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

Meera Mahadevan*

September 20, 2019
Click for Latest Version

Abstract

Political capture of public electricity provision may benefit targeted consumers through informal subsidies. However, this causes leakages in utility revenues, inhibiting their ability to reliably supply electricity to the broader consumer base. Using a close-election regression discontinuity design, and a confidential dataset on the universe of geo-coded electricity bills from a large state in India, I show that *billed* electricity consumption is lower for constituencies of the winning party after an election. However, *actual* consumption, as measured by satellite nighttime lights, is higher for these regions. I find new evidence to explain this discrepancy – politicians illicitly subsidize their constituents by systematically allowing the manipulation of electricity bills. To measure changes in welfare, I develop a method to estimate demand elasticities in the presence of data manipulation, by leveraging exogenous variation from policy-led price changes and predictive analytics. The net deadweight loss I estimate is large enough to power 3.7 million rural households over an electoral term.

JEL: Q41, Q48, O13, O17, P48

Keywords: Electricity, political economy, corruption, consumer welfare, India

*Department of Economics, University of Michigan, meeram@umich.edu, meeramahadevan.com. I am grateful to Ryan Kellogg, Dean Yang, Hoyt Bleakley and Catherine Hausman for insightful advice and guidance, and Achyuta Adhvaryu, Prashant Bharadwaj, Gaurav Khanna, Brian Min, Shaun McRae, Robyn Meeks, Faraz Usmani, Jeffrey Weaver and Pinghui Wu for invaluable suggestions. Thanks to the seminar participants at Michigan, Energy Fest 2018 (Northwestern), WCERE 2018 (Gothenburg), Summer Workshop 2018 (UCSD), CU Environmental & Resource Economics Workshop 2018 (Vail), Canadian Resource and Environmental Economics Workshop 2018 (Montreal), the Empirical Methods in Economics Conference 2018 (Northwestern), Development Day 2019 (U Chicago), PACDEV (2019), SETI Workshop (2019), AERE Summer Conference (2019), Rutgers University, University of Melbourne, University of California (Irvine), University of Western Ontario, American University, National University of Singapore for useful comments, and to the Michigan Institute for Teaching and Research in Economics (MITRE) for research funding.

1 Introduction

Electricity is critical for the production of goods and the functioning of households. Despite its economic significance, supply in developing countries is plagued with poor infrastructure (Ryan, 2014), a lack of investment (McRae, 2015), frequent blackouts, and voltage fluctuations (Allcott et al., 2016). Generation has increased exponentially in countries like India (Central Electricity Authority, 2018), but state-owned providers continue to lose almost 23% of generated electricity to technical and commercial (billing) leakages (Ministry of Power, 2018). In democracies, politicians may contribute to these leakages by selectively rewarding constituents with subsidized electricity. While low prices or reduced supply shortages benefit targeted consumers, the losses to utilities may lead to future outages and intermittent supply to the detriment of many. In what ways do politicians systematically cause market distortions in public electricity provision, and how do such interventions affect the welfare of consumers? That is the focus of this paper.

It is challenging to rigorously verify the anecdotal evidence and press reports on political corruption. Many reports show that politicians turn a blind eye to energy theft committed by voters or allies (The Telegraph, 2014; The Washington Post, 2012), tacitly support violence against officials who clamp down on energy theft (The Times of India, 2017), and are directly involved in violent retaliation against such officials (Hindustan Times, 2016).[1] A clear understanding of the mechanisms and magnitudes of corruption is crucial to designing future policy in electricity provision. Yet, in order to systematically investigate these allegations, and obtain a sense of the magnitudes involved, we need to derive well-identified relationships from administrative data.

I overcome each of these challenges in my work. First, I use satellite nighttime lights, and confidential quarterly billing records for 17 million accounts from an electricity utility, to show that electricity is used as a tool to garner votes. Leveraging a close election Regression Discontinuity Design (RDD), I show causal evidence that regions aligned with the governing party receive indirect electricity subsidies. Second, I identify the mechanisms of corruption through which politicians use electricity supply and prices as a form of patronage. Importantly, I use the micro data to investigate how the distribution of reported consumption

[1] "Vote-hungry local politicians protect the thieves....At its worst, Indias power sector is the perfect example of populism and patronage trumping sound economics, analysts say." The Washington Post (2012) *Power Thieves Prosper in Indias Patronage-based Democracy.*

changes systematically with political alignment. Finally, I estimate the magnitudes of the welfare consequences both for consumers and providers, this time by exploiting exogenous changes in prices to estimate the elasticity of electricity demand.

I causally identify widespread corruption in electricity billing practices. I highlight efforts by incumbent politicians to favor their voters post-election, by under-reporting their consumption in electricity bills. I find that shortly after a state-level election, there is an increase in electricity consumption, as measured by satellite nighttime lights data, for regions that voted for the winning party. Alone, this evidence appears to indicate selectively higher levels of electricity access for these regions, possibly owing to politicians redirecting electricity. These same regions, however, have discontinuously lower levels of *billed* consumption, as reported by the electricity provider. This billing evidence alone, may suggest that politicians instead redirect electricity to regions where they lost elections, possibly in a bid to win over voters. Together, however, the evidence from the nighttime lights and billing data paint a different picture: Politicians engage in patronage towards their constituencies by under-reporting electricity consumption, even as their constituents consume higher actual amounts of power. The magnitude of under-reporting is large, constituting a discount greater than 40% of billed consumption for consumers at the cutoff.

Yet, the mere existence of corrupt practices does not inform the types of policy regulations required to stem them if there is little understanding of the mechanisms. I highlight the mechanisms behind under-reported bills by identifying how anomalies in the consumption distribution change at the RD cutoff. I show evidence that billing data is more likely to be manipulated in areas aligned with the ruling party. First, there is a greater divergence in the observed distribution from what would be expected, as predicted by Benford's (1938) Law, which is commonly used to detect data fraud.[2] Secondly, I observe that a discontinuously higher number of bills in the winning party's constituencies are multiples of ten, reporting consumption amounts such as 20, 30, and 40 units. These results are consistent with a patronage hypothesis where local, incumbent politicians reward voters post-election and consolidate power by allowing the manipulation of actual consumption to appear lower than what was consumed.

These corrupt practices, however, tell us little about the size of the economic losses. The

[2]Benford's (1938) Law predicts a frequency distribution of the first digit of naturally occurring, unmanipulated sets of numerical data, such as consumption data.

magnitudes of the welfare gains to consumers, and the deadweight loss are important to measure as they determine the distributive consequences and policy urgency of the problem. To identify the welfare implications on each of the parties involved, I measure both the gains in consumer surplus from receiving subsidized electricity, and the lost revenue to the provider due to under-reported bills. I estimate the size of the loss in producer surplus from RD estimates of under-reported bills. The magnitude of change in consumer surplus, however, requires computing the price elasticity of electricity demand. The estimation of these elasticities for all consumers is challenging as there is clear evidence of data manipulation in the bills. I, therefore, develop a method for calculating price elasticities of electricity demand in the presence of data manipulation.

I leverage policy-changes in tariffs and predictive analytic techniques in order to estimate price elasticites of demand. I first divide the data into two sets – the set of regions where the data is plausibly unmanipulated by political influence, and the set that contains under-reported bills. The regions for which I reject the hypothesis of data manipulation span both constituencies of the winning party as well as those that voted for the opposition. I estimate the price elasticity of electricity demand for the regions with unmanipulated data, exploiting exogenous variation in policy-led electricity prices as instrumental variables, similar to Ito (2014). To elaborate, I use variation in the electricity price schedule (set by independent regulators) over time, across consumer categories, and across different tiers of the price schedule, to determine consumption responses to changes in the marginal price. One advantage over previous studies is that with micro-level data, I can do this for each consumer category while still having sufficient statistical power to estimate demand elasticities.

I then use machine learning methods, specifically the post-double-selection Ordinary Least Squares (PDS OLS) procedure (Ahrens et al., 2018; Belloni et al., 2016), to build a predictive model of regional elasticities based on census village-level demographics. Using OLS alone may lead to biases arising from overfitting or omitted-variables. Compared to previous work, I show that ignoring data manipulation, using aggregated data, or not leveraging policy instruments, leads to biased estimates of demand elasticities.

Using the estimated under-reporting in consumption, and elasticities, I find that the loss to the electricity provider ($57 million) outweighs the gain in consumer surplus ($22 million) for regions near the RD cutoff.[3] Simple calculations show that the net welfare loss of almost

[3]Even a figure as high as $57 million is ultimately an underestimate of the overall losses as I restrict the

$34 million is sufficient to power 3.7 million rural households. Yet, this may not capture the full extent of the welfare costs, as such political favoritism could exacerbate inequalities as well (Asher and Novosad, 2017; Burgess et al., 2015; Fisman, 2001).

I develop a political patronage model based on a combination of features present in Stromberg (2004), and include consumer decisions to highlight the importance of price elasticities in such a setup. Consistent with the model, I find that politicians target consumer categories that have inelastic demand, and groups with greater access to electricity-using infrastructure. Indeed, there is little preferential targeting for commercial rural consumers who have both sufficiently elastic demand and less electricity-using infrastructure.[4]

In theory, politicians may be able to target basic services to their voters who need it the most, increasing their consumer surplus (Brender and Drazen, 2005). Indeed, democracy could play an important role in ensuring the efficient allocation of government inputs in an effort to garner votes (Burgess et al., 2015). However, it could also result in misallocation (Khwaja and Mian, 2005), electoral cycles (Cole, 2009) or preferential access (Asher and Novosad, 2017). Such politically-motivated market distortions impose a burden on the provider and other consumer groups, simply exacerbating the already poor quality of electricity supply in several regions. The welfare effect of political involvement in electricity is therefore, theoretically ambiguous. A contribution of my work is to resolve this ambiguity by estimating the magnitude of the welfare costs and benefits. I find that in the case of Indian electricity, the producer loss outweighs the gains in consumer surplus by more than 2:1. This speaks to current debates in several parts of the world about the privatization of electricity provision, the merits of better metering and questions of equitable access. Furthermore, such political favoritism is not only limited to developing countries (Albouy, 2013; Ansolabehere and Snyder, 2007), and given its wide prevalence, it is important to understand the features that allow political patronage to take place, in order to design policies to counter it.

analysis to constituencies within the RD bandwidth. While the RDD provides a causal claim of political patronage for constituencies near the cutoff, it does not rule out the possibility of patronage in constituencies where the ruling party won with a larger margin. My estimates are about 40% of the losses to the provider in 2016 as measured by the Comptroller and Auditor General of India. (Business Standard, 2018) reports that the provider "suffered a revenue loss of Rs 175.85 crore owing to delay in raising bills, theft of electricity and unauthorised use."

[4] Politicians may also target consumers with greater influence politically, and urban consumers are more likely to have such influence. This is consistent with studies such as Badiani et al. (2012) who show evidence of politicians wooing rich and influential farmers by guaranteeing free or cheap electricity.

Importantly, I identify mechanisms of political patronage, such as manipulation of billing records, through which patronage can occur despite the lack of direct political control over electricity prices. Prior research finds suggestive evidence that politicians increase electricity supply before elections, to sway voters (Baskaran et al., 2015; Min and Golden, 2014), and they pressure authorities to keep tariffs low (Chatterjee, 2018; Millennium Post, 2017; The Economic Times, 2015). However, due to the unavailability of micro-level electricity consumption data, past studies often rely on satellite lights or regional aggregates as proxies for electricity consumption and leakages. Such macro-level aggregates, while informative, conceal the underlying corrupt practices, and are limited in identifying the net consequences of such actions. For instance, I show evidence of higher electricity consumption in regions supporting the ruling party, based on satellite data. However, this is only half the story. By uncovering widespread corruption in the micro-data, my findings stress the importance of having both measures of actual electricity consumption (satellite nighttime lights) and reported electricity consumption (billing data), to reveal the method of corruption. The mechanisms of corruption are now revealed to be through billing-data manipulation, which constitutes an indirect subsidy and may lead to over-consumption.

I take advantage of the monthly or quarterly billing data to study corruption in a way that was not feasible in previous work. First, it allows me to estimate price elasticities after accounting for an individual's consumption tier, and intra-year changes in tariffs. This was not feasible in previous studies relying on aggregated data at year level (Saha and Bhattacharya, 2018). Second, and more importantly, it allows me to estimate the entire consumption distribution at the regional level to test for data manipulation – a feature absent from analyses that rely on aggregate data. I can therefore directly measure under-reporting and estimate the corresponding shortfall in the revenues of the electricity provider. This would not be possible with satellite data alone, and helps me elicit the true welfare costs and benefits of political interference.

The rest of this paper is laid out as follows. Section 2 provides background information on the Indian electricity sector, and the surrounding institutional and political structures. Section 3 discusses the empirical strategy for identifying evidence of political corruption in electricity billing. The next section, Section 4, describes the data sets used in this analysis. I show evidence of corruption in Section 5. Section 6 discusses the welfare implications of corruption by politicians, and Section 7 concludes.

2 The Electricity Sector in India

Electricity supply is a critical issue in India, 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), and both the price and quality remain important election issues (Chatterjee, 2018).

In this paper, I focus on West Bengal, a large Indian state where the transmission and distribution sectors are state-owned.[5] 55% 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 millions accounts. In 2003, the central Electricity Act reforms created a state regulatory commission, responsible for setting electricity tariffs and overseeing the functioning of the utility. This particular provision was made specifically to separate the control of the electricity sector from increasing political influence. I analyze whether such mandates are sufficient to enforce political separation in reality, given weak enforcement and auditing mechanisms. 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.

Whether political interference in electricity occurs depends on the incentives faced by politicians, and whether such influence is feasible. There are a number of reasons why politicians may want to control electricity supply. Election surveys in India find that electricity is a key factor in election platforms (Chhibber et al., 2004). While politicians may try to win over new voters by offering cheaper or better access to electricity, there is a well documented pattern of patronage politics (Min, 2015; Nagavarapu and Sekhri, 2014; Sadanandan, 2012) in India, with politicians exerting great effort in consolidating existing votes.

Chatterjee (2018) presents evidence consistent with my model of politicians exerting ef-

[5] With the exception of one privately owned firm which distributes only to the city of Kolkata.

fort to provide cheaper electricity. Interviews with regulatory officials show pressure from politicians in the ruling party to delay or avoid upward revisions in tariffs. Regulators report resisting these attempts, demonstrating the difficulty faced by politicians in directly influencing the price of electricity. This arguably leads politicians to explore other, more indirect means of affecting electricity access and tariffs. Examples of such methods include politicians implicitly allowing energy theft among their constituents (The Telegraph, 2014; The Times of India, 2018; The Washington Post, 2012).[6] Golden and Min (2011) demonstrate how electricity bills are more likely to go unpaid in areas where criminals have political affiliations. Another possible channel is through the middle-men involved in the bill collection process. External inspectors are hired on a contract basis to conduct manual meter readings. Rains and Abraham (2018) highlights the often overlooked policy issue of low revenue collection, due to poor incentives for these contractors. Finally, politicians could selectively encourage lower enforcement of revenue collection in their constituencies, allowing billing centers to make lower bill imputations and under-charge their constituents.

One factor helping governing parties is that while the electricity provider remains state-owned, politicians themselves are not held accountable for its functioning. In several states, electricity distributors have faced mounting losses for several years. This cycle of losses is virtually systematized by the setup of a centrally managed bailout program, Ujwal Discom Assurance Yojna (UDAY), launched in 2015 to help loss-making electricity utilities recover financially. In practice, politicians do not pay any penalty for their state utility making such losses, whereas checks-and-balances that would prevent them from interfering with utility functioning are minimal. In such an environment, state politicians have an incentive to 'informally' provide their voters with access to cheaper and more electricity, following a long tradition of patronage politics in India. The empirical portion of this paper shows evidence of the mechanisms through which politicians provide informal or indirect subsidies.

2.1 Theoretical Predictions

I develop a model in Appendix A to generate testable implications, derive estimation equations, and motivate my welfare analysis. First, I derive a standard equation for electricity

[6] "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).

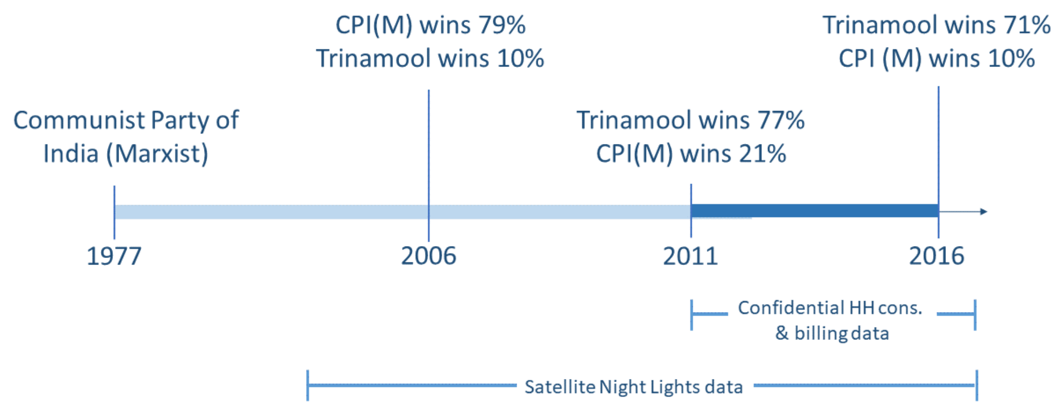
demand given a simple quasilinear utility function, increasing in electricity consumption with a constant elasticity for demand. Access to electricity-using infrastructure also shifts out the demand for electricity, and these consumers vote for politicians that give them higher utility. Second politicians exert effort and influence over utility providers to maximize their probability of winning the next election. Exerting effort comes at a cost, which prevents politicians from indiscriminately targeting all voters. These costs are lower in areas where politicians are in power and aligned with the state government.

This simple set-up allows me to derive testable implications. First, politicians exert more effort and influence in areas where local leaders are aligned with the state government. I measure this influence by looking at evidence on systemic under-reporting of consumption. Second, electricity subsidies and actual consumption (as measured by satellite data) are higher in such areas. Third, politicians target consumer bases with relatively more inelastic demand as they stand the most to gain from informal subsidies. The model also allows me to reproduce standard equation for estimating the price elasticity of demand for different types of consumers, and to test whether politicians do indeed target more inelastic consumers. Fourth, politicians target consumers with access to more electricity-using infrastructure, such as consumers in urban areas. Last, as in standard models, the change in consumer surplus is a simple function of the elasticity of demand. As I show in Appendix A these predictions motivate using, and are testable in a simple RD set up.

3 Close-election Regression Discontinuity Design

I apply a close-election Regression Discontinuity (RD) framework to identify whether politicians in West Bengal informally subsidize electricity. In India, state elections occur every five years and follow a parliamentary style format. States are composed of legislative assembly constituencies (in short, assemblies). The voting population elects constituency level representatives or Members of Legislative Assembly (MLAs), and the political party with the majority of MLAs forms the government. The head of the winning party becomes the Chief Minister of the State.

Figure 1: A timeline of the winners of the state elections in West Bengal from 1977 to 2016



Notes: CPI(M) is the Communist Party of India (Marxist), and Trinamool is the All India Trinamool Congress (AITC) party. These are the two rival parties in West Bengal.

I use the winning margin percentage in assembly elections as the running variable for the RD. I compare outcomes just above and below a zero winning margin RD cutoff to estimate the Local Average Treatment Effect (LATE) of being in a constituency aligned with the ruling government. The winning margin percentage is the fraction of votes by which an MLA from the ruling party wins an assembly election. Asher and Novosad (2017); Bardhan and Mookherjee (2010) and Nagavarapu and Sekhri (2014) use similar close election RDs in India-specific contexts. Constituency level elections in India are competitive, unpredictable and several factors affect their outcomes. Therefore, despite widespread political patronage, the probability of a constituency lying near the RD cutoff is randomly determined in an election. Given the unpredictability of these local elections, particularly in regions close to the RD cutoff, the close election RD is especially valid in this case (Eggers et al., 2015).

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

In the 2011 state elections, the All India Trinamool Congress (AITC) defeated the incumbent Communist Party of India – Marxist (CPI(M)) in a landslide election (Figure 1). Prior to the election, the CPI(M) had been in power in West Bengal since the 1970s. I use state assembly election data from 2006 to 2017, covering elections in 2006, 2011 and 2016, and discuss my data in greater detail in the next section.

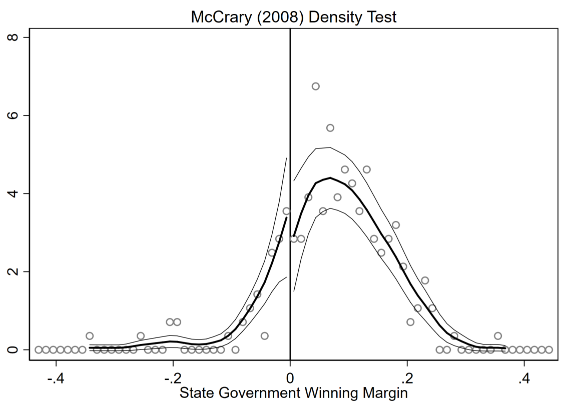


Figure 2: McCrary Test – density of winning margins at cutoff

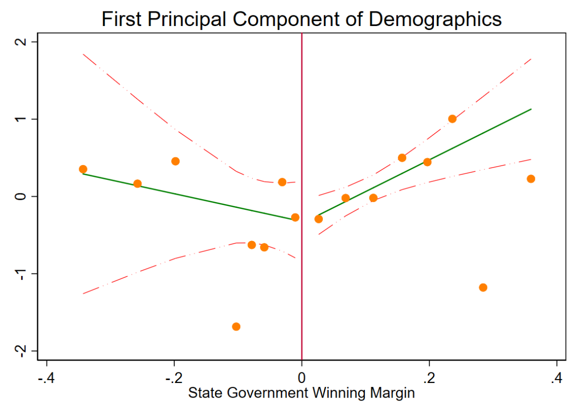


Figure 3: Balance on PCA of age, gender and caste

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 test for discontinuities in demographic characteristics of assembly candidates on either side of the cutoff and find that there are no significant discontinuities in the first principal component of age, sex and caste of the candidates. I also show balance in terms of village characteristics across the cutoff in Appendix C, Figures C3 and C4.

In order to test the validity of the RD design, I run two main checks to test for balance of the running variable and other characteristics on either side of the cutoff. On running a (McCrary, 2008) test, I find no significant discontinuities in the density across the cutoff (Figure 2). Similarly, when comparing candidate characteristics such as age, gender and caste, a measure of the first principal component of the three does not yield any significant discontinuities across the cutoff (Figure 2). This strengthens any causal claims that I make using this RD.

4 Data Description and Variable Definitions

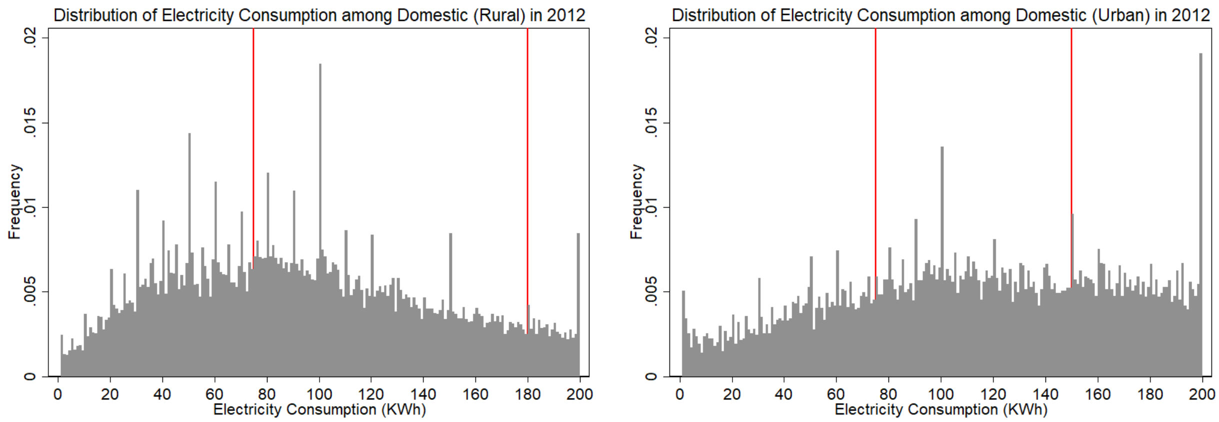
4.1 Administrative data on Electricity Consumption and Billing

I obtain confidential administrative data on the universe of electricity consumption and billing records from the West Bengal State Electricity Distribution Corporation Limited (WBSEDCL). This is a state-owned utility in West Bengal, serving a consumer base of approximately 17 million households, or 72 million customers. These data include consumption for residential, commercial and agricultural users in both rural and urban areas between 2011 and mid-2017. For most consumers, billing is done quarterly, with the exception of a few monthly users with commercial accounts. WBSEDCL faces no competition from other electricity distributors within its purview, and the only area not covered by WBSEDCL is the capital city of Kolkata.

The utility is controlled by an independent regulatory board, the West Bengal State Electricity Regulatory Commission (WBERC). WBERC accepts proposals from WBSEDCL requesting tariff increases to meet their rising marginal costs of providing electricity. After reviewing these reports, WBERC sanctions a tariff revision, that can occur at any time within a year. I compile a dataset of these tariff revisions that include changes across tiers in the pricing structure, as well as different tariff schedules for different consumer categories.

In order to bill consumers, WBSEDCL sends inspectors to account holders' homes to enter their meter readings into the database. Electricity meters function akin to car odometers, where the number on the meter represents the cumulative consumption of the account holder. 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. Indeed, when I plot the consumption distribution for residential and commercial consumers in Figure 4, I observe a highly multi-modal distribution of consumption, with significant bunching at various points. The peaks in the data often appear at round numbers such as 20, 30 or 40 KWh. While it is common for meter inspectors to not conduct readings every billing cycle and make imputations for the periods between their observations, the spikes observed in the data are quite large. Using the RD, I test whether this occurs systematically more in certain areas based on political alignment.

Figure 4: Consumption Distribution for Residential Consumers



Notes: The consumption distributions above are for residential consumers in rural and urban areas. The range of consumption extends from 1 KWh to more than 1000 KWh, but the bulk of distribution lies below 200 KWh, and largely has the shape of a chi-squared distribution. I restrict consumption to under 200 KWh in these graphs. There are several spikes in the distribution particularly at multiples of ten and five.

4.2 Measures of Data Manipulation

Based on the multi-modal consumption distribution, I define two measures to characterize manipulation of the underlying data. Benford's (1938) Law lays out an expected distribution for the first digit of a naturally occurring set of numbers. I measure the normalized distance of the consumption distribution for each assembly-year from the expected distribution. This metric, which is the same as the chi-squared goodness-of-fit statistic provides me with 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, continuously and smoothly distributed, a multiple of ten should not occur discontinuously more just above the RD cutoff.

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. 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.[7] The close-election RD provides a neat way of identifying whether these billing items are systematically different in constituencies supporting the majority party.

Table 1: Summary Statistics for Outcomes in Winning and Losing Legislative Assemblies

	2011		2016	
	Winning	**Losing**	**Winning**	**Losing**
Number of Constituencies	227	67	211	83
Chi-Sq. Square Distance	26.59	11.85	34.42	32.33
Fraction of consumers with whole numbered KWH	0.15	0.16	0.13	0.13
Reported consumption (KWh)	260.55	174.39	270.96	181.27
Sum of all bill components (Rs.)	1533.27	979.10	1754.30	1117.91
Sum of all arrears (Rs.)	90.14	48.79	56.43	33.78
Average energy price per KWH (Rs.)	3.89	3.52	5.45	4.93
Average arrear per KWH (Rs.)	0.42	0.29	0.50	0.45
Total subsidies in Bill (Rs.)	-153.56	-104.56	-109.25	-79.19
Connected Load (KVA)	1.08	0.81	1.13	0.81

Notes: Summary statistics based on confidential billing data. The above table shows the mean level of the outcome variables by legislative assemblies that are aligned ('Winning') and not aligned ('Losing') with the governing party, for each respective election. I show billing outcomes from 2012, when my data begins, under the 2011 column. The 'Chi-Sq Square Distance' is a measure of distance of the reported consumption distribution from the expected distribution. Connected Load refers to a predetermined maximum demand based on the appliances used in a household.

Given that central regulations do not allow political entities any direct control over electricity tariffs, these measures would enable me to test whether they indirectly influence electricity tariffs through the manipulation of the above measures. This may point towards a patronage model of politicians in power wanting to reward their voters. If bills are manipulated to reflect lower than actual consumption, that would amount to an indirect subsidy to constituents.

In the consumption dataset, each account holder is linked to a consumer care center (CCC). These centers are the local administrative offices for WBSEDCL, in charge of billing. I geo-locate each of these 510 CCCs and situate them within their respective legislative

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

assemblies, ending up with 2-3 CCCs per assembly area. Through their CCCs, therefore, all account holders under WBSEDCL are assigned to a particular legislative assembly and I use this setup to run an RD analysis. I hypothesize that if politicians wanted to indirectly subsidize their voter base, they would do so by influencing the local CCCs within their jurisdiction. One possible channel through which they may operate is to selectively not enforce local contactors in charge of meter readings to record observations regularly. Rains and Abraham (2018) identify this as a vulnerability in bill collections due to low incentives of contractors collecting consumption meter readings. Not having regular meter readings allows local billing centers to make their own imputations of consumption, and could be made lower to appease the local MLA.

Table 1 presents summary statistics for the main variables of interest by whether or not the constituency was aligned with the majority party, and also by years 2012 and 2016. In the RD analysis using billing data, I make use of only the 2016 election, due to data availability. All results from this analysis using billing and consumption data reveal political behavior post-elections.

4.3 Satellite Nighttime Luminosity Data

I use nighttime light density as a measure for actual electricity consumption in grid-connected areas, and possible new electrification. This is an un-manipulable measure of consumption, and serve as a barometer for the reported consumption measures from the electricity bills.

Satellites from the United States' Defense Meteorological Satellite Program (DMSP) collect images of the earth twice a day, and they make 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. This data has also been used in the economic literature to capture economic development (Chen and Nordhaus, 2011; Donaldson and Storeygard, 2016; Henderson et al., 2012). Figure C1 in Appendix C shows a map of West Bengal with both (state-level) assembly boundaries and (national level) parliamentary constituencies. I intersperse these administrative boundaries with the luminosity data.

These luminosity measures are also effectively used as a proxy for electrification, often corroborated by actual consumption measures. Min and Gaba (2014); Min et al. (2013) use this data to examine electrification in Vietnam, Senegal and Mali, and validate nighttime lights as a good proxy, particularly for rural electrification. Several papers have use this data in the Indian context specifically to measure electrification rates (Burlig and Preonas, 2017; Mann et al., 2016; Min and Golden, 2014). Mann et al. (2016) apply machine learning techniques to predict daytime electrification, and show nighttime luminosity to be a good indicator of electricity consumption. Min and Golden (2014) and Baskaran et al. (2015) show evidence of electoral cycles in electricity supply using the DMSP data, and Burlig and Preonas (2017) are able to assess the development effects of electrification using this data as a proxy for village electrification. Given the evidence of electricity consumption data manipulation, this data also provides an unbiased measure of electrification.

For this paper, I utilize the average density of lights within each legislative assembly boundary, and test for a discontinuity at the RD cutoff. In the absence of any political manipulation of the utility's consumption data one would expect them to have information similar to lights data.

5 Empirical Evidence of Political Patronage

The RD analysis uses various electricity outcomes to demonstrate potential manipulation of the sector by political agents. I use average nighttime lights density to capture whether the majority party provided more electricity to assemblies that voted for them. Exploring the mechanisms through which this control may be exercised, I test whether the party in power provided differentially cheaper electricity access to its voters. Given the metrics I define, this translates to under-reporting of actual consumption (and providing a tacit discount), understating owed arrears, lowering energy prices, and overstating subsidies in areas that voted for the MLAs from the majority party. In addition, manipulation of bills may affect the distribution of billed electricity, and I may observe a higher degree of deviation of the consumption distribution from the chi-squared distribution in these areas.

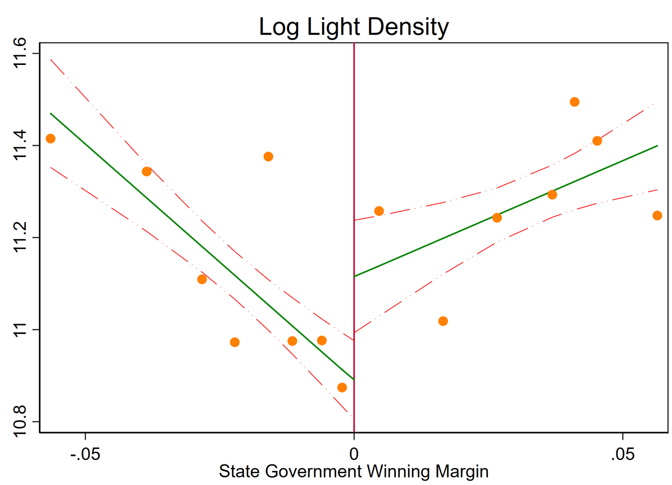
5.1 Average Nighttime Lights Density

In order to test for discontinuities in electricity consumption, I run the following regression 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 number of votes:

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

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 capture the Local Average Treatment Effect (LATE), I can make causal claims for the sub-sample of assemblies close to the winning margin cutoff. This includes swing areas – assemblies where the party in power narrowly won or lost, in which, as the theory suggests in Appendix A, parties concentrate their efforts as the expected payoff may be higher.

Figure 5: RD analysis of average nighttime lights density on either side of the RD cutoff



Notes: Comparing legislative assemblies where the party in government narrowly won to those where it narrowly lost (2012-15), I find a discontinuously higher density of nighttime lights in winning areas. I use the Calonico et al. (2015) method to create optimal bins for observations on either side of the cutoff and a linear specification to fit the data.

Figure 5 demonstrates that there is discontinuously higher light density for assemblies where the chief minister's party narrowly won. Since there was balance across the cutoff on characteristics such as age, gender and caste of the candidates, this discontinuity in electrification points towards differential treatment by the politicians in power.

In order to further investigate this pattern, I use nighttime light density data from 2004-2016, spanning the state elections in 2006, 2011 and 2016. The 2006 elections serve as a control to check whether there was a trend towards discontinuously higher electricity consumption in select regions, following the previous election as well, which was won by another party, the CPI(M). I run the following regression, where β_t is the regression coefficient across years.

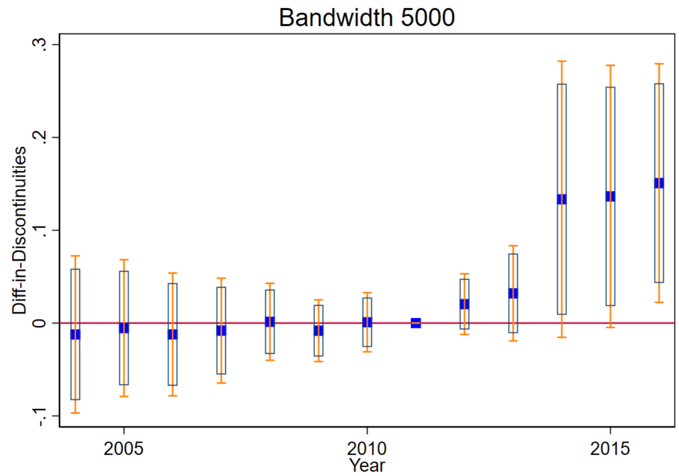
$$Log(Lights)_{at} = \sum_t \beta_t(\mathbb{1}(votemargin > 0)_a \times \gamma_t) + \gamma_t + f(votemargin)_a + \epsilon_{at} \;\; for \; a \in BW \tag{2}$$

In Equation 2, I study how being above the 2011 winning margin cutoff affects light-density both before the elections (2004-2010) and after (2012-2016). We should expect that the pre-election years show no detectable discontinuity, as a falsification test. We should further expect that after 2011, there is an increase in the discontinuity, allowing us to measure the dynamics of this relationship. This regression specification amounts to a difference-in-discontinuities set up, which includes year fixed effects γ_t, restricting the sample to a bandwidth around the cutoff.

On graphing these coefficients in Figure 6, I observe that there was no discontinuity or differential electrification in years before the 2011 elections. Furthermore, after the 2011 elections, there is a clear trend break, and I observe an increase in differential electrification in assemblies where the chief minister's party narrowly won. From this observation, I infer evidence of favoritism shown in electricity provisioning in these areas.

Neither night-lights density data nor utility administrative and consumption data alone can provide a complete picture of electricity use. However, a combination of both, like in this paper, is useful to get a clearer sense of the patterns in the data. Given the stark increases in the RD coefficient for nighttime lights soon after 2011, it is more likely that the effects I observe do refer to electrification outcomes, as opposed to development schemes which typically take longer to have observable effects.

Figure 6: Satellite Night Lights: Difference-in-discontinuities Analysis from 2004 to 2016



Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot the RD coefficients, and confidence intervals of errors clustered at the assembly level. The dependent variable is Log(light density). I plot coefficients over time and find a trend break after the 2011 election, with selectively greater electrification in areas where the governing party narrowly won. For a figure showing the levels of the RD coefficients over time, please refer to Figure C2 in Appendix C

5.2 Data Manipulation in Electricity Billing Records

With the help of the micro-level consumer data, I identify the ways in which politicians influence electricity access. I show evidence in Figure 7 using consumption data on all consumer classes, including households, commercial users, public works, agriculture and irrigation. The only consumer class not present in this dataset is industry, specifically high-tension consumers of electricity, usually large factories. Therefore, aside from factories, which do not commonly operate at night, the nighttime lights data closely corresponds to the consumers captured in the billing dataset. I show evidence of data manipulation in electricity consumption records, in a manner that favors assemblies that voted for the ruling party. I run the following regression specification at the assembly level, where the left hand side includes various measures of data manipulation:

$$y_a = \beta \, \mathbb{1}(votemargin > 0)_a + f(votemargin)_a + \epsilon_a \;\; for \; a \in BW \tag{3}$$

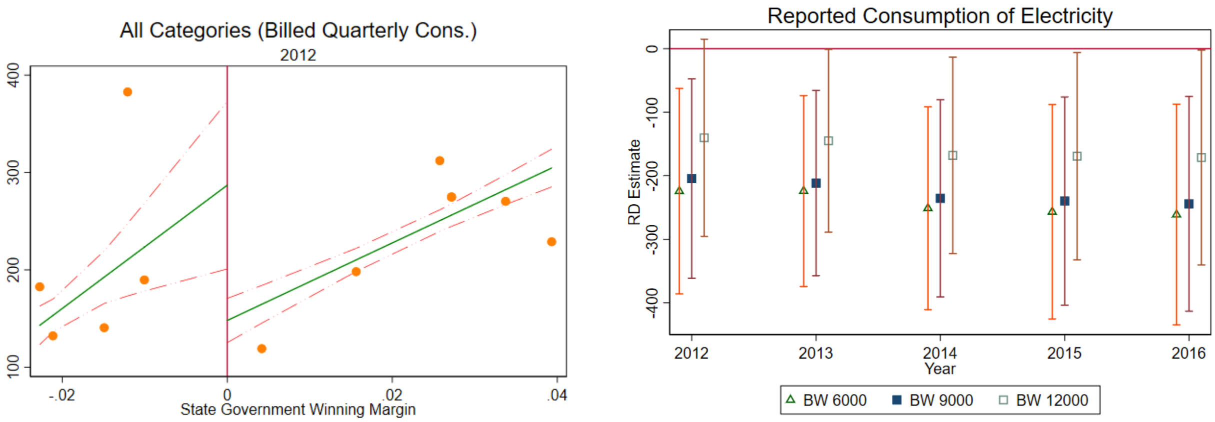
The first variable I study is simply the reported level of consumption in swing-assemblies.

Table 2: Discontinuity in Reported Consumption

	Unit consumption in KWH				
			Residential (Rural)		
RD Estimate	-124.1***	-126.0***	-143.2***	-157.9***	-139.5***
	(24.33)	(20.58)	(21.08)	(22.57)	(23.70)
Observations	7,780	10,457	10,352	10,329	10,213
Bwidth	6,000	6,000	6,000	6,000	6,000
Year	**2012**	**2013**	**2014**	**2015**	**2016**
			Residential (Urban)		
RD Estimate	-311.4***	-366.2***	-382.9***	-401.8***	-433.1***
	(95.28)	(82.32)	(77.72)	(75.35)	(71.69)
Observations	9,630	11,417	11,350	11,260	11,075
Bwidth	6,000	6,000	6,000	6,000	6,000
Year	**2012**	**2013**	**2014**	**2015**	**2016**
			Commercial (Rural)		
RD Estimate	124.8	51.21	81.79	-16.16	107.4
	(99.62)	(78.51)	(70.12)	(80.87)	(88.63)
Observations	3,023	4,120	4,044	4,018	4,010
Bwidth	6,000	6,000	6,000	6,000	6,000
Year	**2012**	**2013**	**2014**	**2015**	**2016**
			Commercial (Urban)		
RD Estimate	-473.4*	-579.9**	-555.3**	-542.6**	-582.3**
	(273.20)	(250.70)	(234.50)	(265.40)	(291.80)
Observations	10,611	12,505	12,227	12,269	12,035
Bwidth	6,000	6,000	6,000	6,000	6,000
Year	**2012**	**2013**	**2014**	**2015**	**2016**

Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), 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 6,000 in terms of the the running variable, winning 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 7: Lower reported consumption in regions where the majority party won (2012-15)



Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot reported consumption of electricity on either side of the cutoff. The running variable for the RD is the winning margin percentage. The left hand side panel uses 11,592 data points, for which I use a 2% sample of the billing data for all consumer categories. In the right panel, I plot the RDD coefficients between 2012 and 2016, and find results robust to other bandwidths – both lower and higher than the optimal bandwidth. Standard errors are clustered at the feeder level.

Given that there is no observable discontinuity in baseline characteristics around the cutoff, there is no *a priori* reason for there to be discontinuities in reported consumption. In Figure 7, 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 government's favor. However, in the previous section, I observe a discontinuously higher level of nightlights density. One possibility is that the *billed* consumption understates *actual* consumption. The magnitudes of these discrepancies are large, amounting to average discounts to constituents of about 50% of their regular bills.[8]

A potential concern with using satellite data is that it may primarily capture an increase 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 four assemblies below and above the RD cutoff, the number of villages receiving electricity connections through this scheme is very similar (5944 compared to 6024 in constituencies

[8]These magnitudes are based on a rough calculation using the estimated effects of being in a constituency of the ruling party and the average electricity consumption at the cutoff in assemblies aligned with the opposition.

of the ruling party).[9] Given that a marginally greater number of villages in constituencies supporting the ruling party received new electricity connections through the scheme, it is all the more striking that their reported billed consumption is discontinuously lower. Another concern with satellite nighttime lights data is that it captures mainly rural electrification. If I focus on only rural consumers in the billing data, I still find evidence of political manipulation for residential consumers (Table 2).

Next, I examine patterns in the data that may shed light on the observed underreporting of electricity consumption. In Figure 8, 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 winning swing assemblies. 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 not completely clear 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. These results are presented for the optimal bandwidth (Calonico et al., 2015) and for bandwidths both smaller and larger.

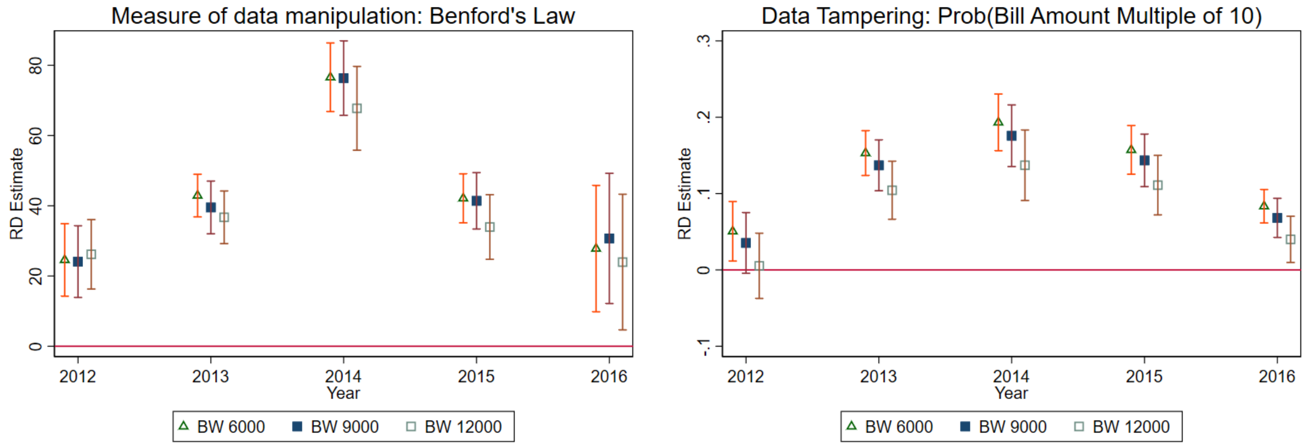
In order to further scrutinize the data manipulation, I examine trends in the RD coefficient for potential manipulation of price variables – namely arrears and subsidy payments in Figure 9. I observe a statistically significantly higher level of subsidies in winning swing assemblies, accompanied by a 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.

5.3 Channels of Political Influence

There are a few possible channels through which politicians reward their voters with cheaper electricity. Electricity meter readings provide one of the few manipulable margins on which to affect electricity price, as the price and total bill estimates are computerized and harder to manipulate without detection. Among several vulnerabilities, Gulati and Rao (2007) identify

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

Figure 8: RD Coefficients for Manipulation Outcomes Across Bandwidths

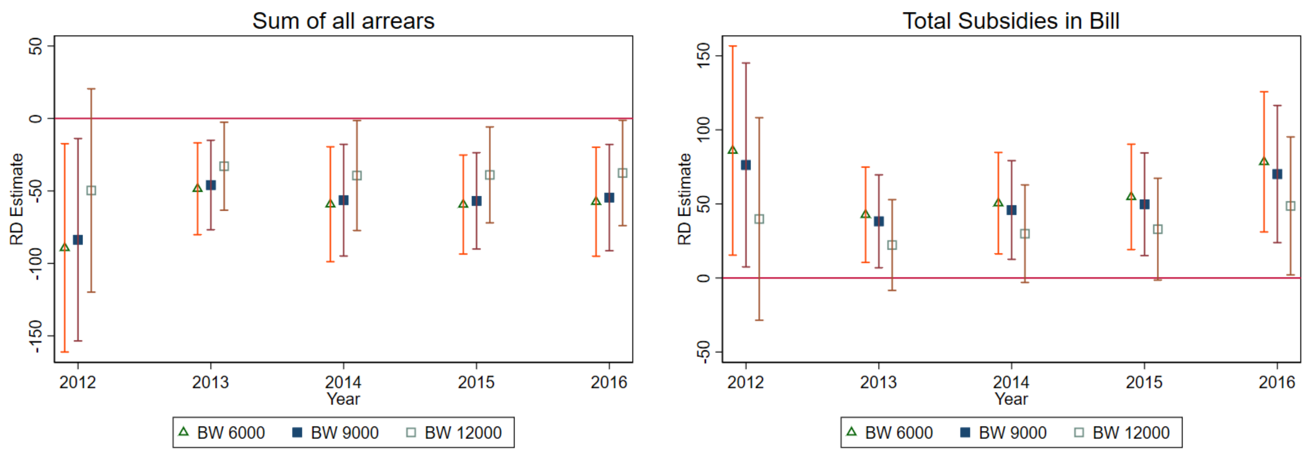


Notes: Using the Calonico et al. (2015) optimal bandwidths and bias-correct RD methodology, I plot coefficients across years for measures of data manipulation, and confidence intervals of robust standard errors clustered at the electrical-feeder level. Specifically I study the distance of the observed distribution from the expected distribution as per Benford's (1938) Law and the fraction of consumers whose consumption was a multiple of ten. 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.

the billing stage as susceptible to political interference, highlighting artificially lowered bills as a specific example. An auditing study carried out by an electricity utility in Uttar Pradesh, another Indian state, identified significant political interference in electricity distribution at local levels (Goenka, 2013). The inspectors who conduct meter readings are often external contractors. They report to a local Customer Care Center (CCC), which enters their reported consumption figures into the digital database. Given that I observe under-reporting of consumption, this appears a likely point where under-reporting occurs. Rains and Abraham (2018) highlight the role of these inspectors in bill collection and how redesigning incentives for them 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 Members of Legislative Assembly (MLAs) to report billed consumption that is lower than actual levels. These billing centers are dispersed all over the state, and it is in narrowly winning assemblies that we observe statistically significantly lower levels of reported consumption.

Another possibility is that politicians selectively discourage utility action against energy theft, tacitly allowing it. Even though I am unable to test this directly, there is a large

Figure 9: Regression Discontinuity coefficients for outcomes across three bandwidths



Notes: Using the Calonico et al. (2015) optimal bandwidths and bias-correct RD methodology, I plot 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.

amount of anecdotal evidence supporting this channel (The Telegraph, 2014; The Times of India, 2017; The Washington Post, 2012).[10]

A centrally mandated independent regulatory authority ensures that it virtually impossible to directly reduce electricity prices. They set tariffs after approving requests by the electricity provider to do so, in response to the changing price of fuels that generate electricity (marginal costs of producing electricity), as well as changes to the composition of the generation stations supplying them. Chatterjee (2017) discusses evidence from interviews with regulators where they report pressure by politicians in government to delay these tariff revisions, but there is little evidence that politicians were able to affect the setting of tariffs themselves.

An alternative explanation for the observed discontinuities is that the reported consumption in swing assemblies where the majority party narrowly lost was over-stated. I cannot

[10] " "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.*

eliminate this possibility, given that the RD analysis provides me with relative changes. Yet, it is unlikely that politicians would expend effort in overcharging consumers in constituencies where they lost elections rather than favoring their own constituents. Over-stating bills is easier to detect and may lead to widespread discontent and protests, and hurt the chances of the ruling party from winning further elections in swing regions.

Another possibility is that rather than manipulating data, electricity distributors provide greater access to electricity for consumers in assemblies where the governing party loses, in a bid to win over new voters. However, this is at odds with the evidence from the night lights data, which shows a discontinuously lower level of actual electrification in assemblies where the governing party narrowly lost (Figure 6 & Figure 7). Lastly, favoring voters in assemblies, where the ruling party lost, is unlikely to win new votes if the beneficiaries credit the MLAs from the losing party (that is in office in areas with better electricity access).

5.4 Falsification Tests and Robustness Checks

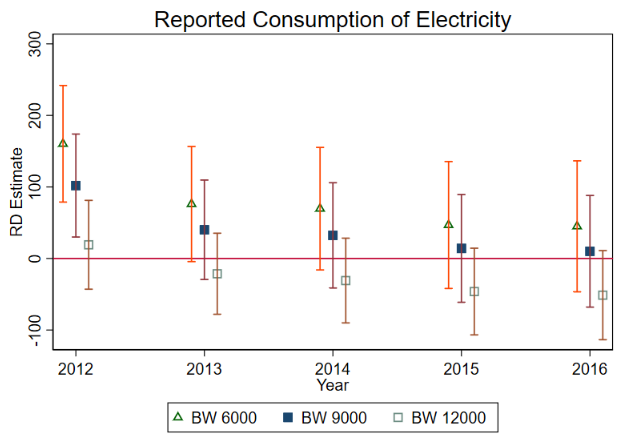
I test for robustness across multiple bandwidths in the RD analysis. I present these figures in Section 5.2 for the RD results on reported consumption, distance from the chi-squared distribution, and bill items, all of which are consistent across different bandwidths.

Next I conduct falsification tests where I use the winning margin and the set of winning and losing assemblies from a previous election (the 2006 election where the CPI(M) party formed the government). If the most likely narrative is that the current political party in power (that ascended after the 2011 elections) induces discontinuities in the consumption and billing data, then I should not observe such discontinuities for assemblies near the 2006 election cutoff in the years after 2011.

In Figure 10, I show the RD results analogous to those in Section 5.2. Using the 2006 election winning margin, I do not observe any robust evidence of a discontinuity. Indeed, the figure shows a slight discontinuity in 2012 perhaps due to some persistence in corruption and manipulation that may have been occurring between 2006-11 under the previous government. After 2012, however, there is no detectable discontinuity.

In Figure 11, I observe a similar pattern of no discontinuities. Again, there is weak ev-

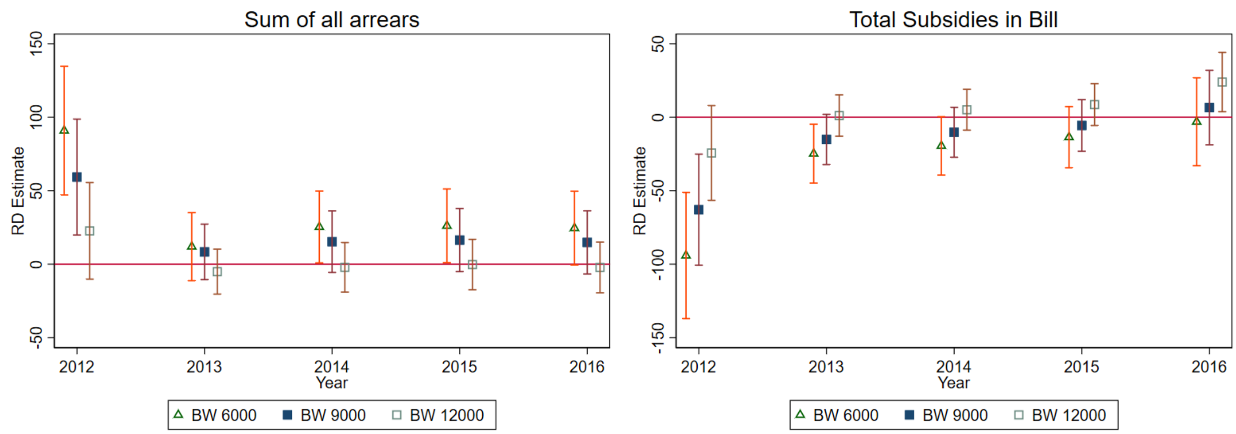
Figure 10: Placebo test: studying discontinuities in reported consumption using the winning and losing constituencies from the 2006 election



Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot RD coefficients for the reported consumption. The winning margin here is defined on the basis of legislative assemblies from the 2006 election, where the CPI(M) party won, and was in power till 2011. This provides a falsification test for the validity of the results using the 2011 election results. The results shown include multiple bandwidths (BW).

idence of a discontinuity in 2012, immediately post the 2011 elections, suggesting possible persistence in manipulation from the previous winning governing party. This points towards evidence that similar political influence in bill items occurred for assemblies where the previous governing 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.

Figure 11: Placebo test: studying discontinuities in bill items (arrears and subsidies) using the winning and losing constituencies from the 2006 election



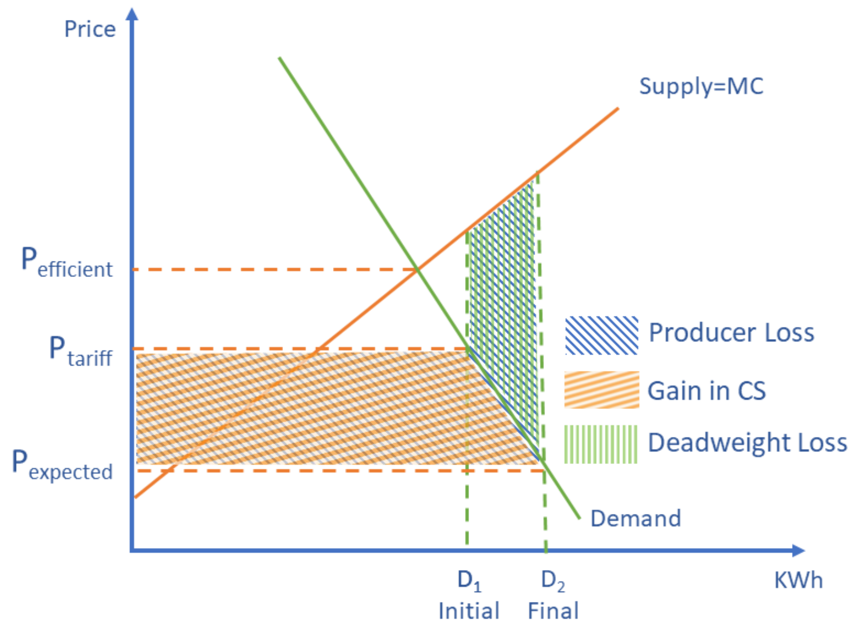
Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot RD coefficients for billing items including arrears and subsidies. The winning margin here is defined on the basis of legislative assemblies from the 2006 election, where the CPI(M) party won. This provides a placebo test for the validity of the results using the 2011 election results. The results shown include multiple bandwidths.

6 Welfare Consequences of Political Patronage

In order to address the widespread corruption in the electricity sector, it is important to understand the magnitude of the problem, as well as the welfare costs that it imposes on society. I have until now, shown causal evidence that politicians selectively manipulate billing data to informally subsidize areas aligned with the ruling party. In this section, I quantify the costs and benefits from these actions.

I rely on a simple demand and supply framework to provide an intuitive account of the welfare implications of political patronage. I characterize the under-reporting in billing data as providing an informal subsidy to constituents of the ruling party. This under-reporting of bills can be approximated as an average price subsidy provided to all consumers in constituencies aligned with the ruling party. In Figure 12, I describe this setup with a downward sloping consumer-demand curve and an upward sloping provider-supply curve, based on the assumption that as supply increases the electricity provider must purchase electricity from progressively more expensive sources. Under an efficient market, the price charged for electricity would be $P_{efficient}$. However, in reality, most electricity providers cross-subsidize residential and smaller commercial users by charging higher prices for industrial

Figure 12: 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 almost any electricity markets. The most common pricing scheme is to cross-subsize 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 face a flat rate of P_{tariff}. Politicians, through corruption, effectively lower this price even further for their constituents, to P_{tariff}. I assume that the marginal cost (MC) curve facing producers is an upwards sloping line, accounting for sourcing electricity from increasingly expensive thermal power plants or gas plants, as the quantity supply increases. The shaded areas show the loss in producer surplus, gain in consumer surplus, and overall deadweight loss. What is clear from the figure is that in order to calculate consumer surplus, I need estimates of the price elasticity of demand.

users. It follows that the price paid per unit of electricity by consumers in my data is lower than $P_{efficient}$, and I refer to this price as P_{tariff}. The price schedule facing residential users in rural and urban areas, and commercial users in rural and urban areas are different. Figure C5 in Appendix C shows the price schedule for these four consumer groups between 2012 and 2016. I focus on these groups for the welfare analysis as they are the majority of consumers, but the complete dataset also consists of agricultural users, irrigation, and publicly owned buildings. It follows that $P_{efficient}$, and P_{tariff} vary across the four consumer groups I focus on. As a consequence of political patronage, consumers in constituencies of the ruling party effectively face a price of $P_{expected}$.

Figure 12 shows the loss in producer surplus, gain in consumer surplus and deadweight loss to society as a result of the informal subsidies provided by politicians to their constituents. These effects are estimated based on the additional market distortions caused by moving from P_{tariff} to $P_{expected}$. In order to estimate the change in producer surplus, which refers to the entire shaded area in the graph, we need a measure of the loss in revenue to producers. I use the RD estimates of the shortfall in consumption reporting at the cutoff, for each consumer group (Table 2) to estimate the potential "unreported" consumption as (i.e. the difference between observed consumption of constituencies on either side of the RD cutoff). Using reported estimates of the marginal cost of providing electricity, I compute the lost producer surplus due to underreporting.

I then estimate the size of the gain in consumer surplus. The difference between the changes in producer and consumer surplus provides the size of the deadweight loss to society. However, as is evident from Figure 12, the change in consumer surplus depends on the price elasticity of demand for electricity. Therefore, I first estimate 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, but that these elasticities do not change over the five-year period.[11]

However, estimating the price elasticities of demand from the consumption data is not straightforward due to evidence of 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. I then compute elasticities for each consumer category for this sub-sample using an instrumental variable approach that leverages exogenous variation in policy-led tariff changes over time. Figure C5 in Appendix C demonstrates the changes in prices over time, across tiers and for the four consumer categories I focus on. These are plausibly exogenous to short-term fluctuations in a consumer's demand as prices are set by independent regulators, and an individual's electricity demand in isolation, cannot directly affect the changes in prices. I use an instrumental variables approach that leverages variation in price tariff changes to estimate the consumption response to changes in marginal price. I calculate elasticities for each consumer group at assembly level, and each assembly is assigned a unique value of elasticity for each consumer group.

[11] It is important to note that these elasticities refer to the price elasticity of demand for *grid-purchased* electricity. This is particularly relevant for commercial users who often own generators and substitute away to non-grid sources of electricity when prices change (The World Bank, 2014).

After computing elasticities for assemblies where bills were not manipulated, the *second step* involves imputing elasticities for assemblies with data manipulation. I build a predictive model of assembly-level elasticities (in the sub-sample of assemblies with unmanipulated data) on village-level characteristics from the census.[12] This model can be used to predict elasticities for constituencies where political interference is detected. In order to estimate a model with higher predictive power, I use a post-double selection OLS (Ahrens et al., 2018). This process uses machine learning tools to select the best set of independent variables from the list of village characteristics that maximizes the predictive power of the model. This improves upon an OLS model which may suffer from omitted-variable biases and overfitting.

The *third step* involves predicting the elasticities for the remaining constituencies where there is evidence of data manipulation. I use the model selected in the second step, with a selected set of village characteristics from the census to project the elasticities for the remaining assemblies. The result is a unique estimate for elasticity for four consumer groups in each assembly in the dataset.

Finally, the *fourth step* uses the full set of estimated and predicted elasticities to calculate the consumer surplus for each consumer class, as a result of the informal subsidy provided by politicians. In order to demonstrate why these steps are necessary, I also derive welfare estimates without accounting for the presence of data manipulation, and show that my estimates are more robust than in prior work. This analysis is described in Appendix B.

6.1 Step 1: Elasticities for Constituencies with no 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 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

[12] The Indian census was conducted in 2011 and consists of individual-level demographic information such as population, literacy status, occupation, age and sex.

individual i and consumer category a level, is the simplest method of estimating elasticity but produces biased elasticities.

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

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 can identify the price-tier corresponding to the marginal level of electricity consumption of each particular consumer, 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 C5 in Appendix C).

For an individual i, in tier t, month m, year y, assembly constituency c, and consumer category a, I use an instrumental variable approach to estimate elasticities. My specification is similar to Ito (2014), but leverages heterogeneity across individuals, and differential changes across price tiers, instead of relying on a simulated IV.[13] 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. Conditional on individual fixed effects, tier-by-month fixed effects, and consumer-category-by-month fixed effects, I instrument the marginal price $\log\left(MP\right)$ with the change in tariffs $\Delta\log\left(Tariff\right)$ across years. The first and second stage are respectively:

$$\log\left(MP\right)_{iamtcy} = \sum_a \gamma_{ac}\Delta\log\left(Tariff\right)_{amtcy} + \nu_{mta} + \zeta_{mac} + \eta_i + \varepsilon_{iamtcy} \;\; \forall \;\; a \in A \tag{5}$$

$$\log\left(Cons\right)_{iamtcy} = \sum_a \beta_{ac}\log\left(\widehat{MP}\right)_{iamtcy} + \tau_{mta} + \mu_{mac} + \omega_i + \epsilon_{iamtcy} \;\; \forall \;\; a \in A \tag{6}$$

I estimate β_{ac} separately for all constituencies a that lie in the set A of assemblies for

[13]The simulated IV method would be more appropriate with a longer time period in my panel dataset.

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, month-by-tier fixed effects τ_{mta}, and consumer-category-by-month fixed effects μ_{mac}. The advantage of having individual fixed effects is that it accounts for baseline consumption. The different month fixed effects allow for seasonality in consumption to vary by tier and consumer category. Standard errors are clustered at the consumer level.

Table 3: 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 linear year trend, customer fixed effects, customer-category-by-month fixed effects, and tier-by-month fixed effects. 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.

Table 3 presents results by running the specification in Equations 5 and 6 for all assemblies with unmanipulated data. This table serves only to provide consolidated elasticities for the assemblies, but I estimate this specification separately for each assembly 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) have more elastic demand. The differences in elasticites 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.

6.2 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, 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 C2 in Appendix C shows the final model used in the prediction step.

6.3 Step 3: Predicting Elasticities for all Constituencies

Following the PDS OLS method, I predict elasticities for constituencies that showed evidence of data manipulation. Table 4 shows the mean values of the resulting elasticities. These differ from Table 3 because they represent the mean elaticity for each consumer category taking into account *all* assemblies, those with unmanipulated as well as manipulated data.

Table 4: 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 3 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.

The elasticity estimates in Table 4 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 am able to observe the marginal price paid by the consumer, and the price tier that they consume over 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 Appendix Section C Table B1. The results in column 1 of Table B1 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 range 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 4). By estimating elasticities in only those regions where there was no evidence of manipulation, provides more precision and removes the biases is 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 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 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.

6.3.1 Targeting Inelastic Consumers

My model (Appendix Section A) predicts that politicians target consumer groups who have inelastic demand, and also regions that have infrastructure conducive to electricity usage. These results are intuitive. Consumers with inelastic demand are usually those who will benefit most from reduction in billed electricity quantity. Therefore, if politicians intended their subsidies to have large impacts, it follows that they would target those with inelastic demand. The model result that politicians target areas with greater access to infrastructure may empirically translate to urban areas, which have more infrastructure, and are wealthier. Arguably living in such areas, in contrast with rural areas that lack access, would also be correlated with greater political influence. This is consistent with studies showing how politicians use electricity prices to target influential groups to curry favor.[14]

When studying the effects of manipulation by consumer category in Table 2, I find no statistically significant discontinuity in reported consumption for commercial users in rural areas. Commercial rural areas have the most elastic demand (Table 4), and also lack the infrastructure (a proxy for influence) to use a constant supply of electricity. It follows that they do constitute the most attractive group for politicians to expend effort targeting subsidies towards. I observe a large discontinuity in reported consumption for residential (both urban and rural) consumers and commercial users in urban areas. Given that the elasticity for residential users is quite low, on average, -0.41 (Table 4), it is not surprising that politicians target them as they would be more affected by tariff increases.

The fact that there is a discontinuity for commercial (urban) consumers in Table 2 is consistent with the model prediction that politicians also target consumers in regions with more infrastructure and higher wealth levels. Commercial users in urban areas have the highest baseline consumption, (a mean of 420 KWh/quarter as compared to 184 KWh/quarter for commercial users in rural areas). The true estimate of electricity consumed for commercial

[14](Badiani et al., 2012) show evidence of politicians wooing rich and influential farmers by guaranteeing free or cheap electricity.

urban accounts is likely higher, given the evidence of under-reporting of bills for that group. Given their location and implied influence based on being the the highest consumers, it follows that politicians justify under-reporting their consumption or avoiding clamping down on energy theft for commercial urban users more so than commercial rural consumers.

6.4 Step 4: The Costs and Benefits of Political Patronage

In the absence of any political involvement in electricity provision, we would expect the electricity markets to perform relatively efficiently. Yet, the evidence presented in this paper demonstrates a combination of under-reporting of consumption and unchecked energy theft in areas where the ruling party narrowly won. A government subsidy in a previously efficient market results in a deadweight loss. However, the more inelastic the demand, the smaller this deadweight loss. If, as I show, the government targets consumer bases with relatively inelastic demand, the deadweight loss is minimized.

A direct advantage of having manipulated data is that I can measure the amount of under-reporting at the cutoff, and thereby calculate the loss to the utility in regions around the cutoff. In simple calculations for the loss in utility revenue caused by these political actions, I take a conservative estimate of the under-reporting in bills. I only consider a narrow bandwidth of assemblies where the majority party narrowly won (the first five closest to the cutoff) to calculate effects of political action. Using Table 2, I calculate an average level of under-reporting of bills per year for each consumer category. Applying this average level of under-reporting (details in Table C3), I calculate the aggregate level of under-reporting for all consumers in the selected assemblies. Using this information, I back out the total loss in revenue for the utility, using the 2015 level of marginal cost of producing a single KWh of electricity in the state. This combination of consumption under-reporting and allowing of energy theft produces a yearly loss to the electricity provider of $57 million.

To measure the benefits of such actions for consumers, I use the discontinuity in the lights data to estimate the increase in consumption in response to the under-reporting (interpreted as an informal subsidy). Figure 7 shows that on the one hand, there is a discontinuously lower reported electricity consumption in areas where the majority party narrowly won, while on the other, using nighttime lights as a proxy for electricity yields the opposite result in

Table 5: Utility Loss Calculations

Consumer Class	**Producer loss (Million Rs./year)**	**Gain in surplus (Million Rs./year)**
Residential (Rural)	₹295.84	₹101.27
Residential (Urban)	₹323.77	₹177.80
Commercial (Urban)	₹111.41	₹11.76
Total (Million Rs./year)	₹731.01	₹290.83
Total (Million Rs./year)	₹3660.05	₹2401.95
Total (Million $ for 5 years)	**$ 57.11**	**$ 22.72**

Notes: To calculate producer losses, I use the estimates of lower reported consumption in areas supporting the governing party from Appendix Section C Table C1. An average of these estimates for each consumer category provides an estimate of the shortfall for producers in terms of how much electricity they supply and how much they get paid for. Multiplying these shortfall estimates with the total consumer base in these regions and the difference between price paid and marginal cost of producing electricity, gives me the final numbers for producer losses. I take the marginal cost of providing a KWh of electricity for the utility as Rs. 3.97 based on 2015 spot market data. The consumer base is restricted to twelve assemblies within the RD bandwidth of 12,000 votes from Appendix Section C Table C1. Commercial (Rural) consumers are excluded as there was no detectable change in reporting or consumption for this sub-group. For consumer surplus, I find the product of the consumer surplus per consumer and the total consumer base in the relevant regions. Each consumer's change in surplus is the found by multiplying the base level of their quarterly bill payment with the percentage increase in consumer surplus and average household size. Finally, the percentage change in consumer surplus is derived from the change in $\Delta \log(prices) * (1 - 1/\epsilon)/(1/\epsilon)$. ϵ refers to the elasticity estimates I calculate in Table 4. $\Delta \log(prices)$ is $\Delta \log(consumption)$ in Appendix Section C Table C1 divided by elasticity ϵ.

Figure 5. If the under-reporting of consumption is indeed seen as an informal subsidy, the result with nighttime lights may be interpreted as the consumption response to this subsidy.

I first need to find the elasticity between night-time lights and actual consumption. I do so by, once again, restricting myself only to regions that did not show evidence of manipulation. To estimate this elasticity I regress $\log Light\ Density$ on $\log Consumption$ at the assembly-by-year level, with year fixed effects. This regression includes all consumer categories, as it is not possible to separate the light density for each consumer class. Figure C6 in Appendix C shows this relationship in graphical form.[15]

Based on the elasticity between night-lights and actual consumption, and the increase in nightlights at the cutoff, I obtain a value for $\Delta \log (Consumption)$, and my results indicate

[15] An advantage of having such geo-coded micro-data allows me to estimate these elasticities, which may, in other contexts, be used to project electricity consumption in other parts of the world.

that there was a 1.7% average increase in consumption at the cutoff.

Using the estimated increase in consumption, the elasticity estimates by consumer category allow me to estimate the change in consumer surplus per person using Equation 21. I aggregate this figure based on the consumer base of the selected assemblies in Table C3. The aggregate increase in consumer surplus due to such informal subsidies is $22.72 million over the election term of the ruling party.

The welfare losses from these political actions are more than twice the gains in consumer surplus. As this estimate comes from a subset of constituencies within a bandwidth (in terms of winning margin) of 12,000 votes, the true losses could be much higher. However, both measures are more nuanced than these figures may indicate. Receiving greater electricity access is associated with numerous benefits in terms of labor force participation (Dinkelman, 2011), economic development (Lipscomb et al., 2013; Rud, 2012) and health (Barron and Torero, 2017), among several others. If consumers do not price these gains into their demand for electricity, then I would be under-estimating the increase in consumer surplus.

In a similar fashion, producer losses have several other consequences, not measured in my estimates. These include limited investment in maintaining and adding new infrastructure, leading to increasing blackouts and other electricity quality problems which are not quantified here. Blackouts and poor quality electricity-supply hinder manufacturing activity and other investments. As described before, if states are not directly held responsible for electricity provider losses, and they are bailed out by centrally funded schemes, there is not necessarily a direct negative consequence within the state. Yet, the bailout will affect taxes paid from other parts of the country, and therefore have distributional consequences.

7 Conclusion

This paper highlights an important fact: that regulation advocating a separation between politics and service provision may not achieve the desired consequences in the face of poor auditing or enforcement. The Electricity Act of 2003 mandated that state electricity utilities would be overseen by independent regulatory authorities who would set prices for electricity provision and supervise the running of the utilities. However, the incentive for politicians to favor their voters remains strong in spite of this. There is a history of patronage politics that runs through the Indian system (Baskaran et al., 2015)), and cheaper electricity access is often on election platforms, particularly at the state and lower levels of governance. In the absence of detailed auditing and scrutiny of consumption records, there may be channels through which politicians could still provide their voters with cheaper electricity access.

A major advantage of having confidential billing data for 76 million customers is that I am able to shed light on the consumption distributions for different categories of consumers, as well as for regions with different political affiliations. Having information on this distribution allows me to demonstrate evidence of data manipulation and to characterize these observations. While aggregate data may be able to provide general trends, it is limited in identifying data manipulation and its implications.

I find evidence that politicians do favor their voters both in terms of providing electricity access, and in subsidizing them. Using average nighttime light density data, I show that there is higher electricity consumption in areas aligned with the party in power. Politicians provide their voters with cheaper electricity access by indirectly subsidizing them through under-reporting their actual consumption. As a result, they are billed for consumption that is lower than what they actually consume. Using a close-election RD analysis, I find a statistically significantly higher level of subsidies, and lower level of arrears owed. Consistent with the hypothesis that political agents may influence intermediaries to manipulate the data, I find that in swing assemblies where the governing party narrowly won, there are greater anomalies in the consumption distribution. The fraction of consumers whose consumption is reported to be a multiple of ten is also higher in these assemblies. This explains the large number of modes in the consumption data at multiples of ten.

These patterns are consistent with a model of patronage politics where the party in power rewards its voters and consolidates its power in a very competitive setting. Using the micro-level data, I estimate price elasticities of demand for various consumer classes, developing a method to do so when faced with manipulated consumption data. I find that, consistent with previous work, residential consumers are highly inelastic. In the case of commercial consumers, I am able to provide a more accurate estimate than previous studies, and find that these groups have a high price elasticity of electricity demand. Consistent with the model, I find that politicians target consumer bases that have less elastic demand and more electricity-using infrastructure.

Using the estimates for under-reporting of consumption, I calculate the total loss to the electricity provider (for the group of assemblies near the RD cutoff) as $57 million per year. With the help of elasticities by sub-group, I find that the gain in consumer surplus is only $22 million per year. The deadweight loss alone would be enough to power 3.7 million new consumers in rural areas. The welfare consequences of these interventions, however, are more complex. Targeted voters in winning constituencies may benefit from cheaper electricity, as demonstrated by the increase in consumer surplus. Yet, the loss to the provider may be distributed widely to the tax base, and indirectly hurt other voters elsewhere. As taxpayers are richer, there may be a redistribution to poorer sections of society. However, if the funds used to bail out the utilities cut into the government's developmental budgets, then these bailouts may just be detrimental to poorer sections of society.

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.

Albouy, D. (2013). Partisan Representation in Congress and the Geographic Distribution of Federal Funds. *Review of Economics and Statistics*, 95(1):127–141.

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.

Ansolabehere, S. and Snyder, J. (2007). Party Control of State Government and the Distribution of Public Expenditures. *Scandinavian Journal of Economics*, 108(4):547–69.

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

Badiani, R., Jessoe, K. K., and Plant, S. (2012). Development and the Environment: The Implications of Agricultural Electricity Subsidies in India. *The Journal of Environment & Development*, 21(2):244–262.

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.

Barron, M. and Torero, M. (2017). Household Electrification and Indoor Air Pollution. *Journal of Environmental Economics and Management*, 86(C):81–92.

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.

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

Brender, A. and Drazen, A. (2005). Political budget cycles in new versus established democracies. *Journal of Monetary Economics*, 52(7):1271 – 1295. Political economy and macroeconomics.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padr i Miquel, G. (2015). The Value of Democracy: Evidence from Road Building in Kenya. *American Economic Review*, 105(6):1817–51.

Burlig, F. and Preonas, L. (2017). Out of the Darkness and into the Light? Development Effects of Rural Electrification.

Business Standard (March 8, 2018). WBSEDCL Suffered Rs 175.85cr Revenue Loss in Fy16: Cag Report. *Staff Reporter*.

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.

Central Electricity Authority (2018). Executive summary on power sector. Technical report, Ministry of Power.

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.

Chen, X. and Nordhaus, W. D. (2011). Using Luminosity Data As a Proxy for Economic Statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

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

Cole, S. (2009). Fixing Market Failures or Fixing Elections? Agricultural Credit in India. *American Economic Journal: Applied Economics*, 1(1):219–250.

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.

Donaldson, D. and Storeygard, A. (2016). The View from above: Applications of Satellite Data in Economics. *Journal of Economic Perspectives*, 30(4):171–198.

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.

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

Fisman, R. (2001). Estimating the Value of Political Connections. *American Economic Review*, 91(4):1095–1102.

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.

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.

Hindustan Times (August 11, 2016). Discom Staff Checking for Electricity Theft Assaulted by BJP Worker in Gurgaon. *Rashpal Singh*.

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.

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.

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.

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.

McRae, S. (2015). Infrastructure Quality and the Subsidy Trap. *American Economic Review*, 105(1):35–66.

Millennium Post (December 26, 2017). Electricity Office Ransacked After Local Dies Trying to Remove Electrical Hooking. *Staff*.

Min, B. (2015). *Power and the Vote: Elections and Electricity in the Developing World*. Cambridge University Press.

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.

Ministry of Power (2018). State-wise Aggregate Technical & Commercial Losses on Yearly Basis. data.gov.in.

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

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

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

Rud, J. P. (2012). Electricity Provision and Industrial Development: Evidence from India. *Journal of Development Economics*, 97(2):352–67.

Ryan, N. (2014). The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market.

Sadanandan, A. (2012). Patronage and Decentralization: The Politics of Poverty in India. *Comparative Politics*, 44(2):211–228.

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.

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

The Economic Times (February 26, 2015). Power Utilities Should Be Freed from Political Interference:west Bengal. *Staff Reporter*.

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).

A Model of Consumer Utility and Political Patronage

I create a political patronage model based on a combination of features present in Dixit and Londregan (1996) and Stromberg (2004), and then include consumer decisions to highlight the importance of price elasticities in such a setup. I model decisions made by consumers or voters, and their political parties. The model generates testable implications and estimation equations that I investigate empirically, with implications for consumer welfare.

A.1 Consumer and Voter Decisions

A household living in assembly a under the rule of party i has a utility that depends on the consumption of electricity z_{ia} and a combination of other goods c. Political parties understand that households derive utility in the following quasi-linear manner:

$$U_{ia} \;=\; v(z_{ia}) + c \;\equiv\; \frac{exp^{\beta x_{ia}}}{1-\epsilon} z_{ia}^{1-\epsilon} + c \tag{7}$$

Here the consumer chooses z_{ia} amount of electricity given prices. $exp^{\beta x_{ia}}$ is a taste-shifter, where x_{ia} is a vector of consumer-base characteristics, like amenities, infrastructure and regional income distributions. $\epsilon > 0$ will affect the price elasticity of demand, and thereby also the voters' responsiveness to subsidies. Importantly, it is a sufficient statistic for changes to consumer welfare in response to informal price subsidies.

The "effective" electricity price faced by households p_{ia} also varies under party rule and assembly. The bundle of other goods is assumed to be the numeraire, and in equilibrium a household always consumes a non-negative amount of the other good (i.e. basic food, shelter, etc.). From the household's first order conditions, under these assumption, it is straightforward to show that the equilibrium demand curve is:

$$log\ z_{ia} = \frac{\beta}{\epsilon} x_{ia} - \frac{1}{\epsilon} log\ p_{ia} \tag{8}$$

In Equation 8, $\frac{1}{\epsilon}$ determines the price elasticity of demand, but thereby also the responsiveness of any subsidies. Furthermore, an increase in electricity-using infrastructure and wealth distributions (captured by x_{ia}) will increase the demand for electricity. For instance, urban areas have more infrastructure conducive to using electricity, and therefore demand a higher amount of electricity for a given price.

A.2 Decisions by Political Parties

Over and above the economic benefits, voters do care about which party is in power. While economic preferences are common, voters differ on ideological grounds. Voter j has a η_{ija} (positive or negative) preference for the party that is the opposition at the state-level. Additionally, they credit the party in power at their assembly level for their increase in utility from electricity. They attach a weight $exp^{\gamma D_{ia}}$ to the electricity component of their utility, where $\gamma > 1$ and $D_{ia} = 1$ if the party in the majority party is in power at the assembly level. They reward the incumbent party in power with a *vote* if:

$$vote = \begin{cases} 1 & \text{if } exp^{\gamma D_{ia}} v(z^*_{ia}) > \eta_{ija} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

A party can allocate more electricity and more subsidies (directly affecting p_{ia}) by influencing the utility at the assembly level. This influence comes at a cost e_{ia}, both in effort and resources, and the cost function is given by:

$$e_{ia} = p_{ia}^{-\alpha} \quad , \tag{10}$$

where $\alpha \leq 1$. Given the demand function, we can solve for the electricity component of utility as a function of the effective price (including the subsidy):

$$v_{ia} = \frac{\left(exp^{\beta x_{ia}}\right)^{1+\frac{1}{\epsilon}}}{1-\epsilon} \, p_{ia}^{\frac{-(1-\epsilon)}{\epsilon}} = \frac{\left(exp^{\beta x_{ia}}\right)^{1+\frac{1}{\epsilon}}}{1-\epsilon} \, e_{ia}^{\frac{\alpha(1-\epsilon)}{\epsilon}} \tag{11}$$

Equation 11 shows that consumer utility rises with greater effort made by the party to subsidize consumption. Since voters reward the party for an increase in consumer surplus, the party is motivated to provide more effort in subsidizing voters. The party can allocate resources and effort subject to spending less than their total resources E_i. They wish to maximize their total vote share subject to their resource constraint:

$$\max_{e_{i1},\ldots e_{1A}} \sum_a Pr\left(exp^{\gamma D_{ia}} v(z^*_{ia}) > \eta_{ija}\right) \quad s.t. \quad \sum_a e_{ia} \leq E_i \tag{12}$$

Parties are unaware of a specific voter's preferences, but they have learned over time that the ideological preferences η_{ija} are distributed uniformly with mean μ_{ia} and density ϕ_a. Given this assumption, the problem can be re-written as:

$$\max_{e_{i1},\ldots e_{1A}} \sum_a \phi_a \left(exp^{\gamma D_{ia}} v(z^*_{ia}) - \mu_a\right) \quad s.t. \quad \sum_a e_{ia} \leq E_i \tag{13}$$

This set-up yields the following Nash equilibrium conditions (with respect to each cost e_{ia}) for a given Lagrangian multiplier λ:

$$\frac{\alpha\phi_a}{\epsilon} exp^{\left(\beta\frac{\epsilon+1}{\epsilon}x_{ia}\right)} exp^{\gamma D_{ia}} e_{ia}^{\frac{\alpha(1-\epsilon)-\epsilon}{\epsilon}} = \lambda \quad \forall\ a \tag{14}$$

The optimal amount of effort in assembly a depends on whether or not the party is in power there D_{ia}, the density of voters ϕ_a, and other assembly level features x_{ia}, such as the amount of electricity-using infrastructure:

$$log\ e_{ia} = \frac{\epsilon}{\epsilon - \alpha(1-\epsilon)} \left[log\frac{\alpha}{\lambda\epsilon} + log\ \phi_a + \beta x_{ia} + \gamma D_{ia}\right] \tag{15}$$

Since prices (and thereby subsidies) depend on the effort and resources made by the party to subsidize consumption, we can derive expressions for both electricity prices and consumption:

$$log\ p_{ia} = \frac{-\epsilon}{\alpha\left(\epsilon - \alpha(1-\epsilon)\right)} \left[log\frac{\alpha}{\lambda\epsilon} + log\ \phi_a + \beta x_{ia} + \gamma D_{ia}\right] \tag{16}$$

$$log\ z_{ia} = \frac{1}{\alpha\left(\epsilon - \alpha(1-\epsilon)\right)} \left[log\frac{\alpha}{\lambda\epsilon} + log\ \phi_a + \beta x_{ia} + \gamma D_{ia}\right] + \frac{\beta}{\epsilon}x_{ia} \tag{17}$$

A.3 Comparative Statics and Estimation Equations

Equations 15 through 17 produce some interesting comparative statics and testable equations. First, whether or not the the party increases effort in providing more subsidies in response to various factors, and the responsiveness of demand to these subsidies depends on the price elasticity of demand $\frac{1}{\epsilon}$. Second, for sufficiently inelastic demand $\frac{1}{\epsilon} < \frac{1+\alpha}{\alpha}$ the party will target areas with more swing voters, represented by a higher density in the assembly ϕ_a.

Most importantly, however, the majority party increases their subsidization efforts in assemblies in which it is in power $D_{ia} = 1$. As voters reward the party in power in their assembly for electricity supply, for sufficiently inelastic demand, the party increases efforts in winning over such voters. This will be one primary equation of interest. To causally isolate this impact, it is necessary to control for all the other factors in Equation 15, with the help of a standard regression discontinuity equation:

$$\log\ e_{ia} = \delta_0 + f(\text{vote share of i in a}) + \tau_0 D_{ia} + \varepsilon_{ia} \tag{18}$$

Here, δ_0, captures all things constant across assemblies, like $\frac{\epsilon}{\epsilon-\alpha(1-\epsilon)} log\frac{\alpha}{\lambda\epsilon}$. The term $f(.)$ is a polynomial in the vote share of party i in assembly a, flexibly varying across the RD cutoff. This polynomial controls for all other assembly level features that may change continuously at the cutoff (like the density of voters ϕ_a or other assembly level features x_{ia}). The error term ε_{ia} is uncorrelated with D_{ia} conditional on the polynomial, and the coefficient of interest is τ_0 which is a function of ϵ and γ.

The model predicts that for consumer bases with inelastic demand $\frac{1}{\epsilon} < \frac{1+\alpha}{\alpha}$, the estimate of $\widehat{tau_0} > 0$. To measure efforts $log\ e_{ia}$, I create measures of influence and manipulation that I discuss in the empirical section below. These measures include the anomalous bunching of certain round number values for reported consumption, and different non-standard distributions of consumption amounts.

Similarly, Equations 16 and 17 motivate regression equations on the form below, where we would expect $\widehat{\tau_1} < 0$ and $\widehat{\tau_2} > 0$:

$$\log\ p_{ia} = \delta_1 + f(\text{vote share of i in a}) + \tau_1 D_{ia} + \omega_{ia} \tag{19}$$

$$\log\ z_{ia} = \delta_2 + f(\text{vote share of i in a}) + \tau_2 D_{ia} + \xi_{ia} \tag{20}$$

To measure the changes in prices and subsidies p_{ia}, I use data on subsidies in the billed amounts and the total amount of arrears. As actual consumption is systematically misreported on the bill, I utilize night-time luminosity to measure changes in z_{ia}.

Additionally, the model details three other important equations. The first is Equation 8, the demand equation, which I employ to estimate the price elasticity of demand $\frac{1}{\epsilon}$. Combining the measure of ϵ, with Equations 16 and 17, allows me to measure the credit that voters give to local leaders for providing them cheaper electricity γ.

The second prediction is that politicians target areas that have higher electricity-using infrastructure and amenities (x_{ia} shifts out the taste for electricity). As these are mostly urban areas, a testable implication is that urban areas are targeted more than rural areas.

The third equation of interest is Equation 11, which determines consumer welfare absent any changes to taxation, where the elasticity is a sufficient statistic for welfare. Given changes in observable prices and subsidies p_{ia}, along with an estimate of the demand elasticity ϵ, I can measure changes to consumer utility based simply on either prices or consumption quantities:

$$\Delta log\ v_{ia} = (1-\epsilon)\Delta\ log\ z_{ia} = -\frac{1-\epsilon}{\epsilon}\Delta\ log\ p_{ia} \tag{21}$$

This measure of welfare, however, does not capture increases in losses to the electricity provider, and perhaps the corresponding increases in taxes used to bail out the provider. To measure the extent of provider losses, I estimate the under-reporting of consumption at the RD cutoff using a similar set of equations. The advantage of having two measures of consumption – one non-manipulable (nighttime lights), and the other manipulated (reported consumption) – is that I can estimate under-reporting and thereby the loss to the utilities.

B Estimating Elasticities - Counterfactual Exercise

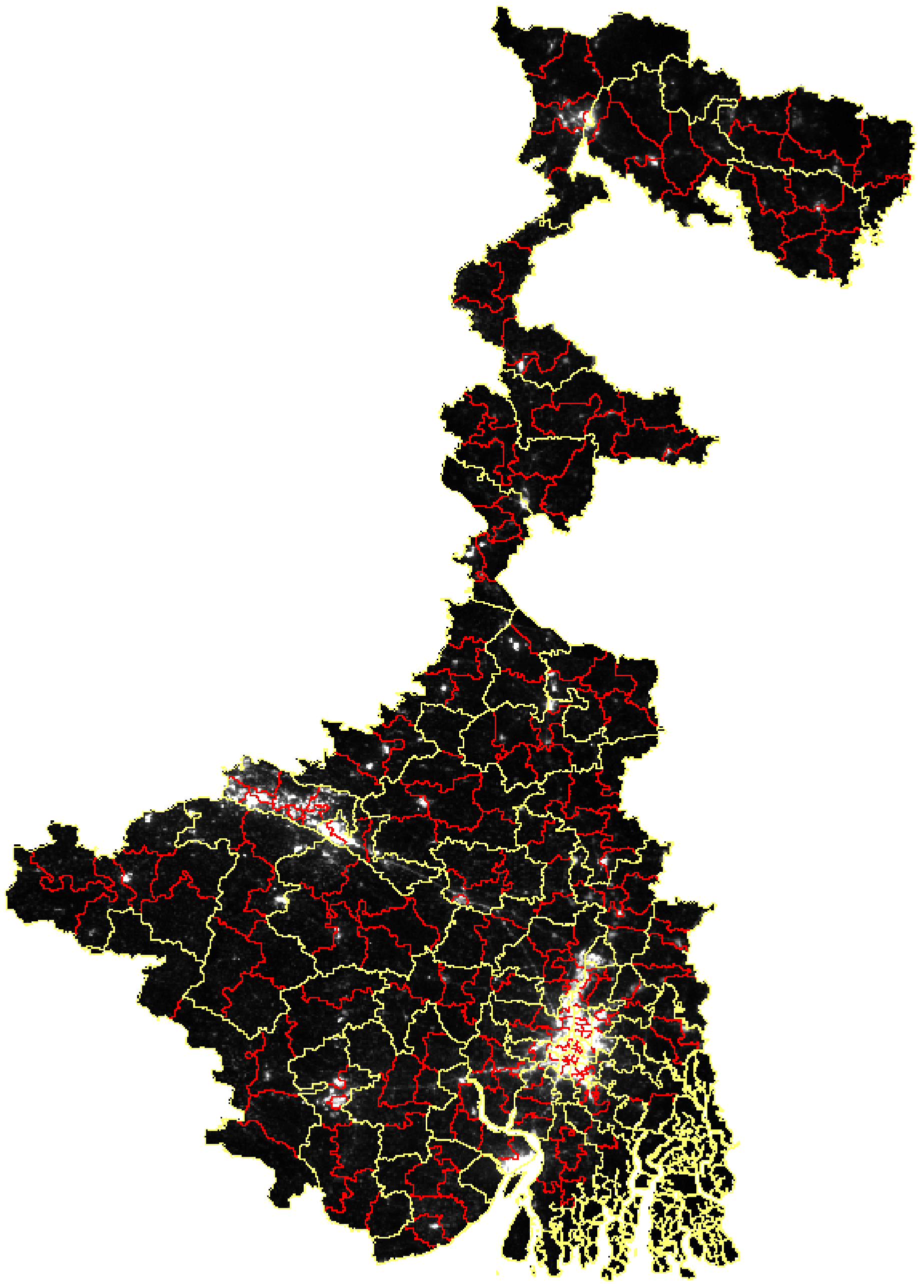
Table B1: Alternative Ways of Calculating Price Elasticities

	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 6. Col 1 shows the elasticity estimates from the running the IV strategy in Table 3 on the manipulated sub-sample (Section 5.2). Col 3 follows Table 3, 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.

C Additional Tables and Figures

Figure C1: Lights density mapped with assembly boundaries



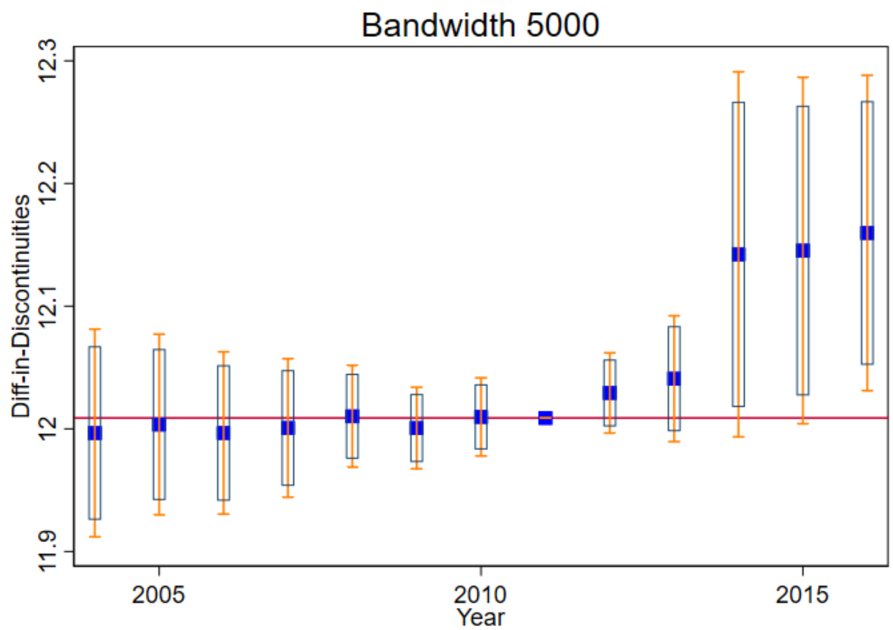
Notes: The figure shows boundaries of state legislative assemblies (in red), national-level parliamentary constituencies (in yellow), and data on nighttime lights density. For each legislative assembly, I calculate the mean value of light density to provide a measure of overall electricity consumption within that area.

Table C1: Discontinuity in Reported Consumption (Bandwidth Winning Margin=12,000 votes)

	Unit consumption in KWH				
Bwidth	12,000	12,000	12,000	12,000	12,000
Year	**2012**	**2013**	**2014**	**2015**	**2016**
			Residential (Rural)		
RD Estimate	-94.50***	-97.16***	-113.1***	-128.4***	-108.4***
	(25.00)	(21.34)	(22.22)	(23.33)	(24.32)
Observations	13,298	17,142	17,053	16,912	16,763
			Residential (Urban)		
RD Estimate	-172.6	-231.6**	-230.4**	-253.0***	-275.2***
	(108.3)	(97.13)	(95.67)	(93.19)	(90.51)
Observations	18,323	21,087	21,031	20,907	20,484
			Commercial (Rural)		
RD Estimate	83.51	38.08	58.04	-17.61	82.68
	(87.86)	(77.65)	(61.97)	(77.70)	(83.38)
Observations	5,151	6,649	6,576	6,544	6,495
			Commercial (Urban)		
RD Estimate	-334.3	-435.4*	-367.9	-369.9	-443.3
	(275.5)	(249.8)	(234.4)	(262.4)	(288.3)
Observations	18,178	20,990	20,616	20,737	20,287

Notes: Using the Calonico et al. (2015) RD methodology, 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 12,000 votes in terms of the the running variable, winning margin. Standard errors in parentheses clustered at electrical-feeder level. *** p<0.01, ** p<0.05, * p<0.1

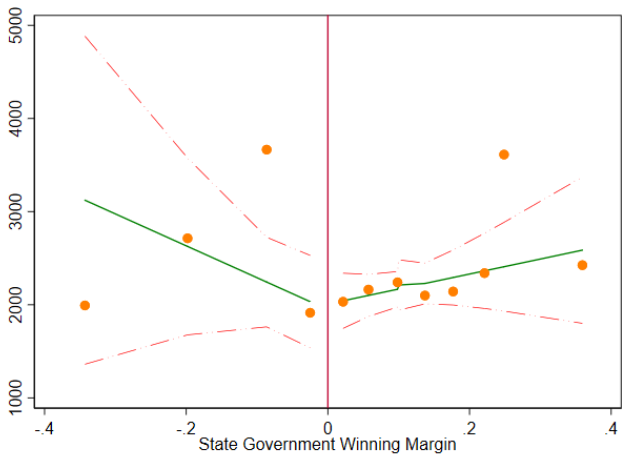
Figure C2: Levels of Nighttime Light Density: Difference-in-discontinuities Analysis



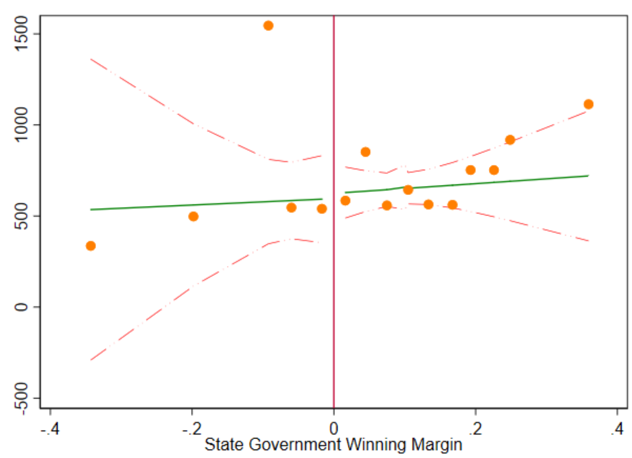
Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot the RD coefficients, and confidence intervals of errors clustered at the assembly level. The dependent variable is Log(light density). I plot coefficients over time and find a trend break after the 2011 election, with selectively greater electrification in areas where the governing party narrowly won.

Figure C3: Balance Across RD Cutoff - Census Village-level Characteristics I

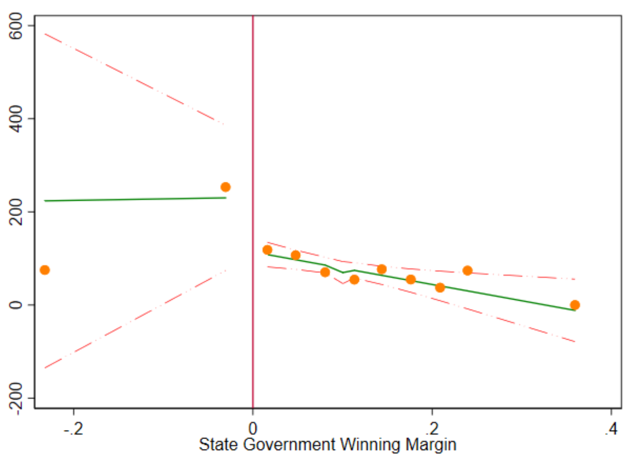
(a) Average Population



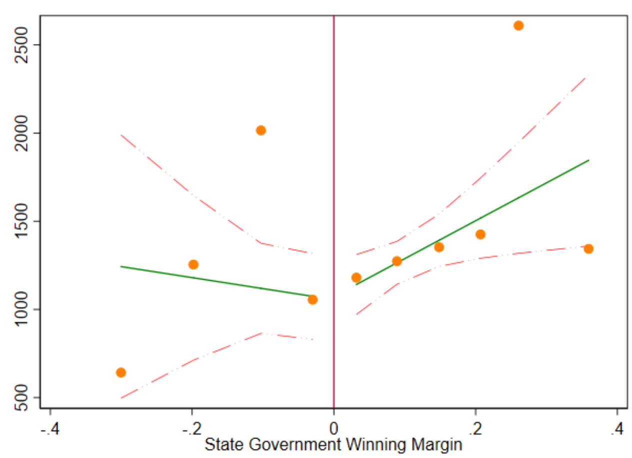
(b) Avg. no. from Scheduled Castes



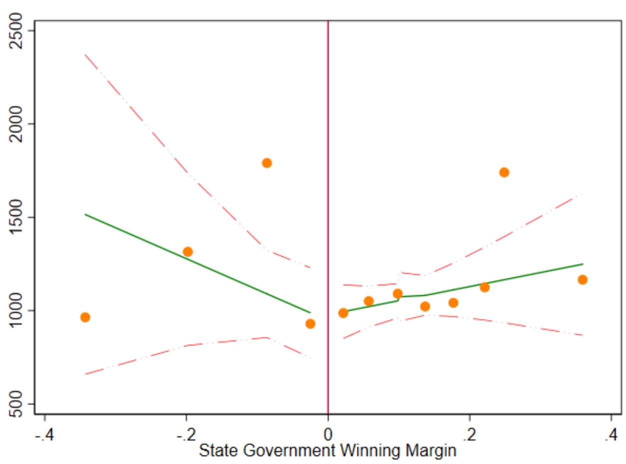
(c) Avg. no. from Scheduled Tribes



(d) Extent of Literacy



(e) Avg. no. of Females



(f) Avg. Population Under 6

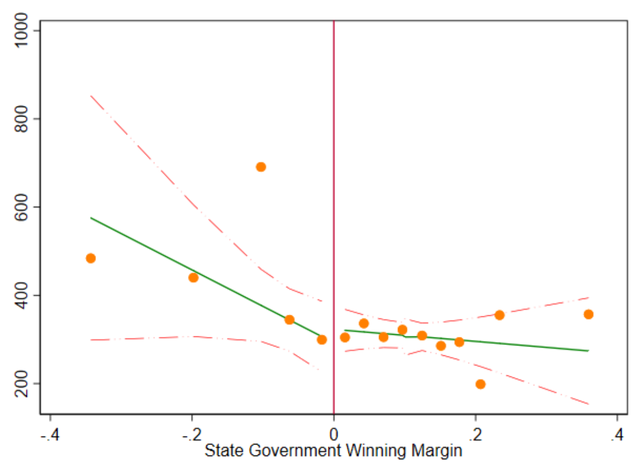
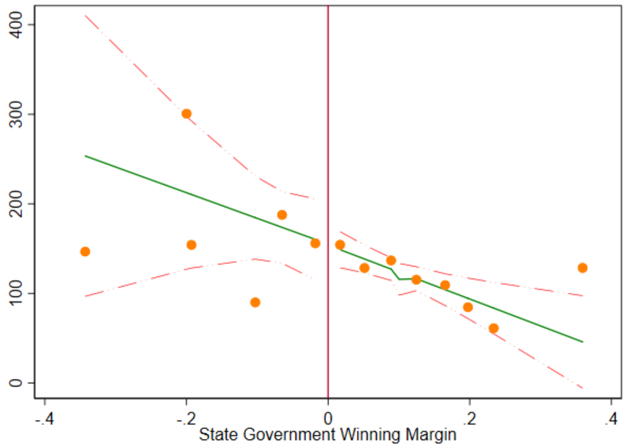
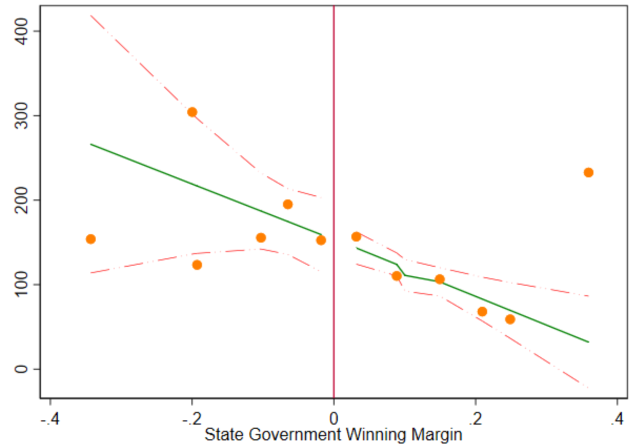


Figure C4: Balance Across RD Cutoff - Census Village-level Characteristics II

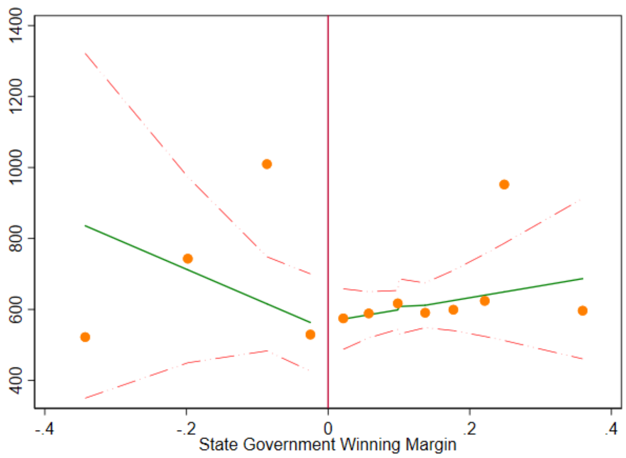
(a) Avg. No. of Agri. Workers



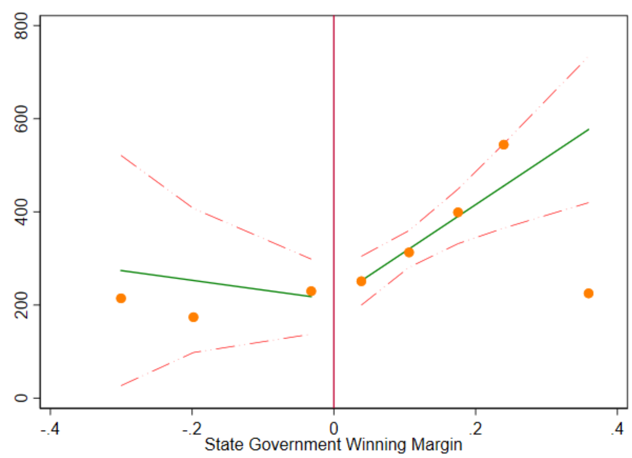
(b) Avg. No. of Cultivators



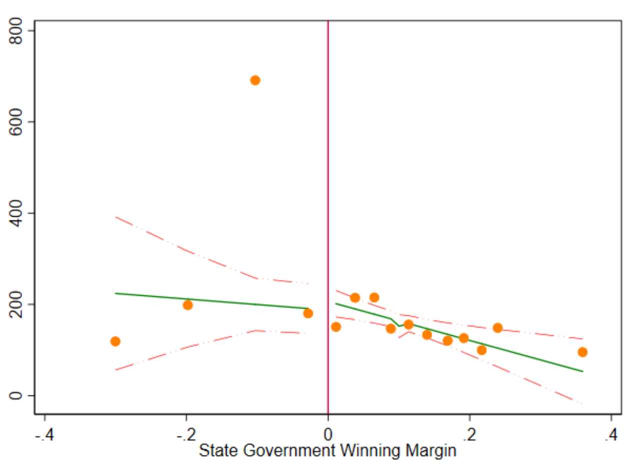
(c) Avg. No. of Manual Laborers



(d) Avg. No. of Other Workers



(e) Avg. No. of Female Workers



(f) Avg. No. of Marginal Workers

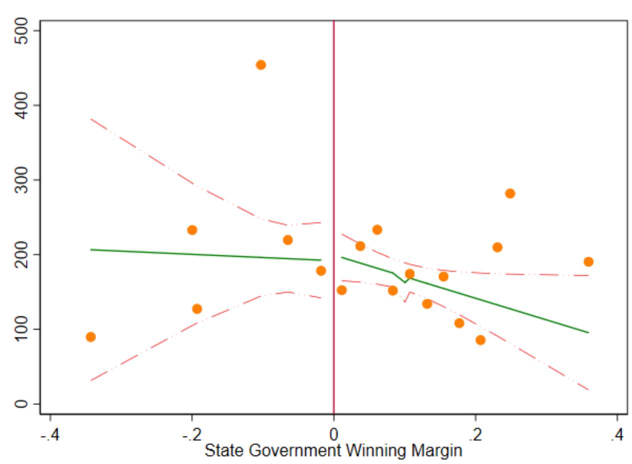
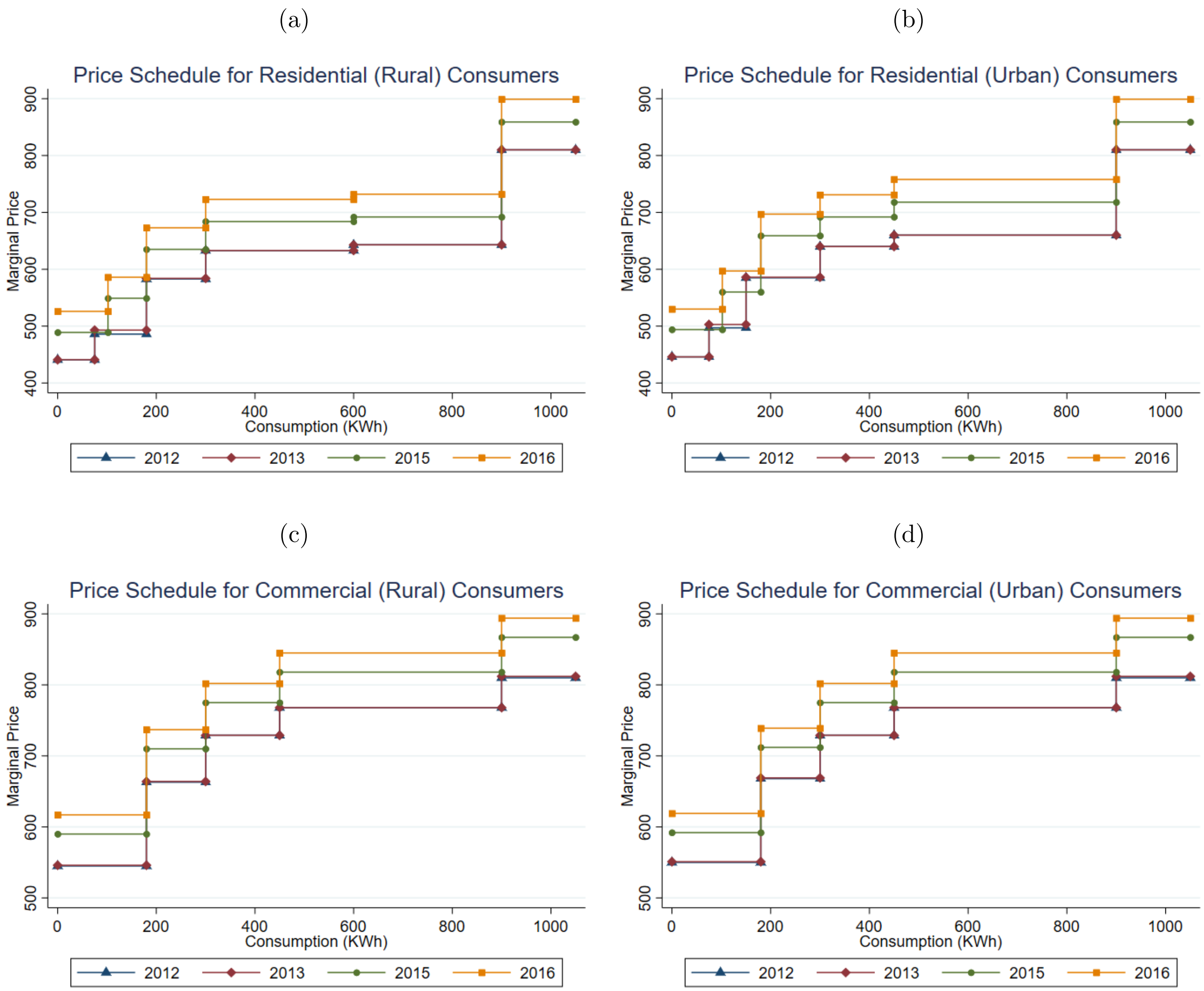


Figure C5: Change in Price Schedule Over Time



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.

Table C2: 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

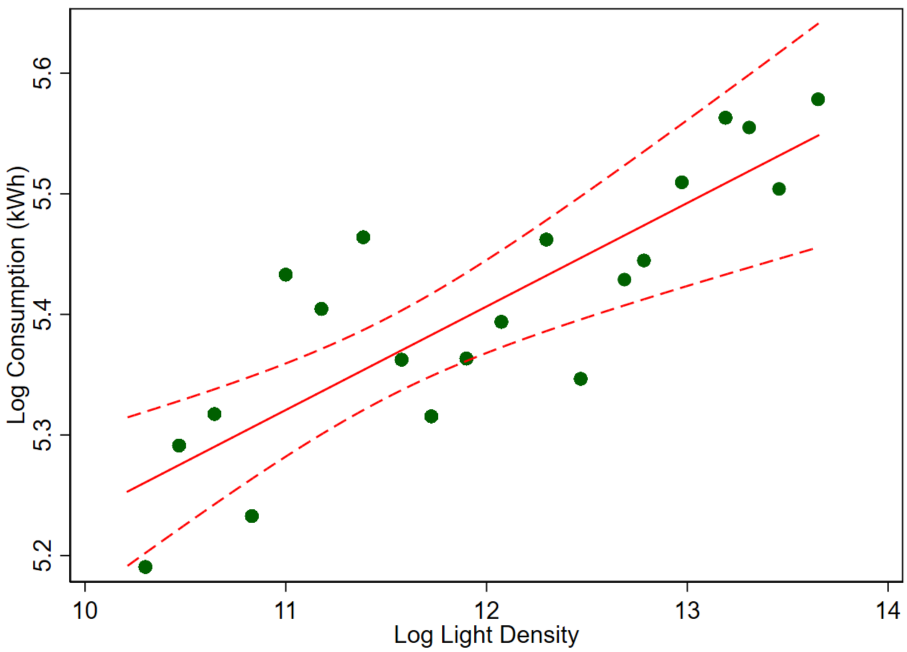
Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. This table shows results of the post-double OLS (Belloni et al., 2016) discussed in Section 6 Sub-section 6.3. Census data provides several village-level demographic characteristics which I use to build a model in order to predict 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 C3: Details for calculation of Welfare Loss and Gain in Consumer Surplus

	Consumer Class	**Residential (Rural)**	**Residential (Urban)**	**Commercial (Urban)**
Winning Margin Bandwidth=6,000	Consumer Base (winning areas near cutoff)	295,982	150,515	37,473
	Estimated under-reporting (KWh/year/customer)	138	379	547
Winning Margin Bandwidth=12,000	Consumer Base (winning areas near cutoff)	688,008	329,441	72,917
	Estimated under-reporting (KWh/year/customer)	108	248	385

Notes: This table shows the total number of consumers in the sub-sample of assemblies located near the RD cutoff, using two different bandwidths from the RD analysis, 6,000 votes on the lower end and 12,000 votes on the higher end. The estimated under-reporting figures are taken from Table 2 and Appendix Section C Table C1 .

Figure C6: Regression of consumption (KWh) on nighttime lights density



Notes: This regression, with year fixed effects, yields a coefficient of 0.08. From Figure 5, I infer an increase in consumption in response to the informal subsidy of 20%. Combined with the coefficient describing the relationship between nighttime lights and consumption, I conclude that the percentage increase in electricity consumption is 1.7%. Finally, I use consumption data for all consumer categories to make these calculations, as it is impossible to isolate the lights density for each consumer group individually.