et al., 2020). Obtaining causal evidence of the mechanisms of patronage at a sufficiently large scale remains challenging (Muralidharan et al., 2016). Yet, identifying the mechanisms behind such practices, and quantifying the welfare costs is first order to designing effective policies and reducing inequities.

This work develops original forensic tools and uses new administrative data to examine the problem of political targeting through the lens of the Indian electricity sector. In many countries, public goods such as electricity and water utilities are state-owned and, therefore, vulnerable to political manipulation. Public utilities offer a steady flow of resources that may be directed by politicians even beyond their initial construction. There are numerous avenues to exploit the heavily bureaucratic and opaque processes behind investment and supply decisions to provide a continued stream of favorable access to preferred voter groups. While there exists some evidence on the costs of misallocation caused by patronage in general (Khwaja and Mian, 2005), the impact on tax-payer-funded institutions themselves and the broader welfare implications are harder to isolate in most contexts. For instance, little is known about how much of the large commercial losses faced by public electricity providers and other state-run entities is attributable to political manipulation.[1] To the extent that such manipulation further hampers utilities' ability to provide reliable electricity, this has ramifications that go well beyond the electricity sector: the social costs of intermittent electricity on economic development (Dinkelman, 2011; Greenstone and Jack, 2015; Lipscomb et al., 2013) and productivity (Allcott et al., 2016; Fried and Lagakos, 2023), and the large opportunity costs of systematically bailing out loss-making electricity utilities (Chatterjee, 2017). Despite the scale of these concerns, there is little evidence of well-identified work describing political manipulation in large utilities (Min and Golden, 2014) and even less on

[1] \$16 billion of tax-payer funds were used to bail out loss-making Indian electricity utilities. Losses amount to more than \$41 billion annually across developing countries (Gulati and Rao, 2007).

its ensuing welfare consequences.

This paper presents causal evidence on how Indian politicians may manipulate public electricity provision to favor a subset of voters, and the large costs this imposes on electricity providers and the economy. I obtain new administrative billing records from the electricity utility of West Bengal, a large Indian state, to measure *reported* consumption. One innovation is to treat this administrative data as distinct from *actual* consumption, which this analysis measures using satellite nighttime luminosity data.[2] I argue that the difference between administrative data and actual usage serves as a proxy for potential manipulation, and this allows me to estimate its welfare implications in a way that is difficult to do in most other instances of corruption.

The paper presents three key results. In the first set of results, the paper provides causal evidence that politicians from the ruling party at the state level favor their constituents by providing them with illicit electricity subsidies after winning an election. I leverage a close-election Regression Discontinuity Design (RDD), a strategy commonly used in political economy research in India and elsewhere (George and Ponattu, 2020; Nellis et al., 2016; Prakash et al., 2019). We infer the existence of illicit subsidies based on two complementary pieces of evidence. First, shortly after a state-level election, there is an increase in *actual* electricity consumption, as measured by satellite nighttime lights data, for regions just aligned with the ruling party.[3] Second, these same regions in West Bengal have discontinuously lower levels of *billed* consumption, as reported by the electricity provider. The magnitude of under-reporting is large, with favored account holders paying for only 60% of their billable consumption. Politicians appear to favor their constituencies by under-reporting electricity consumption, even as their constituents consume higher actual amounts of electricity.

The second set of results uncovers the mechanisms by which bill manipulation takes place, a key feature to understanding how politicians may conceal data manipulation. First, we observe that a discontinuously higher number of bills in the ruling party's constituencies are multiples of ten, reporting consumption amounts such as 20, 30, 40 KWh, and so on – saliently visible in the underlying data. Given that each electoral district consists of 3-4 billing centers answering to the elected representative from the ruling party, these patterns

[2]Satellite nighttime luminosity data has been found to be a good predictor of daytime electricity use, particularly in India (Mann et al., 2016).

[3]Throughout the paper, these results examine the role of a newly installed ruling party at the state level in the years before and following the election where they win. The results first focus on West Bengal, particularly post the 2011 elections, but show that these results extend to the rest of the country for elections from 2006-2016.

point towards a top-down approach to manipulating reported consumption in the billing data. Second, to further corroborate these data anomalies, the paper uses Benford's (1938) Law to show that there is a greater divergence between the observed consumption distribution and the theoretically expected one in constituencies represented by the ruling party.[4] These results are consistent with local, incumbent politicians rewarding their constituents by permitting the manipulation of billed consumption to appear lower than actual consumption, a mechanism made possible by the close relationships elected officials have with local billing centers (Chhibber et al., 2004). These findings explain the observed discrepancies between reported and actual electricity consumption, allowing me to identify the affected parties and assess the impact.

Finally, the paper discusses the welfare implications of billing manipulation by politicians. The combination of administrative data and satellite data in a context where corruption is measurable is instrumental in estimating the welfare consequences. We directly observe under-reporting in billing data, and overconsumption of electricity through satellite nighttime lights data from the RD analysis, and use these estimates to compute the size of the welfare numbers. Welfare depends on the loss in producer profits from not recovering sufficient revenue, and gains in surplus to a subset of benefiting consumers. The difference between the two provides a sense of the efficiency loss, if any. The paper estimates a loss to the electricity utility of over \$1.8 billion, while the favored set of consumers gain a sizeable \$1.2 billion, creating a net efficiency loss of \$0.6 billion in West Bengal alone. Further, these figures reflect losses for just a single state, when in fact, electricity utilities in 25 other Indian states share similar vulnerabilities (Gulati and Rao, 2007). Indeed, the paper finds the same patterns of over-consumption by connected constituents using nighttime lights data in other states, and other elections in India.

At the broadest level, the paper contributes to a vast literature that aims to identify political patronage and corruption, and demonstrates evidence at a large scale: first, for a state with a population of 72 million, and then documenting similar patterns for the rest of the country. Other work has demonstrated the extent to which politicians have incentives to favor constituents, motivated by expected rewards in subsequent election cycles (Fujiwara et al., 2020; George et al., 2018; Zimmermann, 2021). However, the resultant welfare implications are ambiguous as reelection incentives might lead to the efficient allocation of

[4] Benford's (1938) Law predicts a frequency distribution of the first digit of naturally occurring, unmanipulated sets of numerical data, such as consumption data, and is commonly used to detect data fraud in survey data collection.

government inputs (Pande, 2003, 2020) rather than, as my results suggest, to misallocation and efficiency losses. Given this ambiguity, documenting the welfare costs in practice is important to design policies to limit manipulation. My paper joins a handful of studies that documents the existence of welfare costs resulting from patronage practices (Khwaja and Mian, 2005), and one of the only ones to my knowledge that considers how large public institutions may be affected.

Measuring the scale of welfare costs from corruption is challenging and given the limited evidence on it (Hicken, 2011), an important contribution of this paper. The few studies that examine the implications of patronage have often focused on preferential misallocation (Khwaja and Mian, 2005). This paper advances the literature by considering an efficiency loss that goes beyond transfers. The unique combination of administrative and satellite data to distinguish between measured and true consumption is crucial in estimating welfare. Alone, the satellite data may indicate selectively higher levels of electricity access or consumption for politically connected regions. On the other hand, the billing data alone suggests instead that politicians redirect electricity to regions where they lost elections. However, taken together, the evidence from both datasets paints a different picture: that politicians may be under-reporting electricity consumption for their constituents, and these consumers respond by over-consuming. These actions lead to a deadweight loss large enough to power almost 91 million additional rural households across the country. While beneficiaries may be gaining from illicit subsidies in the short run, they may bear the consequences of corruption in the long run through frequent outages due to the utility's limited ability to supply reliable electricity on insufficient revenue (Burgess et al., 2020; Mahadevan, 2022). The true efficiency losses are likely greater if one considers the opportunity cost of electricity utility bailouts (Chatterjee, 2017), and the utility's consequent inability to meet electricity needs, affecting economic productivity more broadly (Fried and Lagakos, 2023).

The evidence on manipulation of administrative data for political ends contributes to a large literature in public finance, where discussions around manipulation have often focused on inadvertent measurement error, incentives related to data gathering, misreporting by consumers (Slemrod, 2016) or eligibility manipulation (Camacho and Conover, 2011). However, the role of political incentives to manipulate the measurement of consumption data itself is less studied. The political machine that enabled data manipulation in this context possibly extends to other kinds of administrative data (Jeong et al., 2020), having implications for development policies that rely on those data. While we may be able to observe the effects of patronage on economic growth or policy targeting from a well-identified setting (Asher and Novosad, 2017), I show that more covert forms of patronage may be difficult to detect without comparing external or satellite data with on-the-ground administrative data. Having access to both micro-administrative and satellite data allows me to detect manipulation of the billing data and quantify the costs: Indeed, regular audits of the electricity billing process failed to uncover this mode of corruption (Gulati and Rao, 2007).

The layout of the rest of this paper is as follows: Section 2 presents a conceptual framework and institutional details for the Indian electricity context. Section 4 covers the empirical strategy and Section 3, describes the data used. Section 5 presents evidence of corruption from administrative and satellite data. Section 6 discusses how the results extend to other contexts, and Section 7 discusses the welfare implications. Section 8 concludes.

2 Background and Conceptual Framework

Theoretically, the idea that politicians favor voters, particularly those in highly contested zones, is reflected in models developed by Stromberg (2004) and Dixit and Londregan (1996). On the one hand, they may redirect additional resources to closely contested regions where they lost elections (Callen et al., 2020). On the other hand, they may prefer to reward aligned voters to continue a cycle of electoral victories (Cruz et al., 2020; Mahadevan and Shenoy, 2023). While targeting regions where they won may also be a result of lower costs (i.e., it is easier to influence local officials if the local MLA and state government are aligned), there is ambiguity over whether it leads to higher returns as well. That is, who received credit or blame is an important factor driving the dynamics at play in a democracy like India.

In India, punishing or rewarding representatives for their policies is somewhat common, independent of their party affiliation (Khanna and Mukherjee, 2023; Zimmermann, 2021). In fact, recent work shows local representatives (the MLA or village council leaders) are the primary targets of credit or blame given the frontline role they play in the political machine (Mahadevan and Shenoy, 2023; Shenoy and Zimmermann, 2022). The level of involvement local politicians have may make their party affiliation a more relevant dimension to which voters respond rather than the ruling party. For instance, in South Asia, there is evidence that local leaders take credit for programs that were not even implemented by them, due to the immediate electoral payoff of these claims (Guiteras and Mobarak, 2014). In other parts of Asia, a centrally implemented CCT (with the help of international organizations) increased local incumbent vote shares (Labonne, 2013). In fact, when such cash transfers occur, voters may actually be more likely to reward the local parties (when aligned to the ruling coalition) rather than rewarding the central government parties (Pop-Eleches and Pop-Eleches, 2012). Indeed, some research argues that voters may not be fully informed, as even quasi-random program assignment leads beneficiaries to favor the implementing incumbent (Manacorda et al., 2011). However, affiliation with the ruling party may crucially enhance the ability of local politicians to practice selective targeting (Asher and Novosad, 2017), while hurting the ability of elected representatives from other parties to perform the same actions.

There are a number of reasons why politicians may want to control the electricity supply. Election surveys in India find that electricity is a key factor in election platforms (Chhibber et al., 2004). However, using a large public sector like electricity provision as a tool of patronage may not be as straightforward as misallocating public funding or new infrastructure. Given the formal separation of the electricity sector from political control (Under the 2003 Electricity Act), there are a number of trade-offs politicians may face when attempting to exploit the power sector for electoral gains. First, there may be direct costs to political favoritism in the form of dealing with local civil servants who would rather not engage in corruption, and, when necessary, bribing or threatening local civil servants. Second, it may not be easy to manipulate electricity pricing or supply given the management of utilities by unaffiliated civil servants and their overseeing by independent regulators, offering no direct access to politicians. Third, however, there are a number of weaknesses in the electricity sector that politicians may be able to exploit.

In practice, electricity providers remain state-owned and managed by independent regulators, separating politicians from accountability for electricity sector performance. However, in several states, electricity distributors have faced mounting losses for several years, but an individual state is rarely singled out (Chatterjee, 2017, 2018). Loss-making utilities that draw minimal attention, therefore, offer a potential avenue for politicians to exploit, despite the seeming lack of access. Indeed, (Baskaran et al., 2015) show evidence of electoral cycles in power blackouts in India. Chatterjee (2018) presents evidence consistent with politicians pressuring regulatory officials to avoid upward revisions in tariffs, which regulators report resisting. Other methods politicians may resort to include implicitly allowing energy theft among their constituents (The Telegraph, 2014; The Times of India, 2018; The Washington Post, 2012).[5] Golden and Min (2011) demonstrate how electricity bills are more likely to go unpaid in areas where criminals have political affiliations.

[5] "A [local politician] .... has said that discom officials who penalise farmers for power theft or overloading should be tied to trees", (The Times of India, 2018).

While politicians are not responsible for their state utility making large financial losses, there exist limits to the level of losses a utility can make before drawing unwanted attention. Excessive losses may attract central audits or unwanted media attention. Therefore while politicians may be able to exploit the perpetually loss-making utilities in some ways, there may be an implicit constraint imposed to the extent of patronage by the "equilibrium" level of acceptable financial losses set by historical precedent and compared to other states. As a result, ruling parties may have an incentive to exploit the electricity sector to reward their constituents, or allow their affiliated local representatives to do so. But they may also have the incentive to limit how much members of opposition parties can do given the zero-sum nature of exploiting the electricity sector under an implicit budget constraint.

Electricity supply is a critical election issue in India (Chatterjee, 2018), where 55% of surveyed firms experienced electrical outages and more than half the firms reported being required to provide a 'gift' in exchange for an electricity connection (The World Bank, 2014). A third of the Indian population does not have access to electricity, and even those who do often experience long and frequent blackouts (Pargal and Banerjee, 2014). Poor electricity supply is a major constraint to manufacturing (Allcott et al., 2016).

This paper first presents evidence on West Bengal, a large Indian state where the transmission and distribution sectors are state-owned. The vast majority of the consumers in the state (and most residential and commercial establishments) are supplied by the state-owned West Bengal State Electricity Distribution Company Limited (WBSEDCL), covering a population of about 72 million individuals, through 17 million accounts.[6] In 2003, a central reform created a state regulatory commission, responsible for setting electricity tariffs and overseeing the functioning of the utility, specifically to separate the control of the electricity sector from political influence. This institutional setup is ubiquitous across states in India, and similar to other countries (e.g. Brazil, Bangladesh, Mexico, Sri Lanka, and Kenya), where electricity is a heavily subsidized commodity for households and small commercial establishments, with most state electricity utilities unable to recover their costs.

3 Data Description and Variable Definitions

Table A1 in Appendix A.1, presents summary statistics for the main variables of interest by whether or not the constituency was aligned with the ruling party. The table presents the

[6] With the exception of one privately owned firm which distributes only to the capital.

summary measures for each of the samples used in the data, varied by the bandwidth applied in each RD. In the analysis using billing data, I focus on the 2011 election. All analyses using billing and consumption data analyze political behavior post elections.

3.1 Administrative data on Electricity Consumption and Billing

This paper uses administrative data on the universe of electricity consumption and billing records from the West Bengal State Electricity Distribution Corporation Limited (WBSEDCL). These data include consumption for residential and commercial users in both rural and urban areas between 2011 and 2016. For most consumers, billing is done quarterly, with the exception of a few monthly users with commercial accounts. For the analysis in this paper, I restrict the analysis to a balanced panel of consumer IDs to ensure that I do not count any new accounts that started after 2012, and avoid issues of entry/exit. From the balanced panel of customers, I sample 2% of customer IDs, stratifying by each consumer category.

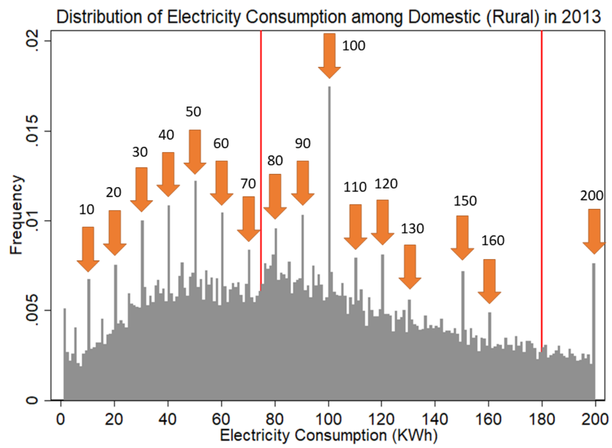
In the consumption dataset, each account is linked to a consumer care center (CCC). These centers are the local administrative offices for WBSEDCL, in charge of billing. I geolocate each of the 510 CCCs and situate them within their respective legislative assemblies, resulting in 2-3 CCCs per assembly area. Through their CCCs, therefore, all account holders under WBSEDCL are assigned to a particular legislative assembly.

3.2 Measures of Data Manipulation

The consumption distribution for residential and commercial consumers in Figure 1 is multimodal, with bunching at specific points. The peaks in the data appear at round numbers such as 20, 30, or 40 KWh. Electricity meters are read before every billing cycle by meter readers employed by the electricity utility. While it is common for meter inspectors not to conduct readings every billing cycle and make imputations for interim periods, the spikes observed are large.

Based on the multi-modal consumption distribution, I define two measures to characterize the manipulation of the underlying data. The first is based on Benford's (1938) Law, which lays out an expected distribution for the first digit of a naturally occurring set of numbers. I measure the normalized distance of the distribution of the first digit of consumption for

Figure 1: Consumption Distribution for Residential Consumers



Notes: The consumption distribution above is for residential consumers in rural areas. The range of consumption extends from 1 KWh to more than 1000 KWh, but the bulk of distribution lies below 200 KWh (restricted to under this level in this graph), and largely has the shape of a chi-squared distribution. The two red lines represent the consumption levels at which the marginal price of electricity goes up. There are several clear spikes in the distribution, particularly at multiples of ten.

each assembly-year from the expected distribution. This metric, which is the same as the chi-squared goodness-of-fit statistic, represents the degree of manipulation in the underlying data. The second measure I use is the fraction of consumers in an assembly, in any given year, who have a reported consumption that is a multiple of ten. Because the consumption data would be, in expectation, smoothly distributed, a multiple of ten should not occur discontinuously more just above the RD cutoff. These measures enable me to test whether there is selective manipulation of administrative data in assemblies closely aligned with the ruling party. If bills are manipulated to reflect lower than actual consumption, that would amount to an indirect subsidy to constituents.

3.3 Satellite Nighttime Luminosity Data

I use satellite nighttime lights data as a non-manipulable measure of electricity consumption, serving as a barometer for the reported consumption from administrative data. I first validate this choice by checking if satellite nighttime lights are a good predictor of billed electricity consumption. Given the novel billing data, I can plot the relationship between selected billed consumption and luminosity data in Figure A5.[7] This figure shows a strong, linear relationship between log lights and billed consumption validating the use of luminosity data as a measure of electricity consumption, and also the use of the log functional form.[8] Luminosity data is also used to represent electricity consumption in other work: e.g. Mann et al. (2016) apply machine learning techniques to predict daytime electrification, and show nighttime luminosity to be a good indicator of electricity consumption. This builds on previous work where luminosity data is used as a proxy for electrification: Baskaran et al. (2015); Burlig and Preonas (2023); Min and Gaba (2014); Min et al. (2013); Min and Golden (2014).

I use the Visible and Infrared Imaging Suite (VIIRS) Day Night Band (DNB) satellite data from 2012 onward. I use a highly processed version that removes much of the noise, including cloud cover, ambient light, ephemeral lights, and background (non-lights), and excludes any data impacted by stray light (Elvidge et al., 2013, 2017). VIIRS satellite data especially improve on low light imaging compared to older satellite data such as the DMSP-OLS (discussed below), making across-region comparisons far more accurate. I measure the average density of lights within each legislative assembly, which is a continuous measure. In the absence of manipulation, the utility's consumption data may be expected to mirror patterns observed with the lights data.

I also use luminosity data from the United States Defense Meteorological Satellite Program (DMSP) for earlier years, which collects images of the earth twice a day and makes available annual composite images by averaging these daily data. They use 30-arc second grids, spanning -180 to 180 degrees longitude and -65 to 75 degrees latitude, and present the data using a 63-point luminosity scale. However, I use assembly-level average values of this luminosity, which yields a continuous measure of the variable.

4 Close-election Regression Discontinuity Design

This paper uses a close-election Regression Discontinuity (RD) design to identify whether politicians indirectly subsidize electricity. In India, parliamentary-style state elections occur

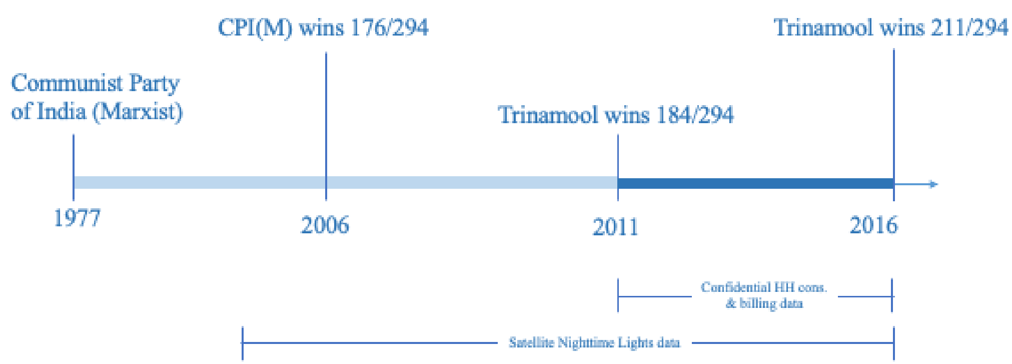
[7]I use assemblies where the consumption data passes the Benford's Law test to ensure I am only using data with a low probability of having been manipulated in any way.

[8]I follow the literature in using log as my main functional form: this seems the standard practice in seminal papers across the literature, whether it is used either at the national level (Henderson et al., 2012, 2011), at the sub-national level (Alesina et al., 2016; Hodler and Raschky, 2014; Michalopoulos and Papaioannou, 2013a,b; Storeygard, 2016), or in papers that discuss whether and how to use VIIRS (Gibson et al., 2021).

every five years. States are composed of legislative assembly constituencies (in short, assemblies). The voting population elects constituency-level representatives or Members of the Legislative Assembly (MLAs), and the political party with the majority of MLAs forms the government or the ruling party. This paper uses data on Indian elections from Asher et al. (2021) and Jensenius and Verniers (2017) from 2006-2016.

I compare outcomes just above and below a normalized winning vote margin RD cutoff to estimate the Local Average Treatment Effect (LATE) of being in a constituency aligned with a ruling party, after an election. The winning margin percentage is the fraction of votes by which an MLA from the ruling party wins an assembly election and is used as a running variable in other studies (Asher and Novosad, 2017; Bardhan and Mookherjee, 2010; Nagavarapu and Sekhri, 2014). Constituency-level elections in India are competitive and unpredictable, and several factors affect their outcomes. Given that the probability a constituency near the RD cutoff aligns with the ruling party is close-to-randomly determined, the close election RD may be especially valid in this case (Eggers et al., 2015).

Figure 2: A Timeline of State Elections in West Bengal from 1977-2016



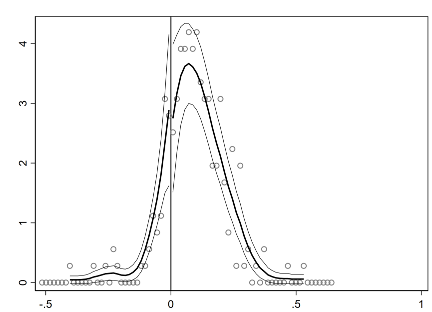
Notes: This figure shows multiple election years for the West Bengal State Legislative Assemblies. Between 1977-2006, the Communist Party of India (Marxist) won an absolute majority in the legislative elections. In 2011, the All India Trinamool Congress won the absolute majority of seats.

In the 2011 state elections, the All India Trinamool Congress (AITC) defeated the incumbent Communist Party of India – Marxist (CPI(M)) (Figure 2) and won an absolute majority. I use state assembly election data from 2006 to 2017, covering West Bengal elections in 2006, 2011, and 2016, and discuss my data in greater detail in the next section. In Appendix Section A, I further discuss details of the West Bengal elections. I focus on the 2011 election for results using administrative data. I use all election years between 2006-2016 to examine patterns in the nighttime lights data across other states in India, and highlight the external validity of the main results.

An important issue when using the RD is the selection of a smoothing parameter (Calonico et al., 2015; Imbens and Kalyanaraman, 2012; Imbens and Lemieux, 2008). In particular, I estimate local linear regressions with a rectangular kernel and employ the optimal data-driven procedure and bandwidth selection in Calonico et al. (2015). I present my results for a wide range of bandwidths to highlight the robust nature of my estimates, varying them from well below the optimal bandwidths to larger bandwidths. Varying the size of the bandwidth and the polynomial order does not affect the results presented in my analysis.

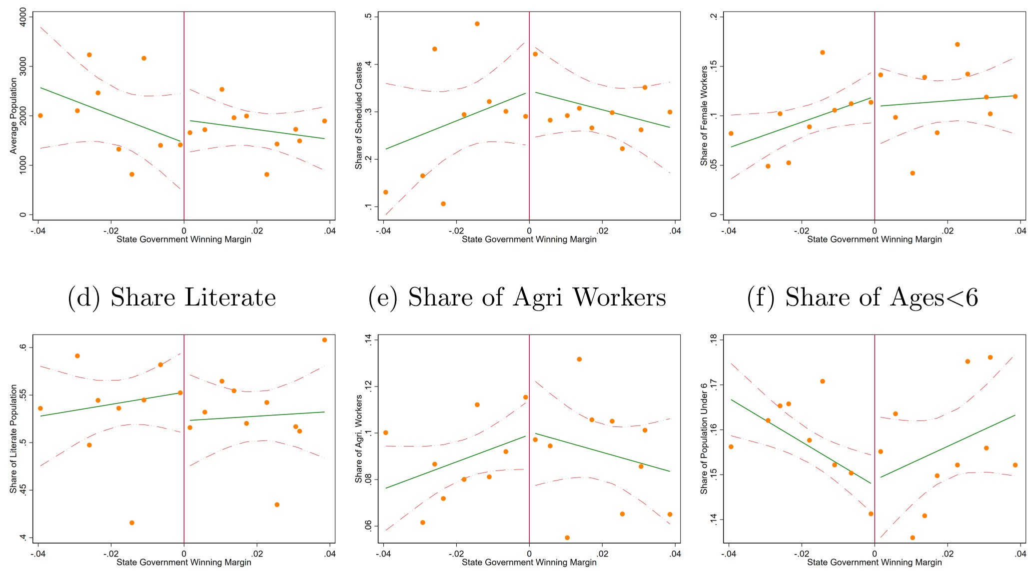
Two checks for balance of the running variable, the winning vote margin, and other demographic characteristics of the assemblies on either side of the cutoff validate the use of the RD. The McCrary (2008) test finds no significant discontinuities in the density of the running variable across the cutoff (Figure 3). Figure 4 further checks for balance across a range of village-level characteristics from the 2011 Indian Census and finds no significant discontinuities in demographic characteristics such as the proportion of certain castes, females, literate people, agricultural workers, and children. In Appendix Section A.2, Figure A1 shows the McCrary test limited to the bandwidth and finds no discontinuities.

Figure 3: Balance across RD cutoff – McCrary test



Notes: I test the smoothness of the running variable density (winning margin in the 2011 state election) and find no discontinuities across the RD cutoff using the McCrary Test. I do not find any discontinuities in the running variable even after zeroing in on the optimal bandwidth, as shown in Figure A1.

Figure 4: Balance on demographic features across RD cutoff



(a) Average Population (b) Share of Scheduled Castes (c) Share of Female Workers

(d) Share Literate (e) Share of Agri Workers (f) Share of Ages<6

Notes: I show balance in terms of village characteristics from the Indian Census 2011 across the RD cutoff for the state of West Bengal.

5 Empirical Evidence of Political Patronage

I leverage the close-election RD to test whether the party in power illicitly provided differentially cheaper electricity access to its voters by comparing electricity provision across the RD cutoff, using both administrative (reported consumption) and satellite data (actual consumption). I also explore the mechanisms behind potential corruption by examining patterns in the within-region distributions of electricity consumption. Throughout the paper, I refer to the party that wins an election as 'ruling party' at the state level in the years before and following the election. I refer to the assemblies where they win or lose as 'winning' or 'losing' assemblies respectively. I also refer to assemblies 'aligned' with the ruling party to be synonymous with the assemblies where the ruling party wins the election.

I use a bandwidth of 4.169 percentage points as the winning margin percentage, corresponding to about 6650-7109 votes. This is the optimal bandwidth under the Calonico et al. (2015) method using the billed consumption from the administrative data. I maintain this bandwidth across outcomes for consistency and ease of interpretation, but show robustness to a range of bandwidths from 1.6 pp up to 7 pp (I show the effective sample sizes and summary statistics in Table A1). To contextualize this in the literature examining close elections in India, these vote margins are lower or comparable to those used in Asher and Novosad (2017) (3-20 percentage points), Brown et al. (2021) (5 percentage points), Prakash et al. (2019) (6.16-7.79 percentage points), Bhalotra et al. (2018) (16-21 percentage points), Lehne et al. (2018) (3-6.2 percentage points) and Clots-Figueras (2012) (6-9 percentage points) who all examine close elections in India.

5.1 Average Nighttime Lights Density

I estimate the following specification at assembly level a, where the vote margin is the net difference in the fraction of votes received by the winning party over the party with the second-highest votes:

$$Log(Lights)_a = \beta\ \mathbb{1}(vote\ margin > 0)_a + f(vote\ margin)_a + \epsilon_a\ \ for\ a \in BW \tag{1}$$

Here, $f(vote\ margin)_a$ controls for the vote margin running variable, and BW is the optimal bandwidth around the cutoff. I test for discontinuities in the average light density around the RD cutoff, allowing for the slope of the vote margin to vary at the cutoff. β measures the RD coefficient. Given that the RD estimates the Local Average Treatment Effect (LATE), I make causal claims for the sub-sample of assemblies close to the winning margin cutoff. Table 1 shows that assemblies narrowly aligned with the ruling party consume 0.44 log points more electricity than those that do not, after the 2011 elections. Given balance across the RD cutoff on the running variable and underlying assembly demographics (Figure 3), this discontinuity suggests differential treatment by the politicians in power.

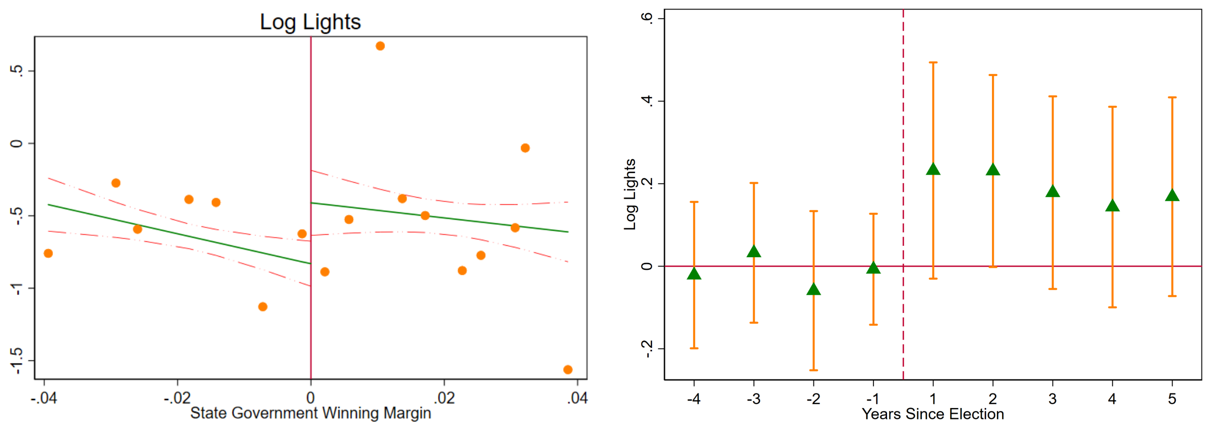
Table 1: Discontinuity in Actual Electricity Consumption (West Bengal)

	log(mean lights)
RD Estimate	0.436***
	(0.110)
Observations	1,356

Notes: I report the RD estimate using a bandwidth of 4.17 pp winning vote margin. This table shows evidence of discontinuously higher actual electricity consumption in West Bengal after the 2011 elections. Figure A11 shows robustness of these results for a wide range of bandwidths. Standard errors in parentheses clustered at the assembly level *** p<0.01, ** p<0.05, * p<0.1

Figure 5: Satellite Night Lights: RD Plot and Difference-in-discontinuities Analysis

(a) RD Plot (b) Difference-in-discontinuities



Notes: In the left panel, there is a discontinuously higher density of nighttime lights in assemblies just aligned with the ruling party after the 2011 election (2012-2016). I use an optimal bandwidth of 4.17 pp. The right panel plots RD coefficients over time (2007-2016) and finds a trend break after the 2011 election in West Bengal, with selectively greater electricity consumption in areas where the ruling party narrowly won in 2011. This graph uses an optimal bandwidth of 3.4 pp for each year. I plot the RD coefficients and confidence intervals of errors clustered at the assembly level. The dependent variable is Log(light density). Figure A11 shows robustness of these results for a wide range of bandwidths.

In Figure 5, the left panel demonstrates that there is discontinuously higher light density for assemblies where the ruling party narrowly won. To further investigate this pattern, I use nighttime lights data from 2006-2016. I study how being above the 2011 winning margin cutoff affects light density both before the elections (2007-2010) and after (2012-2016). The pre-2011 years serve to check whether there was a pre-election trend towards discontinuously higher electricity consumption. The coefficients after 2011 map out the post-election dynamics, as a consequence of the constituency being aligned with the ruling party. I estimate a difference-in-discontinuities specification, described by Equation 2, which includes year γ_t and assembly fixed effects μ_a, and restricts the sample to a bandwidth around the cutoff. β_t is the coefficient of interest across years.[9]

$$Log(Lights)_{at} = \sum_t \beta_t(\mathbb{1}(vote\ margin > 0)_a) + \gamma_t + \mu_a + \epsilon_{at} \quad for\ a \in BW \qquad (2)$$

On graphing these coefficients in Figure 5, the right panel, I observe that there was no

[9]In 2016, West Bengal had 294 assemblies spread across 23 administrative districts.

discontinuity or differential electricity consumption in the years before the 2011 elections. After 2011, there is a break, and I observe an increase in differential electricity consumption in assemblies where the ruling party narrowly won. Taken in isolation, this evidence may imply that there is differential access to electricity that is provided to the constituents of the winning party. However, this alone is not sufficient to understand the underlying dynamics, as I show using the administrative billing data below.

5.2 Data Manipulation in Electricity Billing Records

Administrative individual-level consumer data directly obtained from the state utility provides a useful companion to the satellite data described above. While the satellite data indicates actual electricity consumption, billing data documents consumption as reported by the utility. Similarities or divergences between these two datasets could be useful in understanding potential corruption by politicians. I show evidence of a discontinuity in Figure 6 using consumption data on all consumer classes, including households, commercial users, public works, agriculture, and irrigation.[10] For each post-election year and consumption category, I estimate the following specification at the individual i account level, where the outcome is electricity consumption. I present the consumer-wise RD results in Table 2.

$$y_{ia} = \beta \ \mathbb{1}(vote\ margin > 0)_a + f(vote\ margin)_a + \epsilon_{ia} \ \ for\ a \in BW \tag{3}$$

In Figure 6, using the consumption data reported by the electricity utility, I observe a discontinuously lower level of average electricity consumption in assemblies that narrowly swung in the ruling party's favor. I find this to hold using a wide range of bandwidths from 1.6 pp to 7 pp (Figures in Sections D.1 and D.3). Further, the magnitudes of these discrepancies are large, amounting to average discounts to constituents of about 40% of their regular bills.[11] This result is in contrast to the previous section, where we observed a discontinuously higher level of nightlights density.

A potential possibility with using satellite data is that it may primarily capture an in-

[10]The only consumer class not present in the dataset shared with me is high-tension industrial consumers of electricity (usually large factories). However, this does not present a concern for the results in the paper because, given that factories do not commonly operate at night, the nighttime lights data should closely correspond to the consumers captured in the billing dataset.

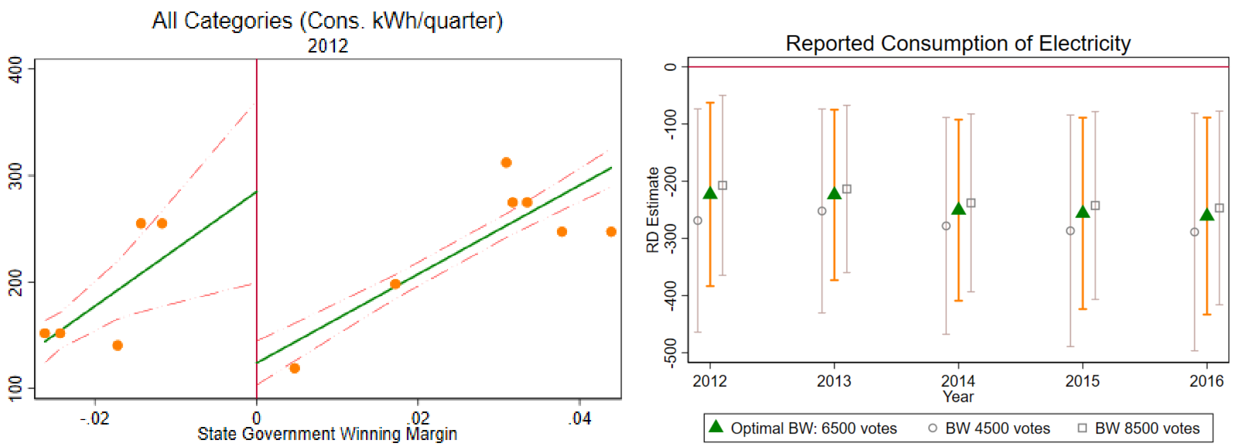
[11]These magnitudes are based on a simple calculation using the estimated effects of being in an aligned constituency, and the average electricity consumption at the cutoff in opposition-party assemblies.

Table 2: Discontinuity in Reported Consumption

			Unit Consumption in KWH			
			Residential (Rural)			
RD Estimate	-135.9***	-120.2***	-123.2***	-140.3***	-155.1***	-136.8***
Std. Error	(21.64)	(24.24)	(20.47)	(21.02)	(22.44)	(23.51)
Observations	53,084	8,526	11,262	11,176	11,128	10,992
Bwidth	6500	6500	6500	6500	6500	6500
Year	**Stacked**	**2012**	**2013**	**2014**	**2015**	**2016**
			Residential (Urban)			
RD Estimate	-376.8***	-304.5***	-361.2***	-377.3***	-396.5***	-428.1***
Std. Error	(80.86)	(96.76)	(83.92)	(79.25)	(76.96)	(73.02)
Observations	58,225	10,316	12,122	12,075	11,965	11,747
Bwidth	6500	6500	6500	6500	6500	6500
Year	**Stacked**	**2012**	**2013**	**2014**	**2015**	**2016**
			Commercial (Rural)			
RD Estimate	62.42	119.2	46.83	77.51	-18.61	101.5
Std. Error	(75.90)	(98.40)	(78.51)	(69.61)	(80.17)	(87.96)
Observations	20,946	3,368	4,465	4,391	4,374	4,348
Bwidth	6500	6500	6500	6500	6500	6500
Year	**Stacked**	**2012**	**2013**	**2014**	**2015**	**2016**
			Commercial (Urban)			
RD Estimate	-546.1**	-464.8*	-577.0**	-549.5**	-537.8**	-579.7**
Std. Error	(258.4)	(273.5)	(250.6)	(234.0)	(264.7)	(291.2)
Observations	66,623	11,975	13,917	13,656	13,686	13,389
Bwidth	6500	6500	6500	6500	6500	6500
Year	**Stacked**	**2012**	**2013**	**2014**	**2015**	**2016**

Notes: I report the RD coefficients across years for reported electricity consumption for each consumer class, controlling for the size of the electorate in each assembly. These results are robust across multiple regression specifications. The results in this table use a bandwidth of 6,500 votes in terms of the running variable, winning margin, corresponding to the optimal bandwidth of 4.17 pp vote share margin. This table shows evidence of discontinuously lower reported consumption for residential (urban and rural) consumers, as well as commercial (urban) users. Standard errors in parentheses clustered at the feeder level *** p<0.01, ** p<0.05, * p<0.1

Figure 6: Lower reported consumption in regions where the ruling party won (2012-15)



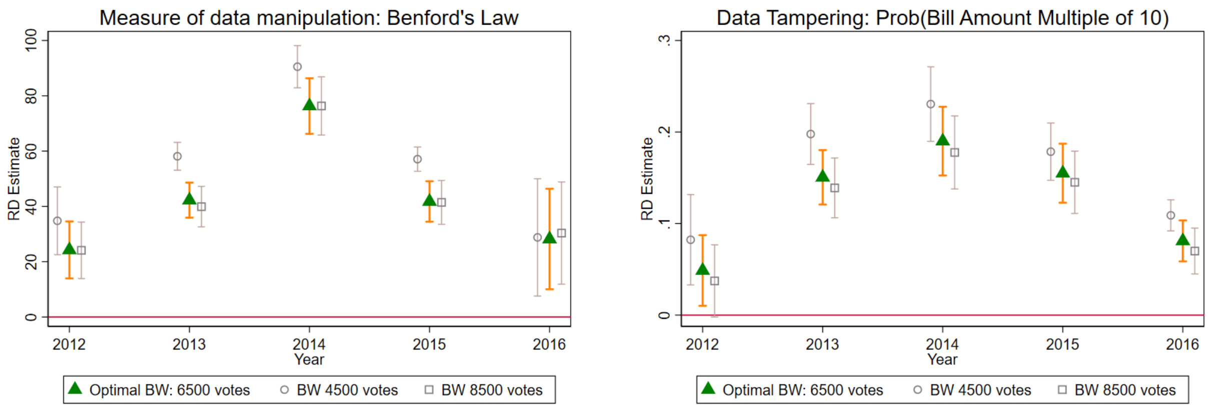
Notes: I plot the reported consumption of electricity on either side of the cutoff. The running variable for the RD is the winning margin percentage. The optimal bandwidth is ± 4.169 percentage point vote margin for the ruling party, which corresponds to a winning margin of -7109 to 6650 votes in various constituencies. In the right panel, I plot the RD coefficients between 2012 and 2016, and find results robust to other bandwidths (in terms of the number of votes) between ± 4500 and ± 8500 votes, showing 95% confidence intervals. Here, a ± 4500 vote margin corresponds to between -2.74 and 3.18 percentage point, 6500 votes corresponds to between -3.88 and 4.45 percentage point, and 8500 votes correspond to between -5.77 and 6.31 percentage point vote margin for the ruling party. These results remain robust to much smaller and much bigger bandwidths as well from 1.6 pp to 7 pp winning margin percentage (Sections D.1 and D.3). Standard errors are clustered at the feeder level and are robust to clustering at the assembly level.

crease in the extensive margin of electricity consumption, which billing records may not capture. Indeed, the Rajiv Gandhi Grameen Vidyutikaran Yojana (RGGVY) in India, launched in 2005, sanctioned the electrification of unelectrified villages all over the country. Looking at the assemblies just below and above the RD cutoff, the number of villages receiving electricity connections through the RGGVY scheme is very similar: 5944 compared to 6024 in constituencies aligned with the ruling party.[12] Further, the bulk of new electrification in India happened before 2011.

Next, I examine patterns in the data that may shed light on the observed under-reporting of electricity consumption, using the measures of data manipulation described in Section 3.2. In Figure 7, I find that the measure of distance (of the consumption distribution) from the expected chi-squared distribution (based on Benford's (1938) Law) is statistically significantly higher in assemblies just aligned with the ruling party. These results are echoed by the RD on the likelihood of billed consumption being reported as multiples of ten, which

[12] Author calculations from statistics by the Ministry of Power, India.

Figure 7: RD Coefficients for Manipulation Outcomes Across Bandwidths



Notes: I plot coefficients across years for measures of data manipulation, and confidence intervals of standard errors clustered at the electrical-feeder level. Specifically, I study the distance of the observed distribution from the expected distribution as per Benford's (1938) Law and the fraction of consumers whose consumption was a multiple of ten. 'BW' indicates the bandwidth size. The bandwidth of 6500 votes corresponds to the optimal bandwidth of 4.17 pp winning margin percentage used throughout this paper. For both outcomes, I plot the RD coefficients between 2012 and 2016 and find results robust to other bandwidths – both lower and higher than the optimal bandwidth of 6500 votes (between 4500 and 8500 votes). These regressions control for the total size of the electorate within each assembly.

is systematically higher in constituencies represented by the ruling party. *Ex-ante*, there would be no reason for these areas to see an anomalously high incidence of KWh that are neatly rounded off, but coupled with Figures 5 and 6, the results point towards politically motivated under-reporting of electricity consumption in assemblies aligned with the ruling party. The degree of data manipulation grows over time, and then the discontinuity falls by 2016, on the eve of the next election. From the available data, it is difficult to distinguish if this occurs because there is a higher degree of data manipulation in losing assemblies as well, or that politicians direct their efforts elsewhere in the run-up to the next election.

5.3 Mechanisms of Political Corruption in Electricity

This paper presents evidence that there may be a systematic under-reporting of electricity consumption in the state of West Bengal by manipulating billed consumption for connected constituents. In the Indian context, the setup of the bureaucracy around electricity provision appears well-suited to control by local representatives, given the close oversight of local electricity billing and distribution centers. There is research in other contexts to suggest that politicians can indirectly influence lower levels of the bureaucracy that may be involved in day-to-day transactions reflected in administrative micro-data (Barnwal, 2019; Lowe et al., 2020; Neggers, 2018; Weaver, 2021).

There are a number of ways that the electricity sector may be susceptible to manipulation. Electricity meter readings provide one of the few manipulable margins on which to affect electricity prices. In order to bill consumers, electricity utilities send meter readers to account holders' premises to manually record consumption. To a large extent, due to the absence of additional checks, reported consumption is up to the discretion of these meter inspectors and the local Customer Care Centers (CCCs) they report to, who then manually enter their reported consumption figures into the database. This is a possible point at which under-reporting occurs.[13] Indeed, among several vulnerabilities, Gulati and Rao (2007) identify the billing stage as susceptible to political interference, highlighting artificially lowered bills as a specific example. An audit study carried out by an electricity utility in Uttar Pradesh, another Indian state, identified significant political interference in electricity distribution and billing at local levels (Goenka, 2013). Rains and Abraham (2018) highlight the role of these inspectors in bill collection and how redesigning their incentives could lead to massive gains in utility revenue. My findings are consistent with a selective lack of enforcement in inspector readings, in order to allow local billing centers under the purview of the MLAs to report billed consumption that is lower than actual levels. Further, it is a relatively low effort to systematically under-report electricity consumption as a part of the routine data entry, making this type of political targeting quick to implement after an election.

Another way politicians may exploit the electricity sector is by selectively discouraging utility action against energy theft. Even though I am unable to test this directly, there is a large amount of anecdotal evidence supporting this channel (The Telegraph, 2014; The Times of India, 2017; The Washington Post, 2012).[14] While this is consistent with lower reported consumption and higher actual consumption, it cannot alone explain the discontinuously higher levels of data manipulation in constituencies controlled by the ruling party.

[13]Over the course of my fieldwork, I observed several instances of meter readers not conducting their inspection rounds for multiple billing periods. While the billing center handbooks recommend a formula to impute consumption from previous readings, there is discretion involved in the data entered. It is also widely acknowledged that MLAs hold a great deal of sway over local government authorities, and, therefore could potentially influence local billing centers. These billing centers are dispersed all over the state, but it is in narrowly aligned assemblies that we observe statistically significantly lower levels of reported consumption.

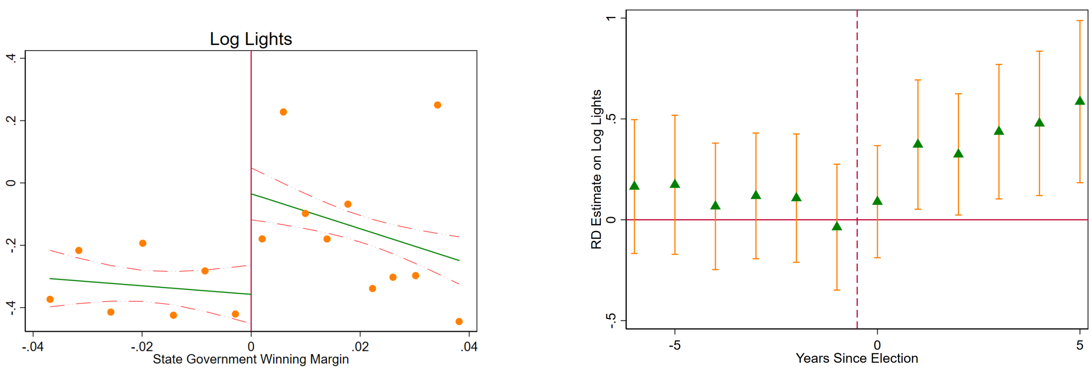
[14]" "Many people known to support the ruling party are allegedly involved in hooking and tapping", a source said.... The chief minister had accused WBSEDCL of "callousness" and questioned the efficacy of such [anti-theft] drives." The Telegraph, July 31st, 2014: *Power Theft Test for Mamata - State Utility to Seek CM's nod to Relaunch Crackdown.*

6 External Validity and Robustness Checks

The empirical results presented so far are focused on West Bengal because of the availability of state administrative billing data from 2012-2016. However, satellite nighttime lights data is available for more years, and for states beyond West Bengal. Below, I present evidence that political patronage using the electricity sector likely extends beyond West Bengal to other parts of the country, spanning multiple elections and political parties.

6.1 The Results Extend to Other States in India Across Elections

Figure 8: Higher actual electricity consumption in ruling party regions (All India 2006-16)



Notes: In the left panel, comparing legislative assemblies where the ruling party narrowly won to those where it narrowly lost (2006-16), I find a discontinuously higher density of nighttime lights in winning areas. This graph uses the optimal bandwidth of 4.17 pp winning vote margin to be comparable to the billing results in the paper. I show this result to be robust across a wide range of bandwidths from 2.5-7 pp (Figure A11). The right graph plots the RD coefficients and confidence intervals over time and finds a trend break after the election year, with selectively greater electricity consumption in areas where the ruling party narrowly won. This graph is based on an optimal bandwidth of 7 pp (winning margin) using Calonico et al. (2015). Errors clustered at the assembly level. This result is robust to other functional forms of nighttime lights, including levels (Figure A15).

Figure 8 presents two panels: one showing a close-election RD for *actual* electricity consumption for all of India, as measured by nighttime lights. The right-hand-side panel shows the RD estimate by year relative to a state election – I include 6 years before and after each election. The graph shows results from elections across multiple states having elections in different years from 2006-2016. These patterns are consistent with what I find for West Bengal: no detectable discontinuities in electricity consumption before an election, but a higher electricity consumption in constituencies aligned with the ruling party after. Table 3 shows the magnitude, indicating that electricity consumption is almost 0.48 log points higher in constituencies aligned with the ruling party.

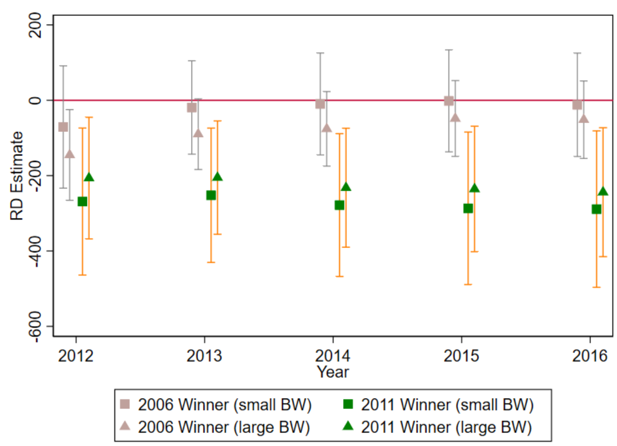
Table 3: Discontinuity in Actual Electricity Consumption (All India)

	log (mean lights)
RD Estimate	0.484***
	(0.127)
Observations	29,747

Notes: Evidence of discontinuously higher actual electricity consumption overall in assemblies that were in close elections and aligned with the ruling party after an election across India in a bandwidth of 4.17 pp vote share margin. Figure A11 shows robustness of these results for a wide range of bandwidths. Standard errors in parentheses clustered at the assembly level *** p<0.01, ** p<0.05, * p<0.1

6.2 Falsification Tests and Lagged Effects

Figure 9: Studying discontinuities in reported consumption using the winning and losing constituencies from the 2006 election



Notes: I plot RD coefficients for the reported consumption in constituencies aligned with the ruling party that won the 2006 election (CPI-M or the 2006 winner) and the 2011 election (AITC or the 2011 winner). The winning margin for the 2006 election results is defined on the basis of legislative assemblies from the 2006 electionand the winning margin for the 2011 election results is defined on the basis of legislative assemblies from the 2011 election, where the AITC party won. However, while the running variables are created from different elections, we examine the outcomes for electricity post-2011. This provides a falsification test using post-2011 electricity. The results shown include multiple bandwidths (BW 4500 votes to 8500 votes).

The billing data begins in 2011 and cannot be used to examine the aftermath of the 2006 elections. But, in 2012, Figure 9 shows suggestive evidence that the CPI(M) (ruling party after the 2006 elections) engaged in the same form of electricity under-reporting in the constituencies they won in 2006. Immediately after the 2011 elections, when the AITC defeated CPI(M), there is evidence of consumption under-reporting in CPI(M) controlled constituencies. However, this small effect fades out quickly, such that after 2012, there is no statistically detectable discontinuity when the AITC may be establishing itself and begins its own machinery to under-report its constituents' electricity consumption. These patterns are mirrored in tests on additional outcomes in Appendix Figure A4.

6.3 Heterogeneity Analysis

I study different sets of states to understand the patterns in discontinuously higher electricity consumption in some places and not others. First, going by the analysis in Pargal and Banerjee (2014), I pick the three worst and best-performing states in terms of utility distribution revenue losses. The three worst-performing states in this regard in 2010-11 were Bihar, Manipur, and Odisha, with the highest losses, and the best-performing states were Chhattisgarh, Karnataka, and Andhra Pradesh. Following the RD design in Table 3, I separately estimate the RD coefficient for these two sets of states in Columns 1 and 2 of Table 4. I find that there are higher discontinuities (statistically significant) across the RD cutoff in states with higher distribution losses. This appears consistent with a narrative that there are higher losses to electricity utilities in states where there is more corruption (or states with higher losses may also be those where there is a lower cost to corruption).

Recent work shows that states with single electricity distributors provide more reliable electricity and are better at revenue collection compared to states with multiple electricity distributors (Mahadevan, 2022). I examine whether there is discontinuously more electricity consumption in ruling-party-aligned constituencies in states with single or multiple electricity distributors. I find in Columns 3 and 4 of Table 4 that there is a large discontinuity in states with multiple distributors, which to a large degree is consistent with what Mahadevan (2022) finds. States with multiple distributors struggle to meet the electricity demand of firms, and perform poorly in revenue collection – perhaps this opens up the demand for more electricity or preferential access, which politicians can take advantage of by favoring their allies. One of the hypotheses discussed in Mahadevan (2022) argues that states with multiple distributors suffer from coordination failures that lead to low revenue collection, as well as poor supply. It is possible that this same reason also makes them more susceptible to corruption.

Finally, using ratings assigned by the Ministry of Power (2019) based on a range of characteristics like finances, complaints files, and audits required, I assign state electricity utilities that received a grade of A+ to the list of best-performing states Karnataka, Gujarat, and Uttarakhand), and those that received a grade of C into the worst utilities (Manipur, Meghalaya, Tripura, Uttar Pradesh and Madhya Pradesh). In Columns 5 and 6, of Table 4, I show that there is a large, statistically significant discontinuity in electricity access in the states with the worst rated utilities, while there is a smaller, statistically insignificant effect in states with the best-rated utilities. These results suggest that greater accountability in the functioning of utilities may deter the capture of utilities for political corruption.

Table 4: Discontinuity in Actual Electricity Consumption - Comparing sets of states with single and multiple electricity distributors, high and low revenue losses from distribution, and good and bad ratings.

	(1) Low Dist Losses log (lights)	(2) High Dist Losses log (lights)	(3) Single Discom log (lights)	(4) Multiple Discom log (lights)	(5) Worst Utilities log (lights)	(6) Best Utilities log (lights)
RD Estimate	0.340***	0.628***	0.0310	0.359***	0.207**	0.120
	(0.0694)	(0.0317)	(0.0633)	(0.0639)	(0.0876)	(0.103)
Observations	8,185	2,960	9,576	23,931	11,469	4,158

Notes: I show heterogeneity in the actual electricity consumption in aligned assemblies across different sets of states. Using Pargal and Banerjee (2014), I find sets of states that had the highest and lowest reported financial losses in their electricity sector; and sets of states that had a single electricity provider in the state and multiple. Finally, using ratings assigned by the Ministry of Power (in 2019) based on a range of characteristics like finances, complaints files, and audits required, I assign state electricity utilities that received a grade of A+ to the list of best-performing states and those that received a grade of C into the worst utilities. Standard errors in parentheses clustered at the feeder level *** p<0.01, ** p<0.05, * p<0.1

6.4 Robustness Checks

In Appendix D, I discuss a range of robustness checks that validate the results in this paper. The RD estimates for a range of billing and lights outcomes are robust to both smaller and larger bandwidths (Appendix Sections D.1, D.3 and D.2). While I present the main results from the administrative data clustering at the electricity feeder level (the level at which electricity supply is highly correlated and governed by a single customer care center), I also show robustness to clustering at the assembly level (Appendix Section D.4). I show the main results for all major consumer categories, but there may be an argument to be made that agricultural consumers are unique given the already large subsidies they receive and their use of electricity primarily for irrigation (not visible using nighttime lights). Therefore I show that the main results remain robust to dropping agricultural consumers, who comprise about 1.5% of the consumer base (Appendix Section D.5. I show that the results using nighttime lights remain robust to using levels instead of the commonly used log transformation (Appendix Section D.6).

7 Welfare Consequences of Political Patronage

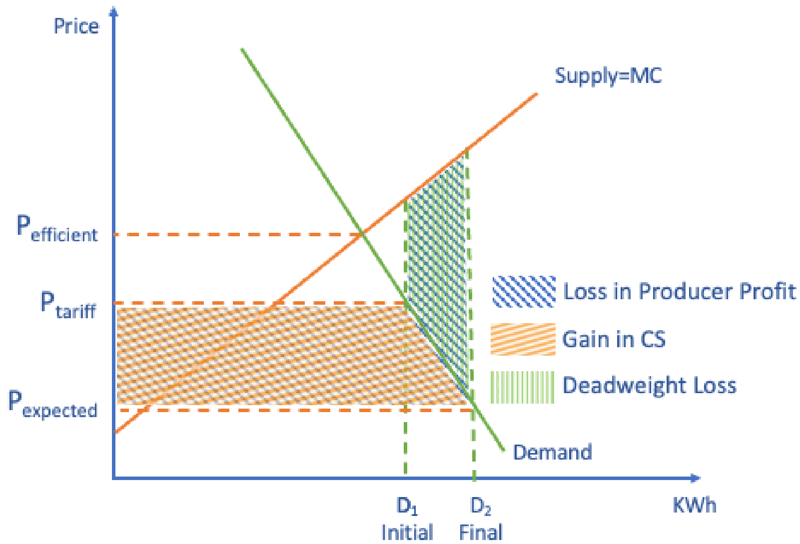
The electricity context sets itself apart from other instances of corruption by allowing for a full accounting of the economic costs involved. The combination of administrative data from the ground and satellite level data to describe the full picture is key to not only detecting corruption but also identifying the welfare consequences. The corruption described in this paper would not be problematic from an efficiency perspective if it involved merely a transfer from one group to another. However, the distortions in electricity billing lead to a deadweight loss that outweighs gains to any group, which makes this form of patronage costly. In this section, I describe how the paper quantifies welfare costs.

This paper characterizes the under-reporting in billing data as providing an informal subsidy to constituents of the ruling party, described in Figure 10. Under an efficient market, the price charged for electricity would be $P_{efficient}$. As a consequence of political patronage, consumers in constituencies of the ruling party effectively face a price of $P_{expected}$.

Figure 10 describes the loss in producer profits, gain in consumer surplus, and deadweight loss to society as a result of the informal subsidies provided by politicians to their constituents. In order to estimate the change in producer profits, I use a combination of the over-consumption of electricity in response to the subsidy from Table 1, and the implicit price subsidy conferred by the under-reporting of consumption. I provide details of these calculations in Appendix Section F.1.

I then estimate the gain in consumer surplus. The difference between the changes in producer and consumer surplus provides the deadweight loss to society. However, as is evident from Figure 10, the change in consumer surplus depends on the price elasticity of

Figure 10: Market Distortions Due to Political Corruption



Notes: I simplify the indirect subsidies by politicians through under-reporting in billed data, by assuming an average level of electricity subsidy for all electricity consumers in regions aligned with the ruling party. $P_{efficient}$ refers to the market clearing price of electricity, but this is not used in electricity markets. The most common pricing scheme is to cross-subsidize residential, small commercial establishments, and agricultural consumers by charging high rates for large industrial users, so usually, consumers face prices lower than $P_{efficient}$. I assume that rather than an upward-sloping block-price schedule, consumers are supposed to face a flat rate of P_{tariff}. Politicians, through corruption, may effectively lower this price even further for their constituents, to $P_{expected}$. I assume that the marginal cost (MC) curve facing producers is upward-sloping; however, this may be simplified to a flat cost curve as well. The shaded areas show the loss in producer profit, gain in consumer surplus, and overall deadweight loss.

demand for electricity, to infer the effective change in marginal price for beneficiaries of the subsidy. I also use this estimate to compute a portion of the loss to producers. Therefore, I first estimate the price elasticities of demand across consumer categories. I allow for the fact that the four consumer categories I focus on, residential rural, residential urban, commercial rural, and commercial urban, each have different elasticities.

However, estimating the price elasticities of demand from the consumption data is not straightforward, given the data manipulation. I, therefore, develop a method of deriving elasticities that accounts for anomalies. As the *first step*, I select assemblies where I statistically reject that the data is manipulated (described in Appendix Section E). I then compute elasticities for each assembly and consumer category for this sub-sample using an instrumental variable approach that leverages exogenous variation in policy-led tariff changes over time (Appendix E.1). The *second step* involves building a predictive model for elasticities in assemblies with no manipulation (Appendix E.2). In the *third step*, the paper predicts elasticities for the remaining constituencies where there is evidence of data manipulation (details described in Appendix E.3). The result is a unique estimate for elasticity for four consumer groups in each assembly in the dataset (Table A4 in Appendix E.3). The advantage of this method over previous estimates of price elasticities using aggregated billing data is that the individual-level billing data allows the distinction of tariff changes within consumer group, tier, assembly, and month for better identification.

The *final step* uses the full set of estimated and predicted elasticities to calculate the consumer surplus and producer loss for each consumer class as a result of the informal subsidy provided by politicians. The elasticity estimates I compute advance the literature by updating the residential and commercial (both urban and rural) elasticity estimates that were in use before (Saha and Bhattacharya, 2018). The elasticities I estimate are within the range of prior work for residential consumers but are significantly higher for commercial accounts. I argue that prior estimates used aggregate data that may conceal problems such as data manipulation. But correcting for manipulation in my estimates, I arrive at figures that may be more reflective of what firms report: having to frequently use generators, and taking advantage of being able to switch in response to high marginal prices of electricity. Section F.2 provides details of how the consumer surplus is calculated.

7.1 Costs & Benefits of Political Manipulation of Electricity Bills

Table 5: Net welfare costs

	Welfare costs and benefits		
		Losses (Mill \$)	**Source**
1	Loss in producer profits for under-reporting and over-consumption response in aligned assemblies	1588	Table A6
2	Gain in consumer surplus for constituents of aligned assemblies	1000	Table A7
3	Net deadweight loss	588	Row 1 - Row 2

Notes: The computations of losses to producers, and gains to consumers hinge on a combination of the administrative data for estimating the implicit subsidy from the under-reporting, and the over-consumption as a result of the subsidy. Importantly, both the losses to producer profits and the gains to consumers depend on being able to connect the satellite nighttime lights data to the administrative data, and translate a luminosity measure to KWh (Figures A5 and A7). Additionally, the welfare analysis uses data on price tariffs (Figure A17) and the Indian Census 2011 (Table A3).

The under-reporting of consumption in bills leads to large welfare distortions not only for the electricity producer and the consumers involved but also more widely for the economy. The combination of the implicit subsidy imposed by consumption under-reporting and the satellite data to identify over-consumption in the same regions leads to a loss of \$1.8 billion in electricity producer profits in West Bengal in a single electoral term (Table 5). My results demonstrate that there is a likelihood that this form of corruption extends to other states (Figure 8), likely pegging the losses to electricity producers magnitudes higher. Conversely, the gains to consumers stand at \$1.2 billion. As a result, the efficiency loss amounts to \$0.6 billion. I also benchmark the welfare estimates using elasticities from prior work. Since the elasticities I estimate are, on average higher (for the small group of urban commercial consumers), I find that, in effect, I compute a lower bound for the gain in the consumer surplus and loss in producer profits. Both these surpluses change by similar amounts, leaving deadweight loss to be only slightly smaller, for instance, when using the elasticities in Saha and Bhattacharya (2018) (Tables A8 and A9).

Losses in electricity producer profits have consequences beyond these estimates, as they disadvantage *all* consumers, even those who the politicians sought to favor. Losses hurt the ability of utilities to meet electricity demand, resulting in poor quality and unreliable supply for all consumers Burgess et al. (2020).

8 Conclusion

This paper demonstrates that political patronage can have adverse welfare implications that go well beyond the transfers caused by selective access to a few favored groups and imposes a significant efficiency loss for society. A major innovation in this paper is to use a combination of administrative billing data for 72 million electricity customers with satellite data to draw a distinction between measured electricity consumption, and actual consumption. I present evidence that politicians favor their voters both in terms of providing electricity access and in subsidizing them by under-reporting their billed consumption. Consistent with the hypothesis that political agents may influence intermediaries to manipulate the data, I find that in constituencies where the ruling party narrowly won, there are greater anomalies in the consumption distribution. This helps me demonstrate evidence of politically motivated data manipulation, as well as isolate the methods used to carry it out. Both of these elements help me capture the impact of political manipulation on electricity providers and consumers.

Understanding the effects on net welfare is a particular contribution of this work, as unlike other settings studied in the literature, the electricity setting is unique in its ability to allow for a full accounting of gains and losses.

I calculate the total deadweight loss as \$0.6 billion. These numbers represent estimates for a single Indian state but could be much larger if scaled to the over thirty Indian states that have similar vulnerabilities to political manipulation. Targeted voters in constituencies aligned with the ruling party may benefit from cheaper electricity, but the loss in profits for the electricity provider, in particular, has wide-ranging implications. If the funds used to bail out the utilities cut into the government's developmental budgets, then these bailouts may be detrimental to poorer sections of society and have wide-ranging welfare consequences. Indeed, the bail-out of electricity utilities in India has been virtually systematized by the set up of a centralized bailout fund through the Ujwal DISCOM Assurance Yojana (UDAY) (Chatterjee, 2017, 2018). Further, increased outages and unreliable electricity that result from insufficient revenue have large implications for growth and productivity (Allcott et al., 2016; Fried and Lagakos, 2023). While research on manipulation in administrative data has explored anomalies arising from measurement error, misreporting by consumers, insufficient incentives for data collectors, and eligibility manipulation, the possibility of politically motivated manipulation remains largely unexplored (Camacho and Conover, 2011; Slemrod, 2016). Given its large impact on policy-making, ability to provide public goods, and measurement of development progress, this is an important area for future study.

References

Ahrens, A., Hansen, C. B., and Schaffer, M. E. (2018). PDSLASSO: Stata module for post-selection and post-regularization OLS or IV estimation and inference. Statistical Software Components, Boston College, Department of Economics.

Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic Inequality. *Journal of Political Economy*, 124(2):428–488.

Allcott, H., Collard-Wexler, A., and O'Connell, S. D. (2016). How Do Electricity Shortages Affect Industry? Evidence from India. *American Economic Review*, 106(3):587–624.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58(2):277–97.

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development Research at High Geographic Resolution: An Analysis of Night Lights, Firms, and Poverty in India using the SHRUG Open Data Platform. *The World Bank Economic Review*.

Asher, S. and Novosad, P. (2017). Politics and Local Economic Growth: Evidence from India. *American Economic Journal: Applied Economics*, 9(1):229–273.

Bardhan, P. and Mookherjee, D. (2010). Determinants of Redistributive Politics: An Empirical Analysis of Land Reforms in West Bengal, India. *American Economic Review*, 100:1572–1600.

Barnwal, P. (2019). Curbing Leakage in Public Programs: Evidence from India's Direct Benefit Transfer Policy. *Working Paper*.

Baskaran, T., Min, B., and Uppal, Y. (2015). Election Cycles and Electricity Provision: Evidence from a Quasi-experiment with Indian Special Elections. *Journal of Public Economics*.

Belloni, A., Chernozhukov, V., Hansen, C., and Kozbur, D. (2016). Inference in High-Dimensional Panel Models With an Application to Gun Control. *Journal of Business & Economic Statistics*, 34(4):590–605.

Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572.

Bhalotra, S., Clots-Figueras, I., and Iyer, L. (2018). Pathbreakers? women's electoral success and future political participation. *The Economic Journal*, 128(613):1844–1878.

Bose, R. K. and Shukla, M. (1999). Elasticities of Electricity Demand in India. *Energy Policy*, 27(3):137–146.

Brown, R., Mansour, H., and O'Connell, S. D. (2021). Does Local Female Political Representation Empower Women to Run for Higher Office? Evidence from State and National Legislatures in India. *The World Bank Economic Review*, 36(1):198–218.

Burgess, R., Greenstone, M., Ryan, N., and Sudarshan, A. (2020). The Consequences of Treating Electricity as a Right. *Journal of Economic Perspectives*, 34(1):145–69.

Burlig, F. and Preonas, L. (2023). Out of the Darkness and into the Light? Development Effects of Rural Electrification. *Journal of Political Economy.*

Callen, M., Gulzar, S., and Rezaee, A. (2020). Can Political Alignment Be Costly? *The Journal of Politics*, 82(2):612–626.

Calonico, S., Cattaneo, M., and Titiunik, R. (2015). Rdrobust: An R Package for Robust Nonparametric Inference in Regression-discontinuity Designs. *R Journal*, 7(1):38–51.

Camacho, A. and Conover, E. (2011). Manipulation of social program eligibility. *American Economic Journal: Economic Policy*, 3(2):41–65.

Chatterjee, E. (2017). Reinventing State Capitalism in India: A View from the Energy Sector. *Contemporary South Asia*, 25(1):85–100.

Chatterjee, E. (2018). The Politics of Electricity Reform: Evidence from West Bengal, India. *World Development*, 104:128–139.

Chhibber, P., Shastri, S., and Sisson, R. (2004). Federal Arrangements and the Provision of Public Goods in India. *Asian Survey*, 44(3):339–352.

Clots-Figueras, I. (2012). Are female leaders good for education? evidence from india. *American Economic Journal: Applied Economics*, 4(1):212–44.

Cruz, C., Keefer, P., Labonne, J., and Trebbi, F. (2020). Making Policies Matter: Voter Responses to Campaign Promises. *Working Paper.*

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

Dixit, A. and Londregan, J. (1996). The Determinants of Success of Special Interests in Redistributive Politics. *The Journal of Politics*, 58(4):1132–1155.

Eggers, A., Fowler, A., Hainmueller, J., Hall, A., and Snyder, Jr, J. (2015). On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races. *American Journal of Political Science*, 59(1):259–74.

Elvidge, C. D., Baugh, K., Zhizhin, M., and Hsu, F. C. (2013). Why VIIRS data are superior to DMSP for mapping nighttime lights. *Asia-Pacific Advanced Network*, 35:62.

Elvidge, C. D., Baugh, K., Zhizhin, M., Hsu, F. C., and Ghosh, T. (2017). VIIRS night-time lights. *International Journal of Remote Sensing*, 38(21):5860–5879.

Filippini, M. and Pachauri, S. (2004). Elasticities of Electricity Demand in Urban Indian Households. *Energy Policy*, 32(3):429–436.

Finan, F. and Schechter, L. (2012). Vote-Buying and Reciprocity. *Econometrica*, 80(2):863–881.

Fried, S. and Lagakos, D. (2023). Electricity and Firm Productivity: A General-Equilibrium Approach.

Fujiwara, T., Kanz, M., and Mukherjee, P. (2020). The Electoral Effects of a Fiscal Transfer: Evidence from Indian Elections. *Working Paper.*

George, S., Gupta, S., and Neggers, Y. (2018). Coordinating Voters against Criminal Politicians: Evidence from a Mobile Experiment in India. *Working Paper.*

George, S. E. and Ponattu, D. (2020). Like Father, Like Son? The Effect of Political Dynasties on Economic Development. *Working Paper*.

Gibson, J., Olivia, S., Boe-Gibson, G., and Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, 149:102602.

Goenka, S. (2013). Tackling Power Theft through Meter Data Management and Quality Analysis - Results from NPCL's AMR Roll Out and AMI Trial.

Golden, M. and Min, B. (2011). Corruption and Theft of Electricity in an Indian State.

Greenstone, M. and Jack, B. K. (2015). Envirodevonomics: A Research Agenda for an Emerging Field. *Journal of Economic Literature*, 53(1):5–42.

Guiteras, R. and Mobarak, A. M. (2014). Does Development Aid Undermine Political Accountability? Leader and Constituent Responses to a Large-Scale Intervention. *Working Paper*.

Gulati, M. and Rao, M. (2007). Corruption in the Electricity Sector. A Pervasive Scourge. pages 115–157.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2):994–1028.

Henderson, V., Storeygard, A., and Weil, D. N. (2011). A bright idea for measuring economic growth. *American Economic Review*, 101(3):194–99.

Hicken, A. (2011). Clientelism. *Annual Review of Political Science*, 14(1):289–310.

Hodler, R. and Raschky, P. A. (2014). Regional favoritism. *The Quarterly Journal of Economics*, 129(2):995–1033.

Imbens, G. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635.

Ito, K. (2014). Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing. *American Economic Review*, 7(3):537–63.

Jensenius, F. R. and Verniers, G. (2017). Studying indian politics with large-scale data: Indian election data 1961–today. *Studies in Indian Politics*.

Jeong, D., Shenoy, A., and Zimmermann, L. (2020). Is Corruption Compensation? Evidence from Local Public Office in India. *Working Paper*.

Khanna, G. and Mukherjee, P. (2023). Political Punishment and Financial Safety Nets: Evidence from India's Demonetization. *Journal of Public Economics*. Forthcoming.

Khwaja, A. and Mian, A. (2005). Do Lenders Favor Politically Connected Firms? Rent Provision in an Emerging Financial Market. *Quarterly Journal of Economics*, 120(4):1371–411.

Labonne, J. (2013). The Local Electoral Impacts of Conditional Cash Transfers: Evidence from a Field Experiment. *Journal of Development Economics*, 104:73–88.

Lehne, J., Shapiro, J. N., and Vanden Eynde, O. (2018). Building connections: Political corruption and road construction in india. *Journal of Development Economics*, 131:62–78.

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

Lowe, M., Prakash, N., and Rajendran, R. (2020). Do Bureaucrats Acculturate? Evidence From a Long-Running Natural Experiment in India. *Working Paper*.

Mahadevan, M. (2022). You Get What You Pay For: Electricity Quality and Firm Response . *Working Paper*.

Mahadevan, M. and Shenoy, A. (2023). The Political Consequences of Resource Scarcity: Targeted Spending in a Water-Stressed Democracy. *Journal of Public Economics*. Forthcoming.

Manacorda, M., Miguel, E., and Vigorito, A. (2011). Government Transfers and Political Support. *American Economic Journal: Applied Economics*, 7(3):1–28.

Mann, M. L., Melass, E. K., and Arun, M. (2016). Using VIIRS Day/night Band to Measure Electricity Supply Reliability: Preliminary Results from Maharashtra, India. *Remote Sensing*, 8(9):711.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714.

Michalopoulos, S. and Papaioannou, E. (2013a). National Institutions and Subnational Development in Africa *. *The Quarterly Journal of Economics*, 129(1):151–213.

Michalopoulos, S. and Papaioannou, E. (2013b). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Min, B. and Gaba, K. M. (2014). Tracking Electrification in Vietnam Using Nighttime Lights. *Remote Sensing*, 6(10):9511–9529.

Min, B., Gaba, K. M., Sarr, O. F., and Agalassou, A. (2013). Detection of Rural Electrification in Africa Using DMSP-OLS Night Lights Imagery. *International Journal of Remote Sensing*, 34(22):8118–8141.

Min, B. and Golden, M. (2014). Electoral Cycles in Electricity Losses in India. *Energy Policy*, 65:619–625.

Muralidharan, K., Niehaus, P., and Sukhtankar, S. (2016). Building State Capacity: Evidence from Biometric Smartcards in India. *American Economic Review*, 106(10):2895–2929.

Nagavarapu, S. and Sekhri, S. (2014). Politics off the Grid: Political Competition, Regulatory Control, and Allocation of Natural Resources. *Working Paper*.

Neggers, Y. (2018). Enfranchising Your Own? Experimental Evidence on Bureaucrat Diversity and Election Bias in India. *American Economic Review*, 108(6):1288–1321.

Nellis, G., Weaver, M., and Rosenzweig, S. C. (2016). Do Parties Matter for Ethnic Violence? Evidence From India. *Quarterly Journal of Political Science*, 11(3):249–277.

Pande, R. (2003). Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India. *American Economic Review*, 93(4):1132–1151.

Pande, R. (2020). Can Democracy Work for the Poor? *Science*, 369(6508):1188–1192.

Pargal, S. and Banerjee, S. G. (2014). *More Power to India : The Challenge of Electricity Distribution*. The World Bank.

Pop-Eleches, C. and Pop-Eleches, G. (2012). Targeted government spending and political preferences. *Quarterly Journal of Political Science*, 7(3):285–320.

Prakash, N., Rockmore, M., and Uppal, Y. (2019). Do Criminally Accused Politicians Affect Economic Outcomes? Evidence from India. *Journal of Development Economics*, 141:102370.

Rains, E. and Abraham, R. J. (2018). Rethinking Barriers to Electrification: Does Government Collection Failure Stunt Public Service Provision? *Energy Policy*, 114:288–300.

Saha, D. and Bhattacharya, R. N. (2018). An Analysis of Elasticity of Electricity Demand in West Bengal, India: Some Policy Lessons Learnt. *Energy Policy*, 114:591–597.

Sen, R. and Biswas, A. (2017). Order of West Bengal Electricity Regulatory Commission in the Matter of Case No. TP(R)-26/16-17. *WBERC*. Forthcoming.

Shenoy, A. and Zimmermann, L. (2022). Political Organizations and Political Scope. *Working Paper*.

Slemrod, J. (2016). Caveats to the Research Use of Tax-Return Administrative Data. *National Tax Journal*, 69(4):1003–1020.

Storeygard, A. (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies*, 83(3):1263–1295.

Stromberg, D. (2004). Radio's Impact on Public Spending. *The Quarterly Journal of Economics*, 119(1):189–221.

The Telegraph (July 31, 2014). Power Theft Test for Mamata - State Utility to Seek CM's Nod to Relaunch Crackdown. *Staff Reporter*.

The Times of India (July 18, 2017). Discom Engineer Death: Why Power Thieves Fear No One. *Staff Reporter*.

The Times of India (March 6, 2018). Rajasthan BJP MLA Backs Farmers Stealing Power. *Staff Reporter*.

The Washington Post (October 4, 2012). Power Thieves Prosper in India's Patronage-based Democracy. *Simon Denyer*.

The World Bank (2014). Enterprise Surveys (http://www.enterprisesurveys.org).

Weaver, J. (2021). Jobs for Sale: Corruption and Misallocation in Hiring. *American Economic Review*, 111(10):3093–3122.

Zimmermann, L. (2021). The Dynamic Electoral Returns of A Large Anti-Poverty Program. *Review of Economics and Statistics*, 103(5):803–817.

Appendix

Table of Contents

A Elections in West Bengal and Sample Details

In 2011, the election that I focus on for the administrative billing results in this paper, the All India Trinamool Congress won the absolute majority in the state legislative assembly elections. They won 184 out of the total 294 seats available (needing 148 to win the majority). Despite not needing to form a coalition to control the government, the AITC formed a post-election alliance with the Indian National Congress. The analysis in this paper, however, focuses on only the AITC as the ruling party and looks at the effects of being in a close-election AITC constituency versus a constituency they narrowly lost. This strategy has a higher generalizability across India, as not all states had such post-election alliances, and understanding the incentives in a coalition setting becomes more complex. Given that the AITC did not need the INC for the majority, they did not have a clear incentive to provide INC-aligned constituencies with the same electricity benefits as their own constituencies.

A.1 Details on data samples used for regressions

In Table A1, I show all the samples used in the regressions run in this paper (where the AITC won and lost), and summary statistics for variables of interest within each sample.

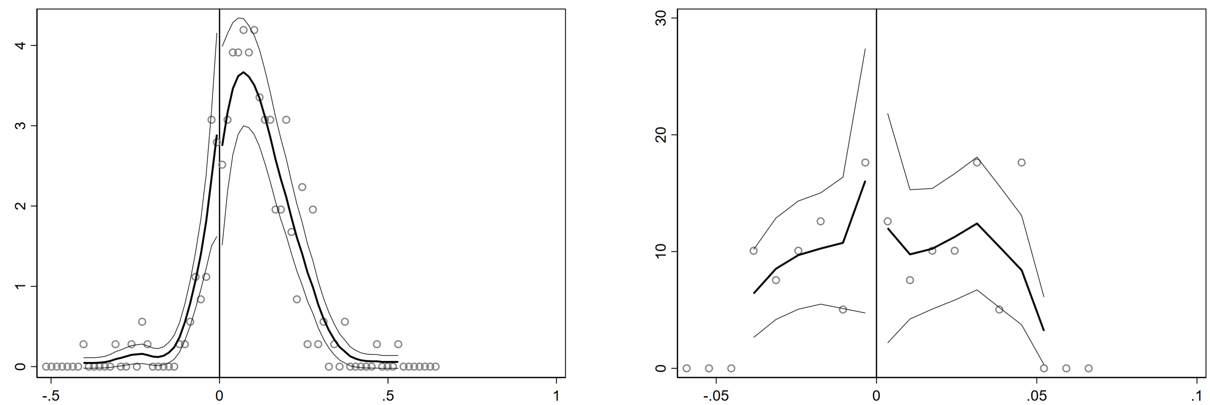
Table A1: Summary Statistics for Outcomes in Winning and Losing Legislative Assemblies

	Full Sample			BW: 4.17 pp			BW: 4500 votes			BW: 6500 votes			BW: 8500 votes		
	Lights data analysis														
	Total	**Won**	**Lost**	**Total**	**Won**	**Lost**	**Total**	**Won**	**Lost**	**Total**	**Won**	**Lost**	**Total**	**Won**	**Lost**
No. of assemblies in each sample	226	184	42	50	25	25	35	17	18	47	25	22	67	41	26
	Mean	**Difference**		**Mean**	**Difference**		**Mean**	**Difference**		**Mean**	**Difference**		**Mean**	**Difference**	
Lights Density	250122.79	274442.59		34895.92	28288.77		28999.65	17327.99		36325.12	27302.38		76690.49	85318.41	
Log Light Density	10.87	1.42		9.74	0.21		9.72	0.23		9.76	0.20		9.85	0.23	
	Billing data analysis														
	Total	**Won**	**Lost**	**Total**	**Won**	**Lost**	**Total**	**Won**	**Lost**	**Total**	**Won**	**Lost**	**Total**	**Won**	**Lost**
No. of assemblies in each sample	184	148	36	41	22	19	32	15	17	41	22	19	58	35	23
	Mean	**Difference**		**Mean**	**Difference**		**Mean**	**Difference**		**Mean**	**Difference**		**Mean**	**Difference**	
Reported cons. of Electricity (KWh)	250.84	101.31		230.80	96.34		156.01	-9.42		230.80	96.34		230.87	93.62	
Data manipulation	25.31	16.42		23.51	16.12		16.47	13.85		23.51	16.12		24.38	16.79	
Total Bill (Rs.)	1473.33	652.65		1352.83	626.81		865.59	-62.92		1352.83	626.81		1350.01	604.86	
Total Arrears in bill (Rs.)	86.56	48.64		82.53	53.10		42.07	-3.04		82.53	53.10		82.83	51.94	
Avg. energy price (Rs./KWh)	3.84	0.43		3.76	0.44		3.41	-0.08		3.76	0.44		3.77	0.44	
Total Subsidies in Bill (Rs.)	-149.66	-56.12		-144.72	-62.96		-98.92	-2.55		-144.72	-62.96		-143.69	-59.79	
Connected load (KVA)	1.07	0.33		1.06	0.21		0.94	0.11		1.06	0.21		1.05	0.20	

Notes: Summary statistics based on administrative billing data. The above table shows the mean level of the outcome variables by a range of bandwidths, including the optimal bandwidth for RD plots of 4.17 percentage points. I also show the difference between the means of legislative assemblies that are aligned ('Winning') and not aligned ('Losing') with the ruling party, for the 2011 election. I show billing outcomes from 2012 when my data begins.

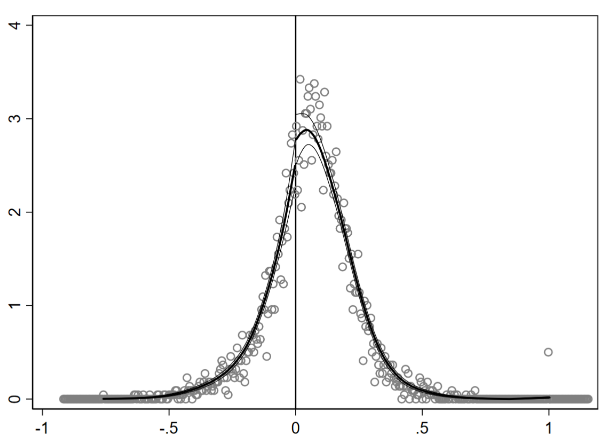
A.2 Balance Tests

Figure A1: McCrary Tests for Full Sample and RD Bandwidth



Notes: In the left panel, I test the smoothness of the density of the running variable (winning margin in the state election (2011)) for discontinuities and find that it is smooth across the RD cutoff. In the right panel, I run the same test but restrict it to the bandwidth of the main results in the paper.

Figure A2: McCrary Test for All India



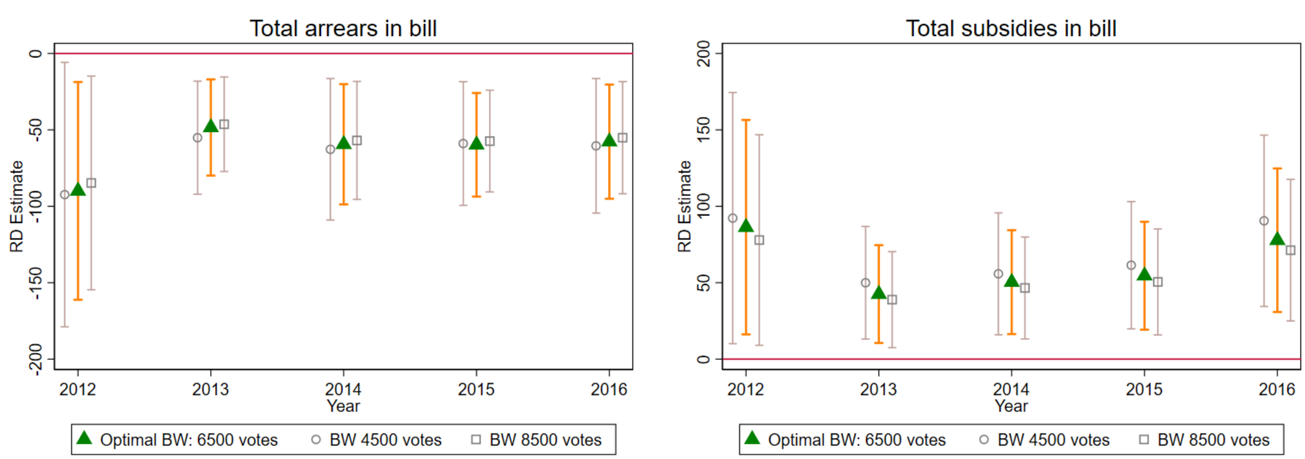
Notes: I test the smoothness of the density of the running variable (winning margin in any given state election between 2006-2016) for discontinuities and find that it is smooth across the RD cutoff.

B Additional Evidence on Data Manipulation

I exploit additional billing items in the data that shed more light on the mechanisms of data manipulation. The electricity bills consist of two items, "arrears" and "subsidies" that have complex formulas, leaving them open to manipulation that is hard to detect. Tariff increases are phased into consumer bills over a five-year period, using a system of arrears, which work like retroactive charges. However, tariff revisions occur every 1-2 years. Therefore the bill item "arrears" consists of components from multiple tariff increases, and anomalies are hard to identify.[15] The close-election RD provides a neat way of identifying whether these billing items are systematically different in constituencies supporting the ruling party.

I examine trends in the RD coefficient for potential manipulation of arrears and subsidy payments in Figure A3. I observe a statistically significantly higher level of subsidies in winning swing assemblies, accompanied by a lower level of arrears. Taken together with the evidence of lower reported consumption, this provides a consistent story. However, under-reporting consumption may translate mechanically to lower bills, with smaller arrears and higher subsidies as well. By under-billing residential users, politicians have effectively subsidized their electricity consumption and increased equilibrium electricity consumption.

Figure A3: Regression Discontinuity coefficients for outcomes across three bandwidths

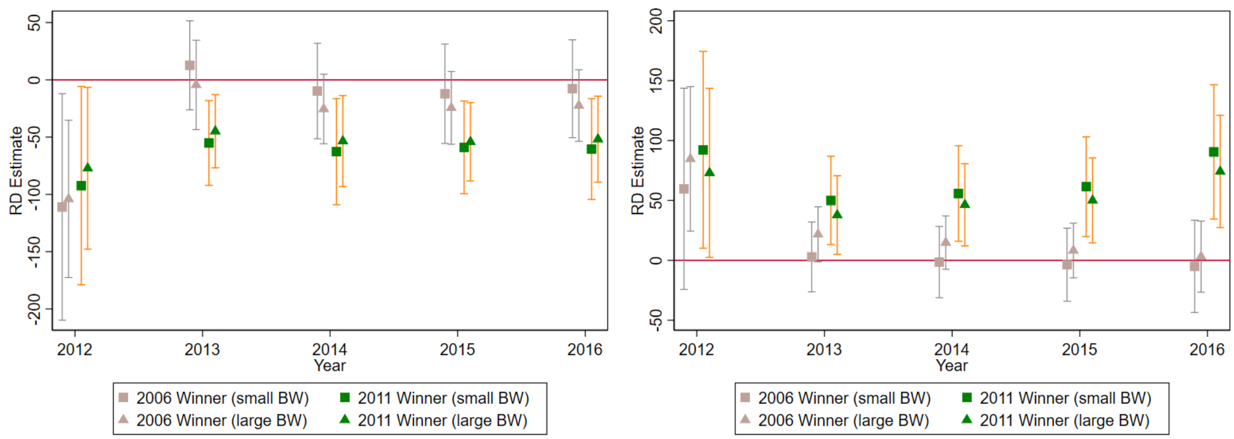


Notes: I plot RD coefficients across years for measures of data manipulation, and confidence intervals of robust standard errors clustered at the electrical-feeder level. Specifically, I study the bill items "total arrears" and "total subsidies". I find these result robust across bandwidths. 'BW' indicates the bandwidth size. The three bandwidths I use in these graphs are slightly lower and higher than the optimal bandwidth. These regressions control for the total size of the electorate within each assembly.

Following the analysis in Section 6.2, Figure A4 shows a similar pattern of no discontinuities using 2006 close-election assemblies. Again, there is weak evidence of a discontinuity in

[15]On speaking with the billing department at WBSEDCL, it was unclear to their IT officers how these variables were calculated, suggesting room for manipulation.

Figure A4: Placebo test: studying discontinuities in bill items (arrears and subsidies) using the winning and losing constituencies from the 2006 election

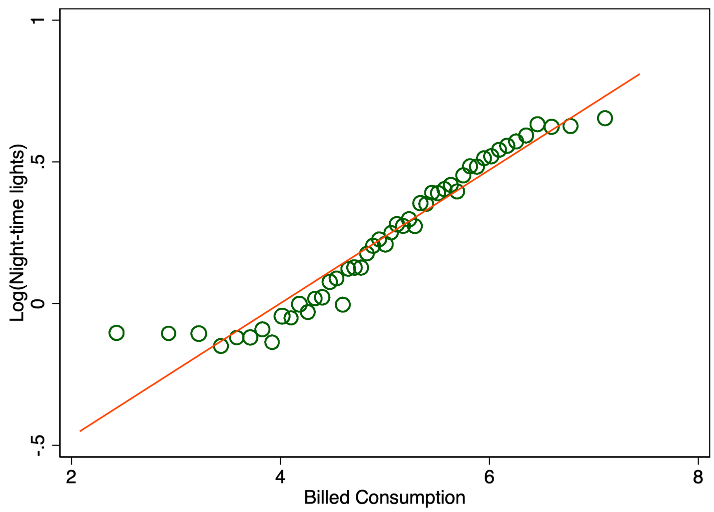


Notes: I plot RD coefficients for arrears and subsidies for the years 2012-2016. The winning margin here is defined for two elections: the 2006 election, where the CPI(M) party won and the 2011 election won by the AITC. This provides a placebo test for the validity of the results using the 2011 election results. The results shown include multiple bandwidths around the optimal bandwidth of +-6500 votes: BW 4500 votes to 8500 votes.

2012, immediately post the 2011 elections, suggesting possible persistence in manipulation from the previous ruling party. This points towards evidence that similar political influence in bill items occurred for assemblies where the previous ruling party won, and this effect peters out, as the actions of the current government, take over. These results provide a validity check for the main results of this paper and also point to possible evidence that politicians in power, across party lines, engage in actions to favor their constituents in terms of electricity access and price.

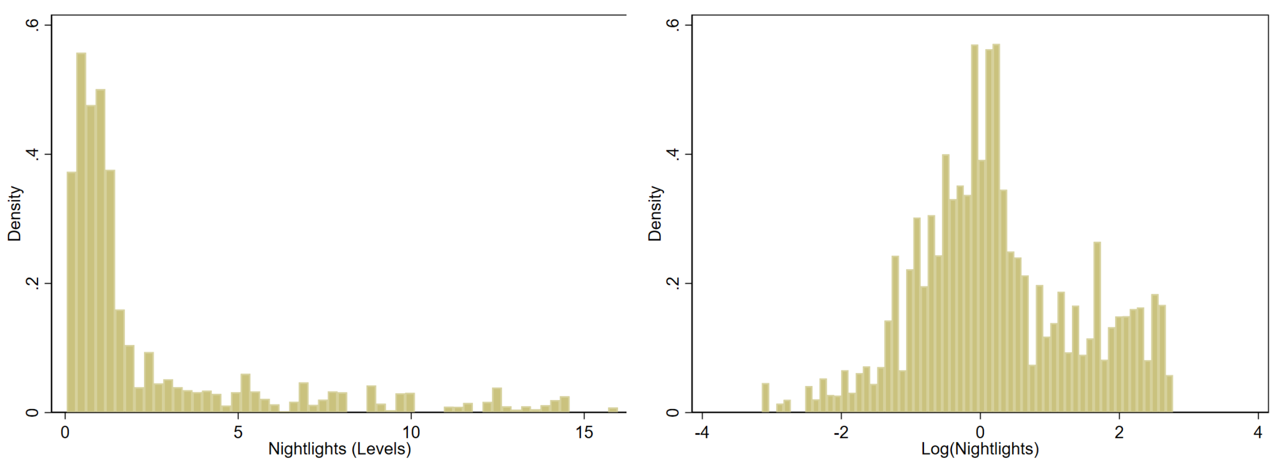
C Additional Tables and Figures

Figure A5: Binned Scatter Plot showing a strong linear correlation between VIIRS Satellite data and Electricity Consumption from Bills



Notes: I use log nighttime lights data from the VIIRS and plot it against log of electricity consumption from the administrative dataset for the years 2012-2016. The figure shows a binned scatter plot between the two variables and finds a strong linear relationship.

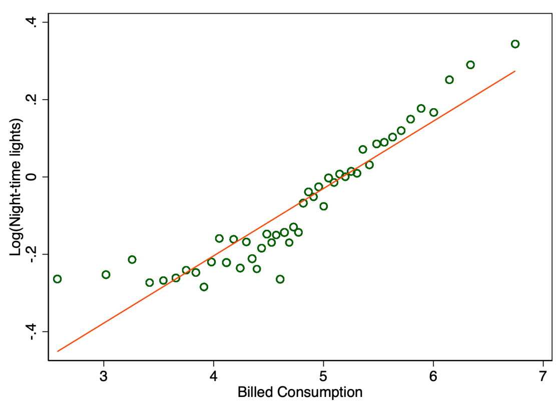
Figure A6: Distribution of VIIRS Satellite nighttime light data in levels and logs



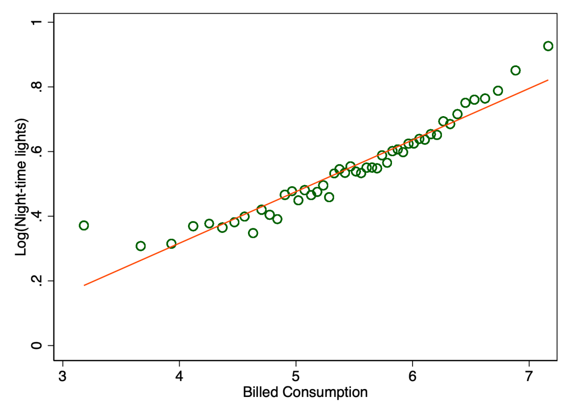
Notes: I use data from the VIIRS satellite and plot the data for 2012-2016 using two functional forms. In the left-hand side panel, I plot the distribution of nighttime lights data in levels, while I plot it on the right-hand side in logs.

Figure A7: Computing a multiplier to translate the change in log(lights) into log(consumption) for the the three main consumption categories

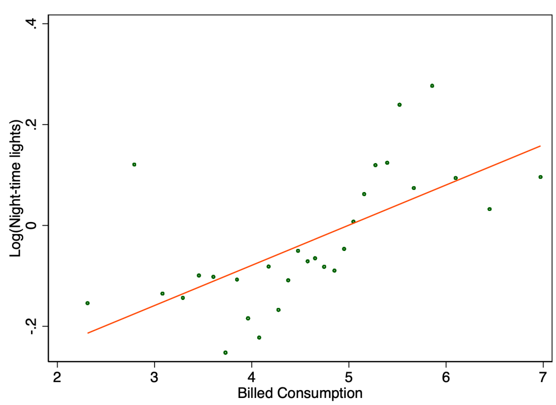
(a) Domestic(Rural)



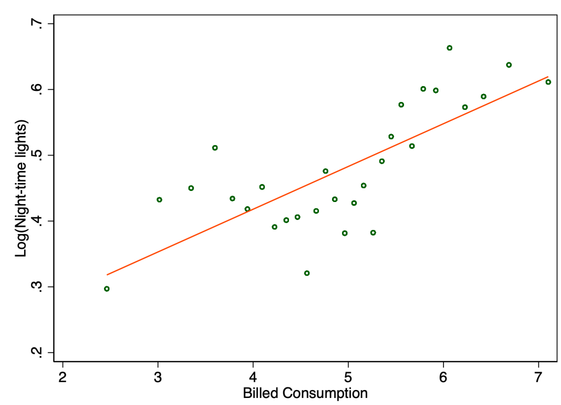
(b) Domestic(Urban)



(c) Domestic(Urban)



(d) Commercial (Urban)



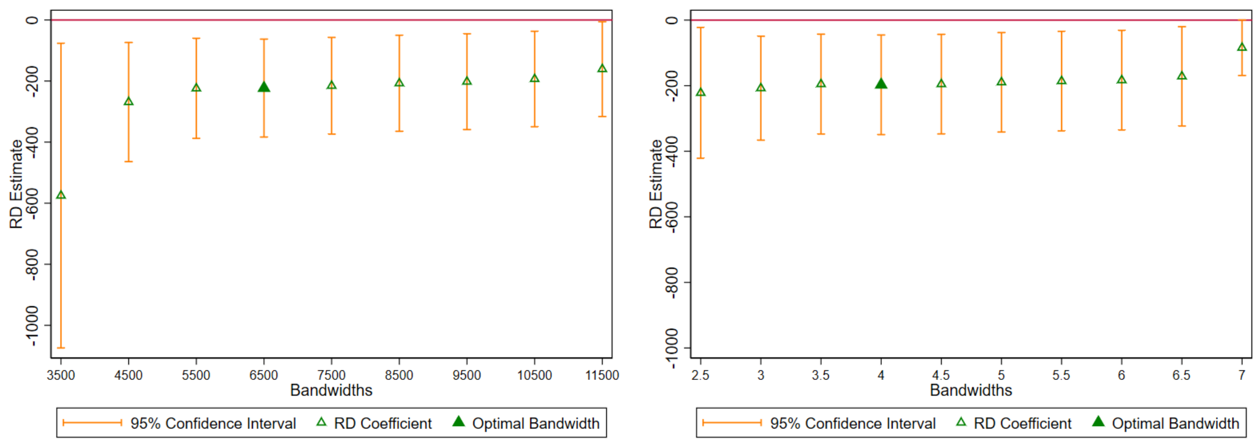
Notes: Each of the 4 panels separately plots log(lights) against log(billed consumption) for a given consumer category. Each of the graphs yields a linear multiplier to transform any change in log(lights) into KWH. The multipliers are 0.15, 0.09, 0.12, 0.05 for Residential (Rural), Residential (Urban), Commercial (Rural) and Commercial (Urban) accounts, respectively to translate 1 log point change in nighttime lights to log(KWh consumption).

D Robustness Checks and Sensitivity Analysis

D.1 Robustness to a wide range of bandwidths

Figure A8: Reported Consumption of Electricity: RD Estimates for smaller and larger bandwidths (vote margin share)

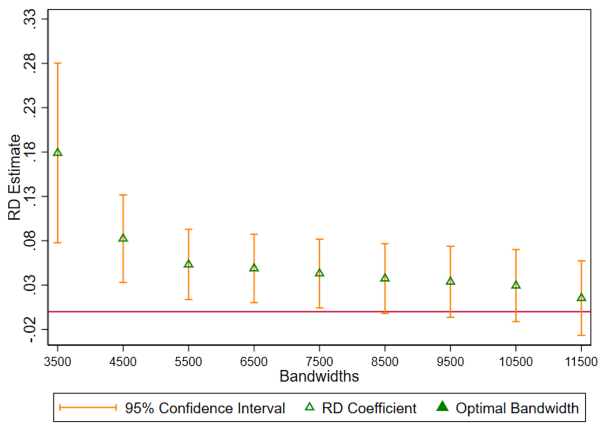
(a) BW: Number of votes (b) BW: Vote share margin



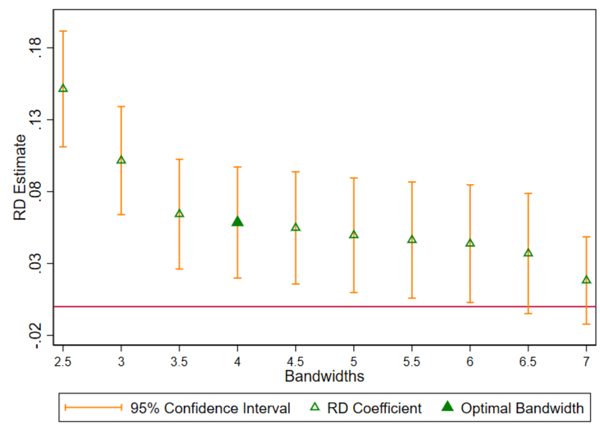
Notes: This figure plots the RD coefficient from regressing reported electricity consumption from administrative data on an indicator for whether the consumer is located in a constituency aligned with the ruling party. I plot this coefficient for a wide range of bandwidths, with the running variable being the number of votes by which the ruling party candidate won or lost elections.

Figure A9: Data Tampering: Prob(Bill Amount Multiple of 10) - RD Estimates for smaller and larger bandwidths (vote margin share)

(a) BW: Number of votes



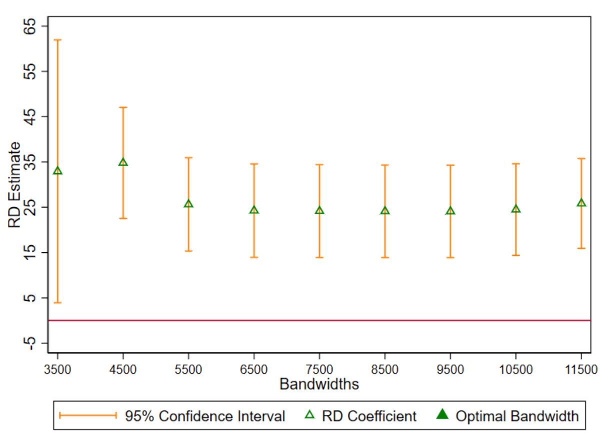
(b) BW: Vote share margin



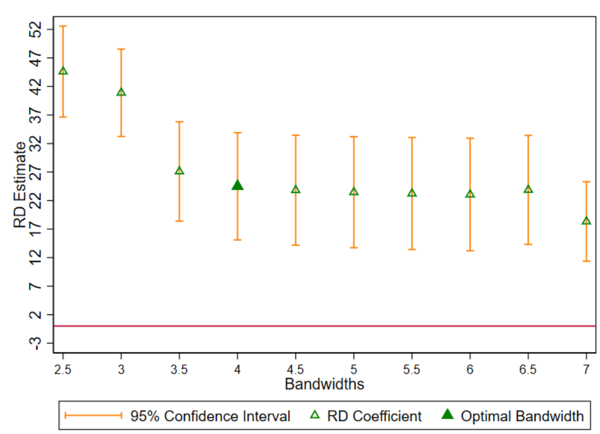
Notes: This figure plots the RD coefficient from regressing the fraction of bills that are a multiple of 10 from administrative data on an indicator for whether the consumer is located in a constituency aligned with the ruling party. I plot this coefficient for a wide range of bandwidths, with the running variable being the number of votes by which the ruling party candidate won or lost elections.

Figure A10: Measure of data manipulation: Benford's Law: RD Estimates for smaller and larger bandwidths (vote margin share)

(a) BW: Number of votes



(b) BW: Vote share margin



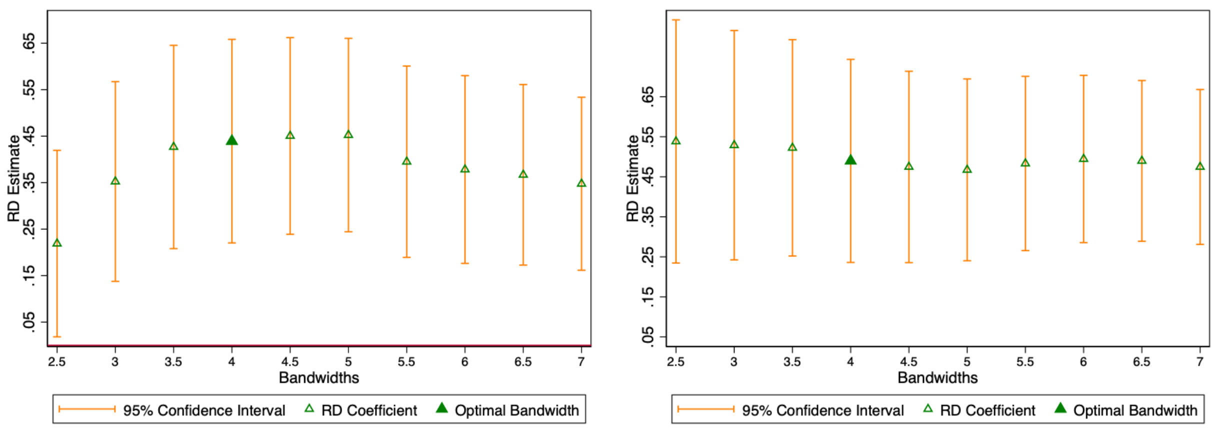
Notes: This figure plots the RD coefficient from regressing the measure of data manipulation based on Benford's Law from administrative data on an indicator for whether the consumer is located in a constituency aligned with the ruling party. I plot this coefficient for a wide range of bandwidths, with the running variable being the number of votes by which the ruling party candidate won or lost elections.

D.2 Robustness of results using log(lights)

Figure A11: Actual Consumption of Electricity: RD Estimates for smaller and larger bandwidths (vote margin percentage)

(a) West Bengal

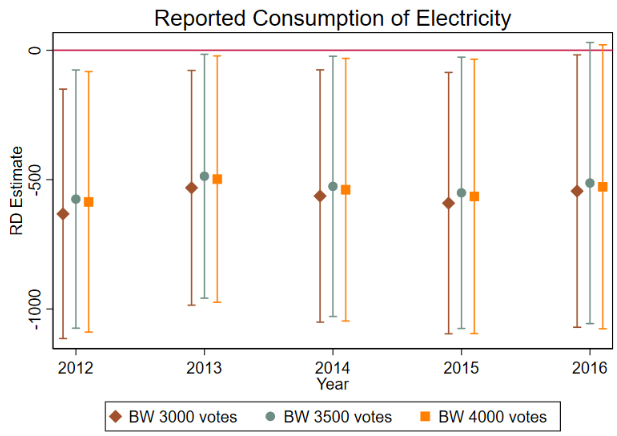
(b) All India



Notes: This figure plots the RD coefficient from regressing log(lights) on whether the consumer is located in a constituency aligned with the ruling party. I plot this coefficient for a wide range of bandwidths, with the running variable being the vote share with which the ruling party candidate won or lost elections. I mark the bandwidth of 4 pp as analogous to the optimal bandwidth used in the billing results.

D.3 Robustness to small bandwidths

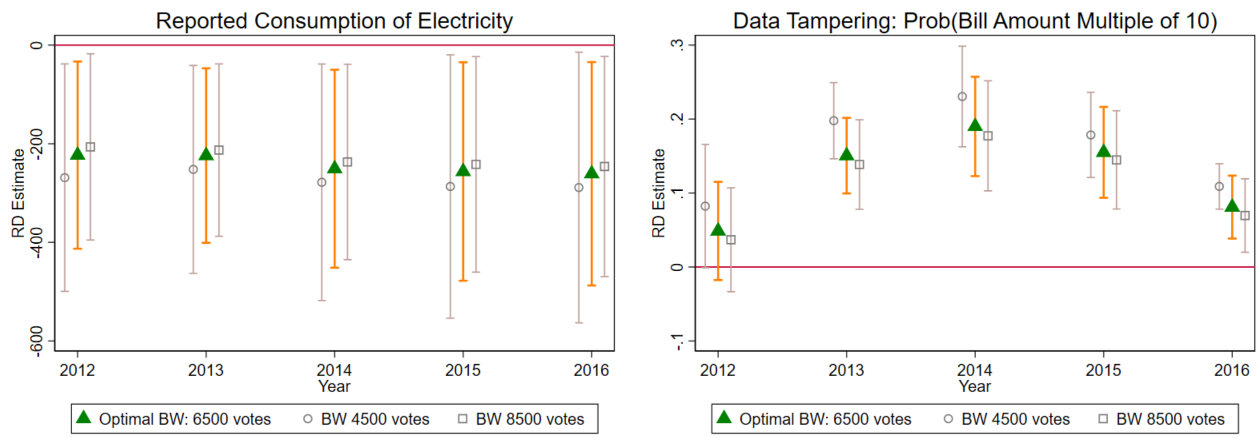
Figure A12: Reported Consumption of Electricity: RD estimates by year



Notes: In this figure, I plot the RD coefficients between 2012 and 2016, and find results robust to other bandwidths (in terms of the number of votes) between ± 3000 and ± 4000 votes, showing 95% confidence intervals. Here, a ± 3000 vote margin corresponds to between -1.87 and 1.6 percentage point, 3500 votes corresponds to between -1.9 and 1.82 percentage point, and 4000 votes correspond to between -2.36 and 2.4 percentage point vote margin for the ruling party. These bandwidths show the robustness of the main result to bandwidths smaller and larger than the optimal bandwidth. Standard errors clustered at the feeder level.

D.4 Robustness to clustering by assembly/constituency

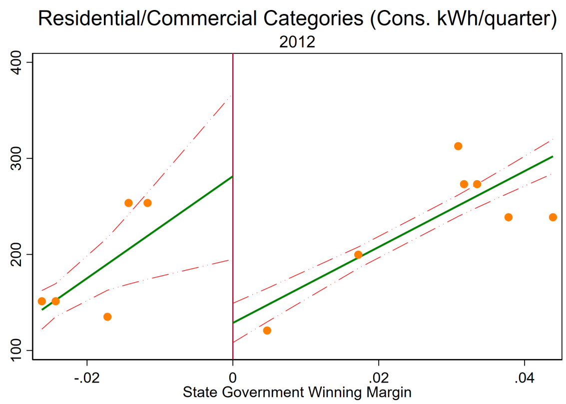
Figure A13: Lower reported consumption and higher measured data tampering in regions where the ruling party won (2012-16)



Notes: I plot the RD coefficients between 2012 and 2016 for two outcomes – reported consumption in the left panel and the probability of data manipulation in the right panel. I find results robust to a range of bandwidths between ± 4500 and ± 8500 votes by which the ruling party won elections, showing 90% confidence intervals. Standard errors are clustered at the assembly level. Here, a ± 4500 vote margin corresponds to between -2.74 and 3.18 percentage point, 6500 votes corresponds to between -3.88 and 4.45 percentage point, and 8500 votes correspond to between -5.77 and 6.31 percentage point vote margin for the ruling party. These bandwidths show the robustness of the main result to bandwidths smaller and larger than the optimal bandwidth.

D.5 Robustness to only using non-agricultural consumers

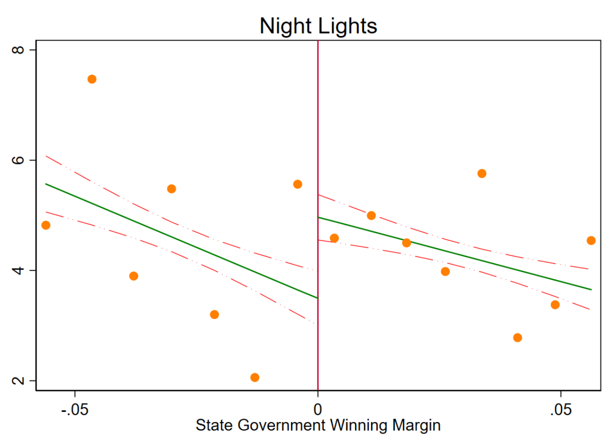
Figure A14: Lower Reported Consumption in Assemblies Aligned with the Ruling Party (non-agricultural)



Notes: This figure compares legislative assemblies where the ruling party narrowly won to those where it narrowly lost (2012-15). I find a discontinuously lower reported consumption from bills in assemblies aligned with the ruling party (excluding agricultural accounts).

D.6 Robustness of discontinuity in nighttime lights using levels

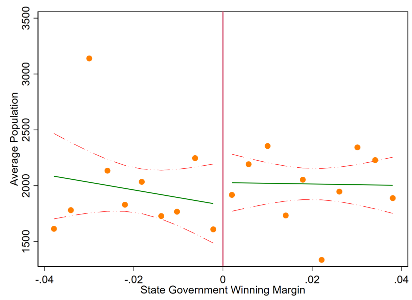
Figure A15: Robustness of RD results using nighttime lights data in levels for All-India



Notes: I present the RD plot showing discontinuously higher actual electricity consumption (using nighttime lights data in levels) in constituencies aligned with the ruling party for all of India from the years 2006-2016.

Figure A16: Balance Across RD Cutoff - Census Village-level Characteristics - All India

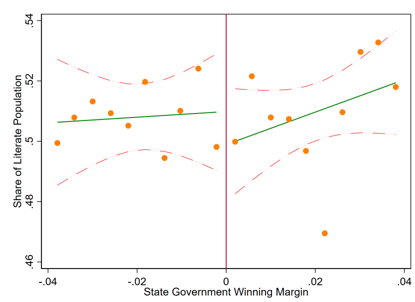
(a) Average Population



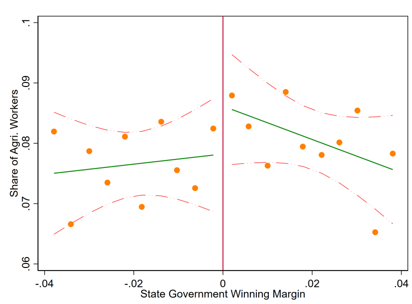
(b) Share of Scheduled Castes

(c) Share of Female Workers

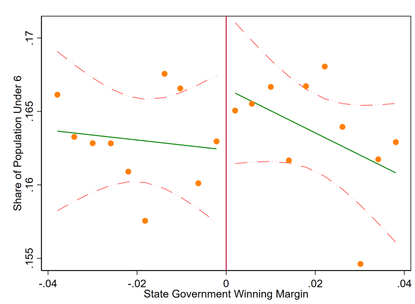
(d) Share of Literate Population



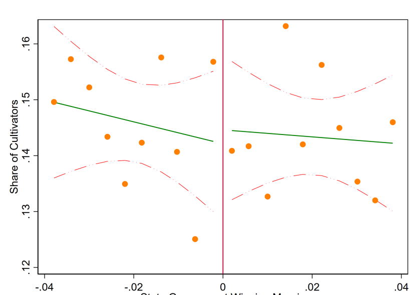
(e) Share of Agriculture Workers



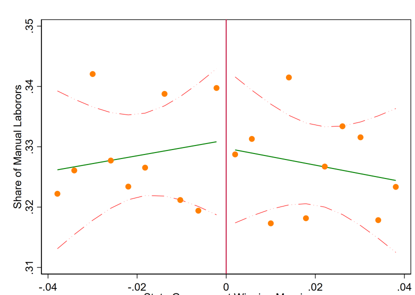
(f) Share of Population under 6



(g) Share of Cultivation Workers



(h) Share of Manufacturing Workers



Notes: I show balance in terms of village characteristics from the Indian Census 2011 across the RD cutoff for all of India. The running variable is vote margin share in state legislative elections.

E Elasticity Estimation

E.1 Step 1: Elasticities for Constituencies with low Data Anomalies

First, I restrict the data to only those assemblies where the distance from the expected chi-squared distribution is not significantly different from 0, at 1% confidence. This is the same measure I use to show evidence of data manipulation in Section 5.2. The micro-level billing data allows me to observe the distribution of consumption for each assembly and I separate these assemblies into those where there is evidence of data manipulation and those where there is no detectable evidence. This results in a dataset with 35 assemblies, for which I reject the hypothesis of data manipulation. For each assembly, I estimate the price elasticity of demand for each of the four consumer categories. The following specification, at the individual i and consumer category a level, is the simplest method of estimating elasticity but produces biased elasticities.

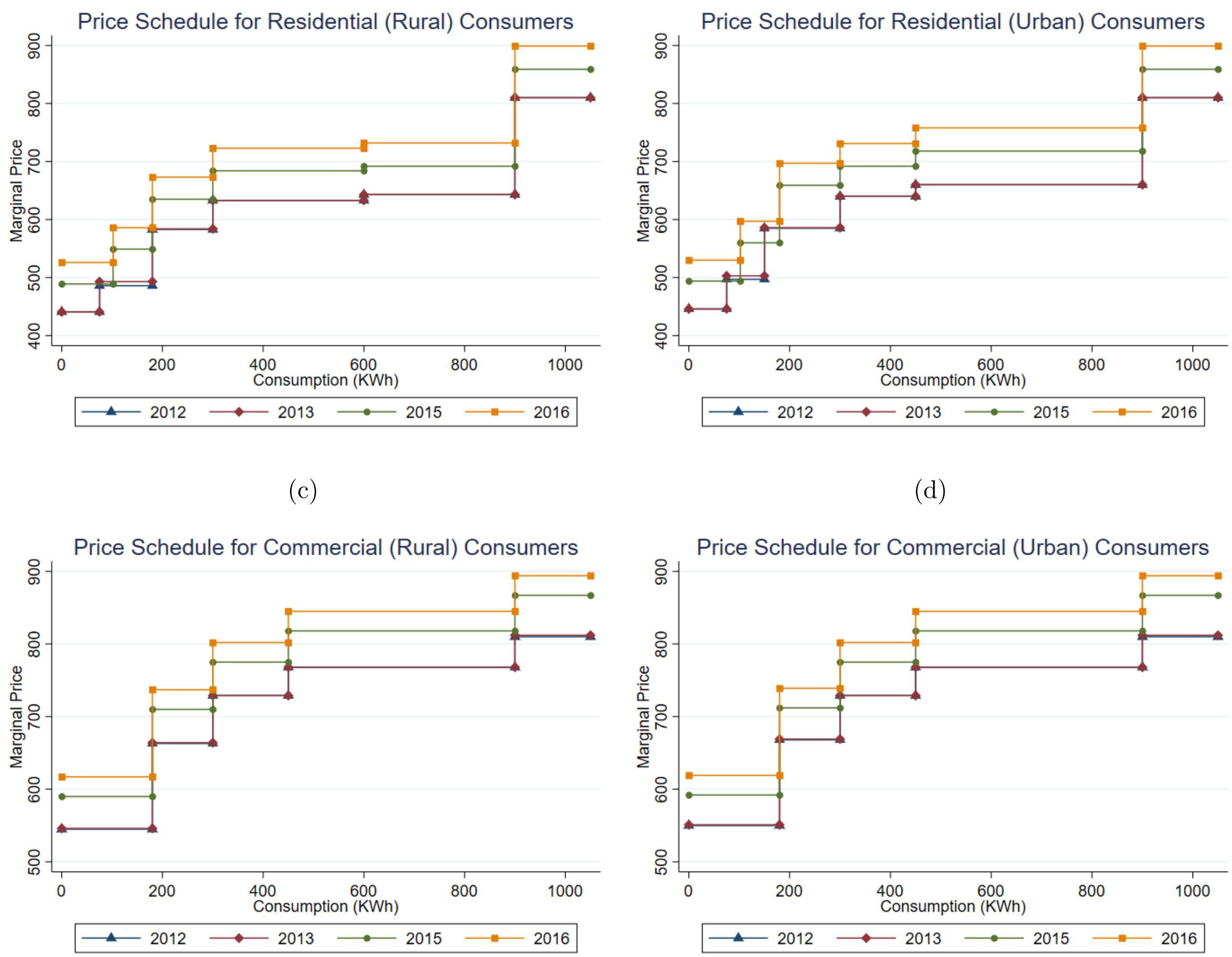
$$\log{(Consumption)}_{ia} = \delta_a \log{(MarginalPrice)}_{ia} + \epsilon_{ia} \tag{4}$$

Given the increasing block price tariff in electricity markets, a higher level of consumption mechanically results in a higher marginal price for higher levels of consumption, resulting in the estimate of δ_a suffering from a simultaneity bias.

In order to address the simultaneity bias arising from an OLS specification, I use an instrumental variable strategy, leveraging exogenous variation in the price schedules of electricity across time and for different consumer categories. With micro-level consumption data, I identify the price tier corresponding to the marginal level of electricity consumption of each consumer type, as well as their consumer category (rural/urban, domestic/commercial). The period for which I have consumption data (2011-2016) spans major tariff revisions, varying across tiers and consumer categories, and this provides me with policy-led, exogenous variation in price (Figure A17).

Figure A17: Change in Price Schedule Over Time

(a) (b)



Notes: The tables show the change in tariffs over time. These changes occurred in different months across different years. The price changes took effect in January 2012, February 2013, May 2015, and November 2016. The choice of instrumental variable in the elasticity estimation step is also prompted by the fact that prices sometimes changed uniformly across tiers. Therefore, instrumenting changes for levels leverages the price variation to greater effect.

My specification is similar to Ito (2014). I instrument the observed level of marginal price faced by a consumer with the policy-led change in marginal prices, in the spirit of Arellano and Bond (1991). I have five major different price regime periods, approximately one for every year of the data. I estimate elasticities for each assembly-by-consumer category, conditional on individual fixed effects, tier fixed effects, and month fixed effects. I instrument the marginal price $\log(MP)$ with the change in tariffs $\Delta \log(Tariff)$ across years. The first and second stage are respectively for an individual i, in tier t, month m, year y, assembly a, and consumer category c:

$$\log(MP)_{imty(ac)} = \gamma_{ac} \Delta \log(Tariff)_{mty(ac)} + \nu_{t(ac)} + \zeta_{my(ac)} + \eta_i + \varepsilon_{imty(ac)} \;\; \forall \;\; \{a,c\} \in A \times C \tag{5}$$

$$\log(Cons)_{imty(ac)} = \beta_{ac} \log(\widehat{MP})_{imty(ac)} + \tau_{t(ac)} + \mu_{my(ac)} + \omega_i + \epsilon_{imty(ac)} \;\; \forall \;\; \{a,c\} \in A \times C \tag{6}$$

I estimate β_{ac} separately for all constituencies-by-consumer categories {a,c} that lie in the set A of assemblies, and C consumer categories, for which I reject the hypothesis of data manipulation. The four consumer categories c are RR (Residential Rural), RU (Residential Urban), CR (Commercial Rural), and CU (Commercial Urban). The regressions include individual fixed effects ω_i, tier fixed effects $\tau_{t(ac)}$, and month fixed effects $\mu_{my(ac)}$. The advantage of having individual fixed effects is that it accounts for baseline consumption. The month fixed effects allow for seasonality (and time trends) in consumption to vary by assembly and consumer category. Standard errors are clustered at the consumer level.

Yet, there are a few underlying assumptions we must rely on, as with any such analysis: (a) there are no income effects, (b) the response is relatively immediate, (c) the short and long-term elasticity is similar, (d) the elasticity is similar across tiers (they are allowed to differ by consumer category and assembly), and (e) individuals have knowledge over the tariff changes.

Table A2 presents results by estimating a modified version of the specification in Equations 5 and 6 for all assemblies with unmanipulated data. Instead of estimating it separately by assembly and customer category, I jointly estimate it for the full non-manipulated sample, interacting the fixed effects with assembly and customer-category indicators. As such, this table is an example for what the elasticities may look like when estimated separately by assembly, and averaged. This table serves only to provide consolidated elasticities for the assemblies, but I estimate this specification separately for each assembly (and consumer category) in order to arrive at elasticity estimates for the prediction exercise. Overall, therefore, in assemblies that do not show evidence of data manipulation, residential consumers have less elastic demand, whereas commercial consumers (that may substitute to alternative sources like generators) have more elastic demand. The differences in elasticities between residential and commercial consumers, for both rural and urban consumers, are statistically different from zero. The high first-stage F-stat demonstrates instrument validity.

Table A2: Demand Elasticity Estimates for Select Regions

	Ln (Cons kWh)
$Ln(MP)_{RR}$ × Residential Rural	-0.240
	(0.293)
$Ln(MP)_{RU}$ × Residential Urban	-0.666**
	(0.310)
$Ln(MP)_{CR}$ × Commercial Rural	-3.158***
	(0.585)
$Ln(MP)_{CU}$ × Commercial Urban	-3.490***
	(0.588)
Observations	83,787
Customers	21,581
R-squared	0.424
P-val test Rural	0.000
P-val test Urban	0.000
F-stat	579.8

Notes: Ln(MP) is the log of marginal price. "Residential Rural" is an indicator for being in the residential-rural sector. Instruments are the change in Log(Marginal Price) for each of the four categories (Residential-Commercial by Rural-Urban). Standard errors clustered at the customer level. Controls include customer fixed effects, month-by-assembly-by-consumer category fixed effects, and tier-by-assembly-by-consumer category fixed effects, for each assembly-by-consumer category. P-val test Rural is the p-value of the test of equivalence of coefficients for the Residential Rural and Commercial Rural elasticities. P-val test Urban is the p-value of the test of coefficients for the Residential Urban and Commercial Urban elasticities.

E.2 Step 2: Predictive Model Selection Using Machine Learning

I use the estimates of assembly-level elasticities in the set A of non-manipulated assemblies and build a model of elasticity heterogeneity. The dependent variable in this model is assembly-level elasticity and the right-hand-side variables include demographic characteristics of assemblies from the 2011 Indian Census. These variables include the total population by gender, the population of Scheduled Castes and Scheduled Tribes (lower social classes and marginalized groups that are a proxy for income levels) by gender, the female literacy rate, and the population of cultivators (a proxy for occupation structures) in each village.

Each assembly has multiple Customer Care Centers (CCCs) set up by the utility and each individual is mapped to the CCC closest to them. As a first step, I map every single village in West Bengal, and assign it to the geographically closest CCC. Following this, I calculate CCC-level means of demographic variables by averaging the village-level aggregates assigned to each CCC. Therefore, each assembly in the dataset consists of 2-3 CCC-level observations with variation in characteristics.

I use the post-double-selection (PDS) method (Belloni et al., 2016) for variable selection. In the presence of several village-level characteristics, an issue with simply using OLS is that the predictive power of the model is compromised if there is omitted variable bias or if the model is overfit. For better out-of-sample predictions, an alternative model selection method is needed. I use the PDS-OLS method discussed in Ahrens et al. (2018); Belloni et al. (2016), which applies the LASSO (Least Absolute Shrinkage and Selection Operator) twice in order to select the set of variables that will maximize out-of-sample predictions. The LASSO is based on a penalized regression form, where shrinkage factors are applied to coefficients of independent variables based on relevance. It is particularly useful in conditions of sparse data, but with many possible independent variables. Applying the LASSO the first time eliminates covariates with the least predictive power, and running it a second time further strengthens model selection. Finally, this is followed by OLS using the limited set of variables selected by the PDS process, as OLS provides the least unbiased coefficient estimates.

In sum, the Census provides several village-level demographic characteristics, and the double-selection process whittles down the number of variables needed for predictive power. The OLS regression is then run (separately for each consumer category) to predict elasticities for all assemblies. Table A3 shows the final model used in the prediction step.

E.3 Step 3: Predicting Elasticities for all Constituencies

Following the PDS OLS method, I predict elasticities for constituencies that showed evidence of data manipulation. Table A4 shows the mean values of the resulting elasticities. These differ from Table A2 because they represent the mean elasticity for each consumer category taking into account *all* assemblies, those with unmanipulated as well as manipulated data.

The elasticity estimates in Table A4 improve upon the previous literature as I have consumer-level data. In most previous studies, estimates have been calculated from aggregate yearly consumption for an entire state, using averaged tariffs. With consumer-level data, I

Table A3: Predictive Model for Elasticity Projection

Independent Variables	Elasticity
Avg. no. of males under 6 yrs	-0.0122
	(0.170)
Avg. no. of females under 6 yrs	-0.000569
	(0.172)
Avg. no. of households	0.0106
	(0.0226)
Avg. no. of working males	-0.0126
	(0.0139)
Avg. no. of working females	0.0330**
	(0.0140)
Avg. no. of scheduled caste females	0.210**
	(0.0861)
Avg. no. of scheduled caste females	-0.197**
	(0.0814)
Avg. no. of scheduled tribe females	0.0153
	(0.0117)
Avg. no. of male cultivators	-0.0279**
	(0.0127)
Avg. no. of female cultivators	0.0339
	(0.0464)
Avg. no. of female workers (other)	0.00114
	(0.0416)
Avg. no. of literate females	-0.0156
	(0.0113)
Sq. of avg. no. of literate females	7.93e-06*
	(4.80e-06)
Constant	-50.99**
	(25.48)
Observations	43

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. This table shows results of the post-double OLS (Belloni et al., 2016) discussed in Section E.2. Census data provides several village-level demographic characteristics which I use to build a model in order to predict out-of-sample elasticities. The double-selection process whittles down the number of variables needed for predictive power. And the OLS regression is run and then used to predict elasticities for all assemblies.

Table A4: Average Demand Elasticities for Entire Consumer Base

Consumer Category	**Elasticity of Electricity Demand**
Residential (Rural)	-0.56
Residential (Urban)	-0.26
Commercial (Rural)	-2.94
Commercial (Urban)	-2.56

Notes: The price elasticities in this table are calculated using an instrumental variables strategy, prediction model selection procedure, and linear prediction model. The demand elasticities for each consumer class from Table A2 are regressed on CCC level characteristics, as described in this section. The coefficients from this regression are then used to predict the elasticities for all the regions where the data is manipulated. These are then combined to produce an average elasticity for each consumer category.

am able to observe the marginal price paid by the consumer, and the price tier that they consume in each month. Not having to aggregate across tiers allows me to use differences in the change in marginal price by tier. Aggregating prices and consumption across tiers may introduce measurement error, attenuating results. Furthermore, tariffs change within the same year, and annual data would need to aggregate tariff changes to the yearly level introducing further noise. This additional heterogeneity in tier and intra-year changes allows me to estimate more accurate elasticities.

Importantly, data that is manipulated will also suffer from measurement error when aggregated. My method allows me to estimate elasticities in regions where there was no evidence of manipulation, providing more robust elasticities. As a counterfactual exercise, I estimate the elasticities of the manipulated sample in Table A5. The results in column 1 of Table A5 confirm that the estimates run on the manipulated sample may suffer from attenuation bias due to classical measurement error. Lastly, the inclusion of individual fixed effects controls for baseline consumption at the individual level.

Price elasticity estimates, using aggregated and annual data, for residential consumers from previous work in India have yielded a range from -0.25 to -0.65, while those for commercial users have ranged from -0.26 to -0.49 (Bose and Shukla, 1999; Filippini and Pachauri, 2004; Saha and Bhattacharya, 2018). The average of the elasticity estimates for residential (rural and urban) consumers from my calculations yields -0.41, which is within this range, while my estimate for average elasticity for commercial (rural and urban) is -2.75, higher than previous estimates (Table A4). By estimating elasticities in only those regions where there was no evidence of manipulation, provides more precision and removes the biases in elasticity estimates.

One primary reason why observing bill-level data for Indian electricity consumers is important is that tariff changes are applied at non-standard times across the years. For instance, tariff changes were applied to bills in May 2013, February 2015, and November 2016, even as the tariff order by the regulator is usually released in December of the previous year. However, the aggregate electricity consumption published by the utility is calculated for every calendar year, and annual data then by construction is less informative about when changes occur.

One of the contributions of this work is to reflect the high elasticity of demand for commercial users in India. This is consistent with the fact that most commercial establishments in India have a kerosene or diesel generator, and therefore can substitute away from electricity if prices rise. Indeed, 46.5% of firms in India own a generator (The World Bank, 2014). The elasticity discussed in this paper is then the price elasticity of grid-purchased electricity. Consequently, this is reflected in their highly elastic demand response to price changes.

E.4 Estimating Elasticities - Counterfactual Exercise

Table A5: Alternative Ways of Calculating Price Elasticities

	Log(Consumption Kwh/Quarter)			
	IV 2SLS **Altered Sample**	OLS **Unaltered Sample**	IV 2SLS **Unaltered Sample**	IV 2SLS **Aggregated to AC Level**
Log Marginal Price Residential Rural	0.388*	1.609***	-0.240	-0.137
	(0.228)	(0.0596)	(0.293)	(0.0972)
Log Marginal Price Residential Urban	0.175	1.395***	-0.666**	-0.019
	(0.220)	(0.0574)	(0.310)	(0.0916)
Log Marginal Price Commercial Rural	-1.364**	0.583***	-3.158***	0.0628
	(0.535)	(0.130)	(0.585)	(0.155)
Log Marginal Price Commercial Urban	-1.800***	0.595***	-3.490***	-0.206
	(0.460)	(0.111)	(0.588)	(0.136)
Observations	120,087	106,937	83,787	13,943
R-squared	0.475	0.450	0.424	0.946
No. of Customers	30,906	21,980	21,581	
Fixed Effects	Month-Class Tier-Acc.	Month-Class Tier-Acc.	Month-Class Tier-Acc.	AC-Month Tier-Class
IV F-stat	704.2		579.8	414.6

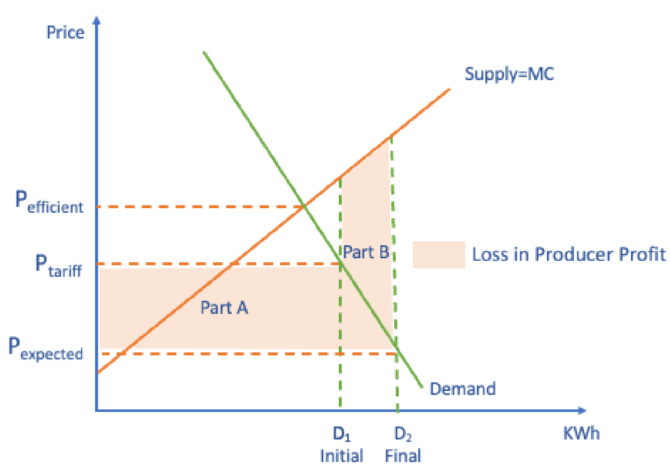
Notes: This table shows the importance of the four-step procedure to calculate welfare as in Section 7. Col 1 shows the elasticity estimates from the running the IV strategy in Table A2 on the manipulated sub-sample (Section 5.2). Col 3 follows Table A2, dealing only with the unmanipulated sub-sample of data, as I do in my welfare analysis. For residential consumers, col 1 show positive elasticities which go against theoretical foundations of demand. For commercial users, this column shows much lower elasticities than column 3. This is possibly because of using aggregated data that suffers from issues such as aggregation of price tariffs, using year-level consumption estimates, and manipulation. Col 4 shows the estimates obtained using aggregated data, like previous studies do. They are much lower than what I obtain even if I restrict the data to the unaltered sample.

F Welfare Calculations

F.1 Losses in Producer Profits

I compute the loss in producer profits by splitting up the area in Figure A18 into two parts: part A and part B. Part A is the loss in profits arising from the under-reporting of consumption, and the implicit subsidy it imposes. D1 initially may be thought of as the consumption in the absence of any subsidies – but because of under-reporting, rather than paying $P_{tarriff}$, the effective price of $D1$ is $P_{expected}$. The magnitude of part A is simply the area of the rectangle labeled Part A in Figure A18, and I compute it using the effective change in price derived from elasticities, shown in Table A7. Part B is the loss in profits arising from over-consumption in response to the implicit subsidy offered by politicians to the aligned assemblies. I use the utility's marginal cost of electricity to compute this component after estimating the scale of over-consumption per consumer group. The marginal cost of purchasing electricity, as reported by the utility WBSEDCL, remained fairly stable across the years studied in this paper, with modest increases in the period 2012-2016 (Sen and Biswas, 2017). I use the 2014 rate, which was the average value. Finally, I combine both pieces and compute an aggregate loss in producer profits in Table A6. I compute a net loss to producers of $1.8 billion.

Figure A18: Losses in Producer Profits



Notes: I first estimate the change in producer profits from Figure 10. I break it up into two parts: part A and part B. Part A is loss due to direct under-reporting, and part B is due to over-consumption.

Table A6: Calculation of Loss in Producer Profits

	Residential (Urban)	Residential (Rural)	Commercial (Urban)
Part A: Change in producer profit from under-reporting			
Initial cons (D1)	716	280	967
Change in price from elasticity P_{Tariff}-$P_{Expected}$ (Table A7)	0.85	0.58	0.06
Loss in profit per account (Rs. annualized)	2442	645	741
Part B: Change in producer profit from over-consumption			
Over-consumption (log(lights)) (Table 1)	0.44	0.44	0.44
Multiplier:Δlog(lights)$\rightarrow$ Δlog(cons) (Figure A7)	0.09	0.15	0.05
Δ(Log(cons))	0.04	0.07	0.02
% ΔCons (KWh) (%(D2-D1))	0.04	0.07	0.02
Loss in profit per account (Rs. annualized)	653	447	1323
Part C: Total loss in profit			
Accounts in AITC controlled areas (millions)	2.74	10.69	0.45
Total Loss (Part A+Part B)	3095	1092	2063
Scaled up by accounts (Mill. Rs)	8475	11677	937
Scaled up by accounts (Mill. $)	146	201	16
Total for electoral term (Mill $)	731	1007	81
Total (Mill $)		**1818**	

Notes: This table shows the aggregate loss in producer profits by combining two bits of evidence from this paper: that there is systematic under-reporting of consumption and an implied price subsidy as a result, and there is over-consumption as a result of the implicit subsidy. I combine both the administrative billing data and satellite nighttime lights data to arrive at these estimates. I compute ΔPrice=(exp(Δlog(price))-1)*P_{Tariff}, where I back out P_{Tariff} as the average price paid for a consumption of D_1 from the bills of consumers in losing assemblies. I compute Δlog(price)=Δlog(cons)/ϵ as shown in Table A7. I use an exchange rate of Rs. 58/$ (2014), and a wholesale marginal electricity purchase price of Rs. 5.70 (2014) as reported by WBSEDCL.

In Table A8, I present figures for losses in producer profit based on the previously established elasticity figures, in Saha and Bhattacharya (2018), which are lower than the estimates in this paper. However, because the loss to consumer profit is also lower by a similar amount, deadweight loss remains largely similar at $0.6 billion.

F.2 Gains in Consumer Surplus

Table A7 presents the calculations for gains to consumers, primarily through computing the area of the trapezoid for consumer surplus in Figure 10. In Table A9, I present figures for consumer surplus based on the previously established elasticity figures, in Saha and Bhattacharya (2018), which are significantly lower than the estimates in this paper for commercial urban consumers. However, this is a small proportion of the consumer base, deadweight loss is only slightly lower than before at a little over \$0.6 billion.

Table A7: Gain in Consumer Surplus from over-consumption (Elasticity from this paper

Consumer Class	**Domestic (Urban)**	**Domestic (Rural)**	**Commercial (Urban)**
Elasticity estimate (ϵ) (Table A4)	-0.265	-0.564	-2.550
Initial cons level (mean qtrly. cons (KWh) - losing assemblies)	716	280	967
Lights-Cons elasticity (Figure A7)	0.09	0.15	0.05
ΔLog(lights) (Table 1)	0.44	0.44	0.44
ΔLog(cons) (log(D2-D1))	0.04	0.07	0.02
ΔLog(price)=Δlog(cons)/ϵ	-0.16	-0.12	-0.01
ΔPrice (P_{Tariff}-$P_{Expected}$)	-0.85	-0.58	-0.06
ΔCS=0.5*(Initial+Final cons)*ΔPrice (Area of trapezoid in Fig 10)	623	167	62
No. of accounts (millions)	2.74	10.69	0.45
Total annualized gain in CS (mill. Rs.)	6832	7138	337
CS for electoral term \$ (Mill. \$)	**1173**		

Notes: This table shows the steps of calculating changes in consumer surplus in Figure 10. The table uses data from administrative billing records and satellite nighttime lights data to arrive at the estimates. I use an exchange rate of Rs. 58/\$ (2014). I compute ΔPrice=(exp(Δlog(price))-1)*P_{Tariff}.

F.3 Welfare using elasticity estimates from other work

This section reproduces the analysis done in Tables A7 and A6 using elasticity estimates from Saha and Bhattacharya (2018), which provide a lower bound on elasticity estimates. While lower elasticity estimates reduce both the estimates of producer and consumer surplus, the deadweight loss remains only slightly lower than before.

Table A8: Calculation of Loss in Producer Profits (Elasticity estimates from Saha and Bhattacharya (2018))

	Residential (Urban)	**Residential (Rural)**	**Commercial (Urban)**
Part A: Change in producer profit from under-reporting			
Initial cons (D1)	716	280	967
Change in price from elasticity P_{Tariff}-$P_{Expected}$ (Table A9)	0.36	0.50	0.60
Loss in profit per account (Rs. annualized)	1042	564	6984
Part B: Change in producer profit from over-consumption			
Over-consumption (log(lights)) (Table 1)	0.44	0.44	0.44
Multiplier:Δlog(lights)$\rightarrow$ Δlog(cons) (Figure A7)	0.09	0.15	0.05
Δ(Log(cons))	0.04	0.07	0.02
% ΔCons (KWh) (%(D2-D1))	0.04	0.07	0.02
Loss in profit per account (Rs. annualized)	653	447	1323
Part C: Total loss in profit			
Accounts in AITC controlled areas (millions)	2.74	10.69	0.45
Total Loss (Part A+Part B)	1694	1011	8307
Scaled up by accounts (Mill. Rs)	4639	10811	3774
Scaled up by accounts (Mill. $)	80	186	65
Total for electoral term (Mill $)	400	932	325
Total (Mill $)		**1657**	

Notes: This table shows the aggregate loss in producer profits by combining two bits of evidence from this paper: that there is systematic under-reporting of consumption and an implied price subsidy as a result, and there is over-consumption as a result of the implicit subsidy. I combine both the administrative billing data and satellite nighttime lights data to arrive at these estimates. I compute ΔPrice=(exp(Δlog(price))-1)*P_{Tariff}, where I back out P_{Tariff} as the average price paid for a consumption of D_1 from the bills of consumers in losing assemblies. I compute Δlog(price)=Δlog(cons)/ϵ. I use an exchange rate of Rs. 58/$ (2014), and a wholesale marginal electricity purchase price of Rs. 5.70 (2014), as reported by WBSEDCL. I use elasticity estimates (ϵ) from Saha and Bhattacharya (2018) to provide a lower bound on the change in producer surplus.

Table A9: Calculation of Gain in Consumer Surplus from over-consumption (Elasticity estimates from Saha and Bhattacharya (2018))

Consumer Class	**Domestic (Urban)**	**Domestic (Rural)**	**Commercial (Urban)**
Elasticity estimate (ϵ) (Table A4)	-0.650	-0.650	-0.260
Initial cons level (mean qtrly. cons (KWh) - losing assemblies)	716	280	967
Lights-Cons elasticity (Figure A7)	0.09	0.15	0.05
ΔLog(lights) (Table 1)	0.44	0.44	0.44
ΔLog(cons) (log(D2-D1))	0.04	0.07	0.02
ΔLog(price)=Δlog(cons)/ϵ	-0.06	-0.10	-0.09
ΔPrice (P_{Tariff}-$P_{Expected}$)	-0.36	-0.50	-0.60
ΔCS=0.5*(Initial+Final cons)*ΔPrice (Area of trapezoid in Fig 10)	266	146	589
No. of accounts (millions)	2.74	10.69	0.45
Total annualized gain in CS (mill. Rs.)	2914	6242	3179
CS for electoral term $ (Mill. $)	**1011**		

Notes: This table shows the steps of calculating changes in consumer surplus in Figure 10 using elasticity estimates from Saha and Bhattacharya (2018), which yield a smaller gain in consumer surplus than in Table A7. The table uses data from administrative billing records and satellite nighttime lights data to arrive at the estimates. I use an exchange rate of Rs. 58/$ (2014). I compute ΔPrice=(exp(Δlog(price))-1)*P_{Tariff}.