Energy Provision and Economic Decision Making in Developing Countries

by

Meera Mahadevan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2019

Doctoral Committee:

Professor Ryan Kellogg, Co-Chair
Associate Professor Dean C. Yang, Co-Chair
Associate Professor Hoyt Bleakley
Assistant Professor Catherine Hausman

Meera Mahadevan
meeram@umich.edu
ORCID: 0000-0001-6812-1624

© Meera Mahadevan 2019

For Gaurav, who took care of me in every way, and kindly pretended that my outrageous behavior like wearing the same pajamas for days, eating ice cream for breakfast and unstoppably whining, was charming and pleasant. I've learnt what happiness is from the most difficult and stressful times in my life.

And, for Ellebow, my cat, who has been with me from the first day of my PhD to my last, and kept me grounded by reminding me that Phd-Shmee-h-D, looking after her was my only real job. Without her clever editing (deleting) from napping and walking on my keyboard, my dissertation would have just been long boring words.

ACKNOWLEDGMENTS

I'd like to thank my chairs, Ryan Kellogg and Dean Yang, and my advisors, Hoyt Bleakley and Catherine Hausman for their incredible mentorship and support throughout my PhD. I'm very grateful to Robyn Meeks, Achyuta Adhvaryu, Shaun McRae and Tanya Rosenblat for always giving me sound advice.

There are so many family members and friends to whom I owe this PhD. My parents, in-laws and siblings for putting up with my grumpiness and still cheerily helping out whenever they could. My research group, without whom I may have never finished this PhD - Pinghui Wu, Luis Baldomero, Oxana Azgaldova and Benjamin Glass. The strong female economists around me who I have learnt so much from - Daniela Morar, Catalina Franco, Tejaswi Velayudhan, Yeliz Kacamak and Salma Khalid. To Laura Flak, Lauren Pulay and Julie Heintz who were rescuing me from crises until the very end. To Austin ”Seth” Davis, Ben ”Dang” Thompson, Tupur Chatterjee, and Caroline D'Angelo for being sympathetic listeners through all of my ”mental gymnastics”.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

ABSTRACT

My dissertation focuses on two crucial issues that face developing economies how political and institutional features affect the public provisioning of electricity; and how major life transitions affect decision-making and behavior amongst the poor. My job market paper uses confidential electricity administrative data and satellite nighttime lights data, in conjunction with a close-election regression discontinuity design, to causally identify widespread political corruption in the Indian electricity sector, and quantify the net welfare implications for the economy. In other work, I study how electricity sector reforms in India, aimed at greater efficiency, affected the productivity of the manufacturing sector. I find that the structure of the distribution network in a state whether there is a single or multiple distribution companies is significant in determining the success of these reforms. My other work, with Catalina Franco, studies how transitioning from student life at university in Colombia, to working can affect economic decision making.

CHAPTER 1

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

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 address this corruption through policy, it is important to measure the size of the welfare losses, and compute demand elasticities. I develop a method to estimate 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.

1.1 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 identify whether electricity is used as a tool to garner votes. Next, leveraging a close election Regression Discontinuity Design (RDD), I show causal evidence that regions aligned with the governing party receive indirect electricity subsidies. Third, 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 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

[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*.

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 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.

[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.

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 $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 targeting for commercial rural consumers who have both sufficiently elastic demand and less electricity-using infrastructure.[4]

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

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). 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.

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

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.

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 1.2 provides background information on the Indian electricity sector, and the surrounding institutional and political structures. Section 1.3 discusses the empirical strategy for identifying evidence of political corruption in electricity billing. The next section, Section 1.4, describes the data sets used in this analysis. I show evidence of corruption in Section 1.5. Section 1.6 discusses the welfare implications of corruption by politicians, and Section 1.7 concludes.

1.2 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, and both the price and quality remain important election issues.

In this paper, I focus on West Bengal, a large Indian state where the transmission and distribution sectors are state-owned. This is with the exception of one privately owned firm which distributes electricity only to the city of Kolkata. 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 almost 17 millions accounts. In 2003, the central Electricity Act reforms led to a state regulatory commission, which is 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. In my analysis, I explore 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, so my findings can be extended to other states, and indeed other countries having similar institutions (e.g. Brazil, Bangladesh, Mexico, Sri Lanka and Kenya). Like these countries, 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 or not depends on the incentives faced by politicians, as well as whether such influence is feasible. There are number of reasons why politicians may want to control electricity supply. Election surveys, particularly in India have often found that electricity remains an important factor in election platforms (Chhibber et al., 2004). Politicians may try to win over new voters by offering cheaper or better access to electricity. However, there is a well documented pattern of patronage politics (Min, 2015; Nagavarapu and Sekhri, 2014; Sadanandan, 2012) that is favored in India, with politicians exerting greater effort in consolidating existing votes.

Chatterjee (2018) presents evidence that is consistent with the way I model politicians exerting effort 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 them 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).[5] 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 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. Another channel of influence that politicians could have is to 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, the 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

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

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.

1.2.1 Theoretical Predictions

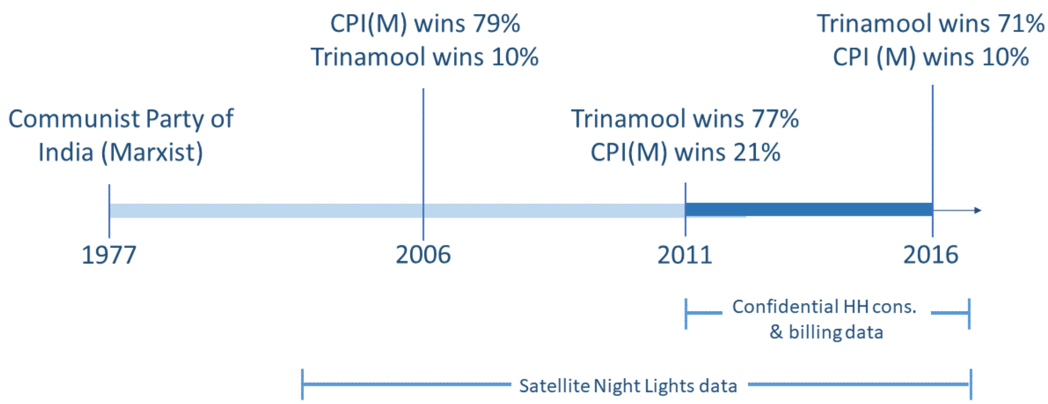
I develop a model in Appendix 1.8.1 to solidify my hypotheses, derive estimation equations, and motivate my welfare analysis. First, I derive a standard equation for electricity demand given a simple quasilinear utility function, increasing in electricity consumption with a constant elasticity for demand. 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 1.8.1 these predictions motivate using and are testable in a simple RD set up.

1.3 Close-election Regression Discontinuity Design

I apply a close-election Regression Discontinuity (RD) framework (Calonico et al., 2015; Imbens and Kalyanaraman, 2012; Imbens and Lemieux, 2008) 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 1.1: 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 Indian 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. 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 1.1). 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.

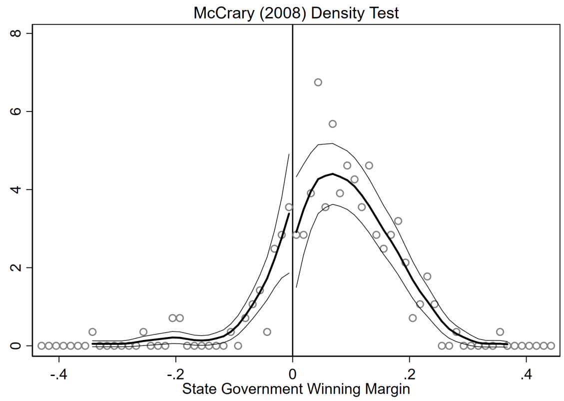


Figure 1.2: McCrary Test – density of winning margins at cutoff

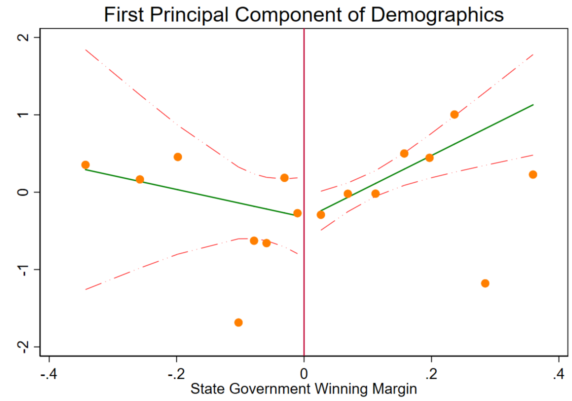


Figure 1.3: Balance on PCA of age, gender and caste

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 1.9, Figures 1.15 and 1.16.

In order to test the validity of the RD design, I run two main checks to ensure balance of the running variable and other characteristics on either side of the cutoff. On running a McCrary test (McCrary, 2008), I find no significant discontinuities in the density across the cutoff (Figure 1.2). Similarly, when comparing candidate characteristics such as age, gender and caste, a measure of the first principal component of the three also does not yield any significant discontinuities across the cutoff (Figure 1.2). This strengthens any causal claims that I make using this RD.

1.4 Data Description and Variable Definitions

1.4.1 Confidential data on Electricity Consumption and Billing

I obtain confidential 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.

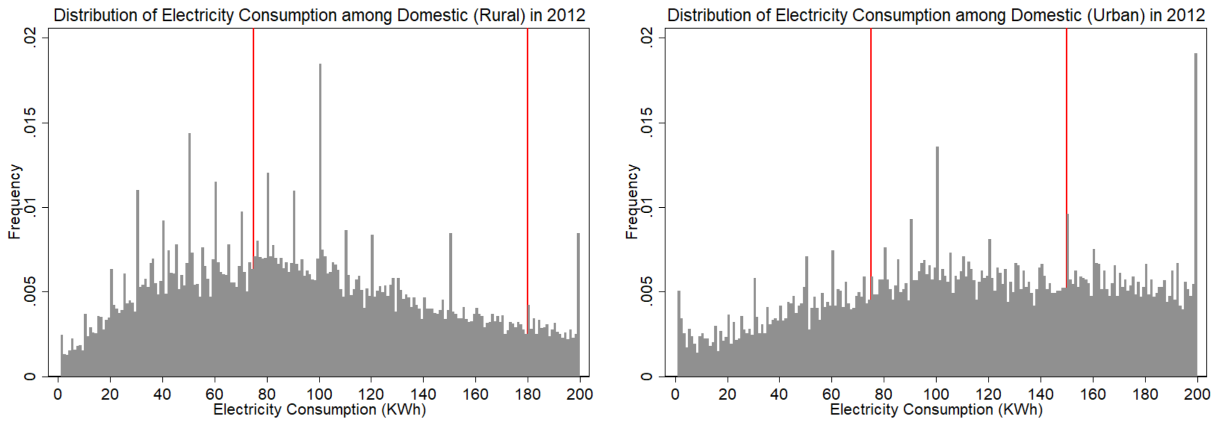
The utility is controlled by an independent regulatory board, the West Bengal State Electricity Regulatory Commission (WBERC). WBERC accepts proposals from WBSEDCL requesting tariff increases to meet their rising marginal costs of providing electricity. After reviewing these reports, WBERC sanctions 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 1.4, 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.

1.4.2 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

Figure 1.4: 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.

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 consists 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.[6] The close-election RD provides a neat way of identifying whether these billing items are systematically different in constituencies supporting the majority party.

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" political

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

Table 1.1: 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.

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 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 1.1 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.

1.4.3 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 1.13 in Appendix 1.9 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.

1.5 Empirical Evidence of Political Patronage

I present results from the RD analysis using measures of possible informal control in the electricity sector by political agents. I use average night lights density to capture whether the majority party provided more electricity to assemblies that voted for them. Exploring mechanisms, I test whether the party in power provided differentially cheaper electricity access to its voters. Given the measures I define, this translates to under-reporting of actual consumption, understating arrears owed, lower energy prices, and overstating subsidies in areas that voted for the MLAs from the majority party. In addition to that, I expect to observe a higher degree of deviation of the consumption distribution from the chi-squared distribution in these areas.

1.5.1 Average Nighttime Lights Density

In order to test for discontinuous electricity consumption, I run the following regression specification at assembly-level a, where vote-margin is represented by the net difference in the fraction of votes received by the winning party:

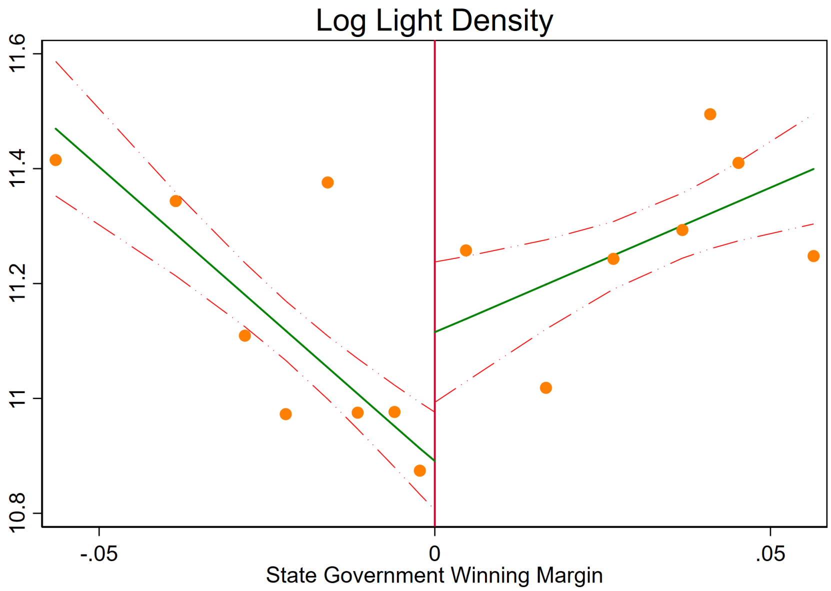
$$Log(Lights)_a = \beta \, \mathbb{1}(votemargin > 0)_a + f(votemargin)_a + \epsilon_a \tag{1.1}$$

I test for discontinuities in average light-density around the cutoff of being above a zero vote margin, allowing for the slope of the vote margin to vary at the cutoff. β measures the RD coefficient. Given that RD analyses result in an estimate of 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 1.8.1, parties concentrate their efforts as the expected payoff is the highest.

Figure 1.5 demonstrates that there is discontinuously higher light density for assemblies where the chief minister's party narrowly won. Given that there was balance across the cutoff on characteristics such as age, gender and caste of the candidates, this discontinuity in electrification points towards favorable 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 selective electricity consumption in the wake of the previous election as well, which was won by another party, the CPI(M). I run the following regression,

Figure 1.5: 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.

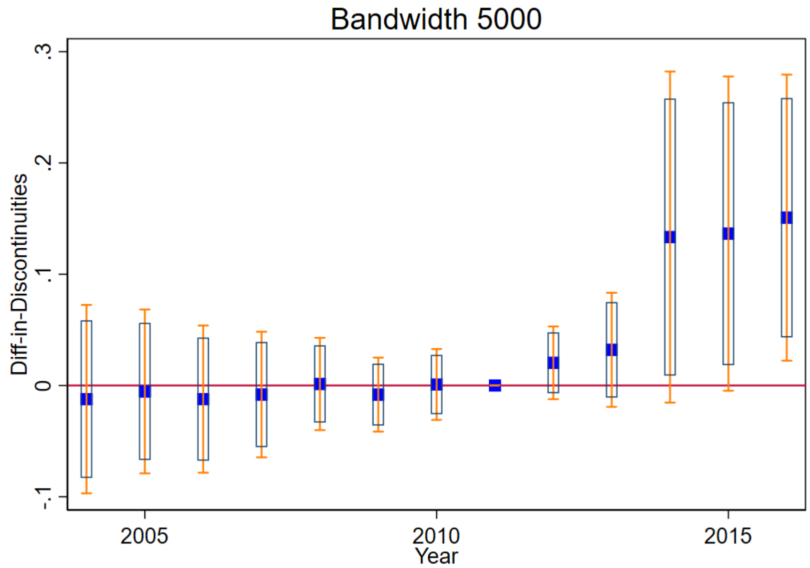
where β_t is the regression coefficient across years.

$$Log(Lights)_{at} = \sum_t \beta_t(\mathbb{1}(votemargin > 0)_a \times \gamma_t) + \gamma_t + f(votemargin)_a + \epsilon_{at} \tag{1.2}$$

In Equation (1.2), 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, and this is a meaningful falsification test. We should further expect that after 2011, there is an increase in this 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.

On graphing these coefficients in Figure 1.6, 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 just won, demonstrating clear evidence that there was favoritism shown in

Figure 1.6: 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 1.14 in Appendix 1.9

electricity provisioning in such areas.

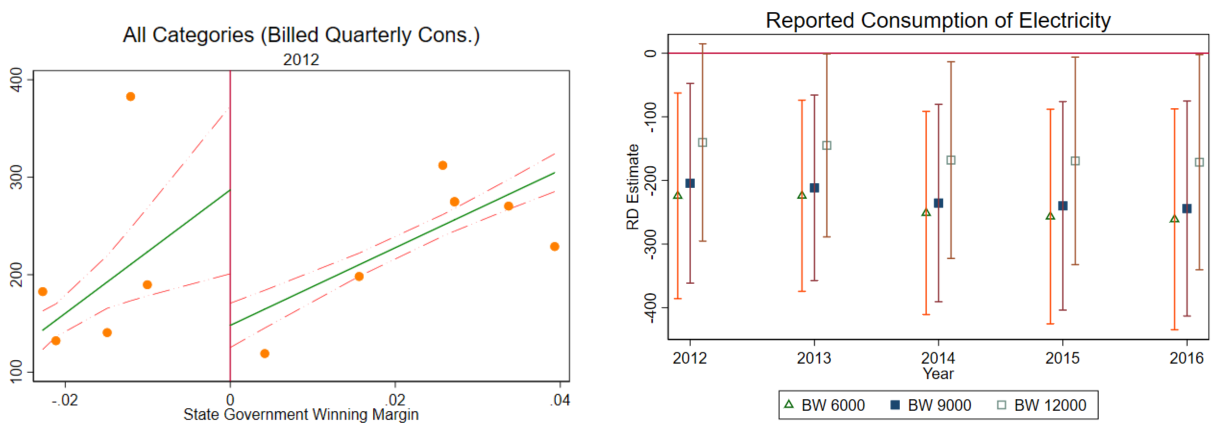
Neither night-lights density data nor utility administrative and consumption data alone can provide a complete picture of electricity use. However, a combination of both, like in this paper, is useful to get a clearer sense of the patterns in the data. Given the stark increases in the RD coefficient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

1.5.2 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 1.7 using consumption data on all consumer classes, including households, commercial users, public works and agriculture and irrigation. The only consumer class omitted from this dataset is industry, specifically high-tension consumers of electricity. This usually includes large factories. Therefore, aside from factories, which usually do not 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 in the government representatives. I run the following regression specification at the assembly level, where the left hand side includes various measures of data manipulation:

$$y_a = \beta \ \mathbb{1}(votemargin > 0)_a + f(votemargin)_a + \epsilon_a \tag{1.3}$$

Figure 1.7: 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.

The first variable I study is simply the reported level of consumption in swing-assemblies. Given that there is no observable discontinuity in baseline characteristics around the cutoff, there is no *a priori* reason for discontinuities to appear in reported consumption. In Figure 1.7, I find evidence that politicians in power did indeed selectively under-bill consumers in assemblies where the chief minister narrowly won. Using the utility reported consumption data, 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 night lights density. The logical conclusion, is that the utility consumption understates actual consumption. The magnitudes if this understatement are large, amounting to average discounts to constituents of over 50% of their regular bills. I calculate this magnitudes based on rough calculation using the large estimated effects of being in a constituency of the ruling party and the

Table 1.2: 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 *** $p<0.01$, ** $p<0.05$, * $p<0.1$

average electricity consumption at the cutoff in assemblies aligned with the opposition.

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 would 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. However, this scheme also included increasing electricity access in villages that already had grid access. Looking at four assemblies below and above the RD cutoff, the number of villages receiving some electrification through this sceme is very similar (5944 compared to 6024 in constituencies of the ruling party).[7] Given that a marginally greater number of villages in constituencies supporting the ruling party received some electrification 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 1.2).

In Figure 1.8, I find that the measure of distance from the expected chi-squared distribution 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. It is unclear 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 and for bandwidths both smaller and larger.

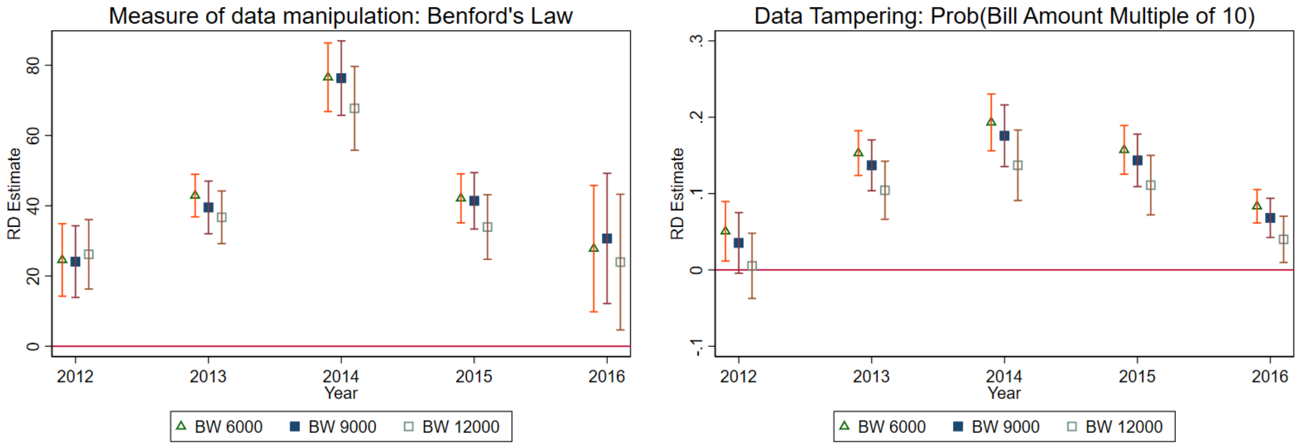
Next I examine the trends in the RD coefficient for potential manipulation of price variables – namely arrears and subsidy payments in Figure 1.9. I observe statistically significantly higher level of subsidies in winning swing assemblies, accompanied by a lower level of arrears. Taken together with 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-billling residential users, the politicians have effectively subsidized their electricity consumption and increased equilibrium electricity consumption.

1.5.3 Mechanisms of Political Patronage

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 change without detection. Among several vulnerabilities, Gulati and Rao (2007) identify the billing stage as sus-

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

Figure 1.8: RD Coefficients for Manipulation Outcomes Across Bandwidths



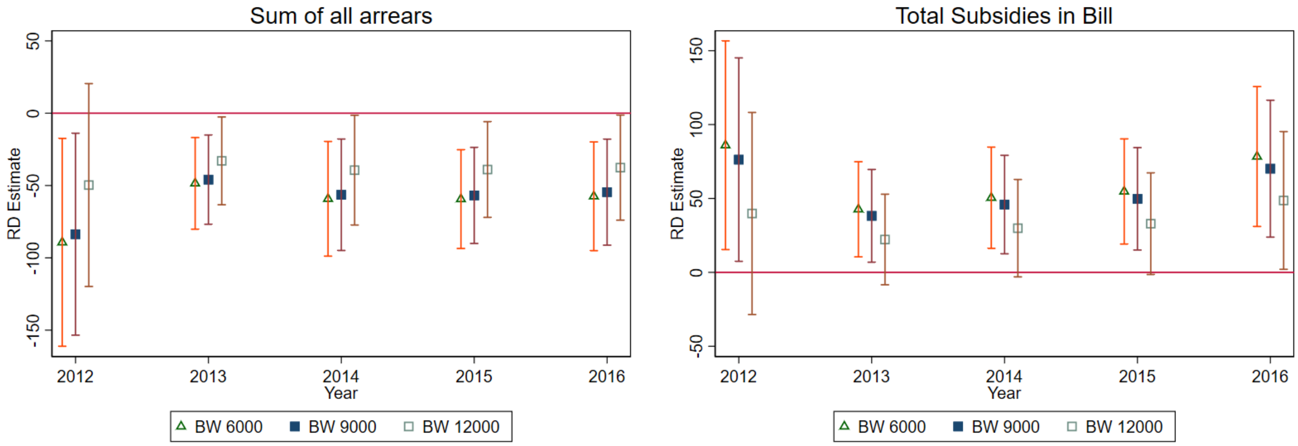
Notes: Using the Calonico et al. (2015) optimal bandwidths and bias-correct RD methodology, I plot coefficients across years for measures of data manipulation, and confidence intervals of robust standard errors clustered at the electrical-feeder level. Specifically I study the distance of the observed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

ceptible to political interference, with billing consumers at lower rates as a specific example in Indian utilities. An auditing study carried out by an electricity utility in Uttar Pradesh, another Indian state, identified significant political involvement in electricity distribution at local levels (Goenka, 2013). The inspectors who conduct meter readings are often external contractors. They report to a local CCC, which enters their reported consumption figures into the digital database. Given that I observe evidence of under-reporting of consumption, this appears a likely channel through which 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 of inspector readings, in order to allow local billing centers under the purview of MLAs from making low consumption imputations. These billing centers are dispersed all over the state, and it is in 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 amount of anecdotal evidence supporting this channel (The Telegraph, 2014; The Times of India, 2017; The Washington Post, 2012).[8]

[8]" "Many people known to support the ruling party are allegedly involved in hooking and tapping", a source said....

Figure 1.9: 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.

As a centrally mandated independent regulatory authority, it is hard to directly influence regulators to reduce electricity prices. They set tariffs after approving requests by the utility to do so, in response to the changing price of fuels that generate electricity, as well as changes to the composition of the generation stations supplying them. Chatterjee (2017) discusses evidence from interviews with regulators of 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 eliminate this possibility, given that the RD analysis provides me with relative changes. Yet, it is unlikely that politicians would expend effort in this way 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. Furthermore, the local MLA is the most likely candidate to pressure local customer care centers, and in losing regions, the MLAs are not aligned with the majority party.

Another possibility is that rather than manipulating data, electricity distributors actually provide greater access to electricity for consumers in losing assemblies, in a bid by the ruling party to

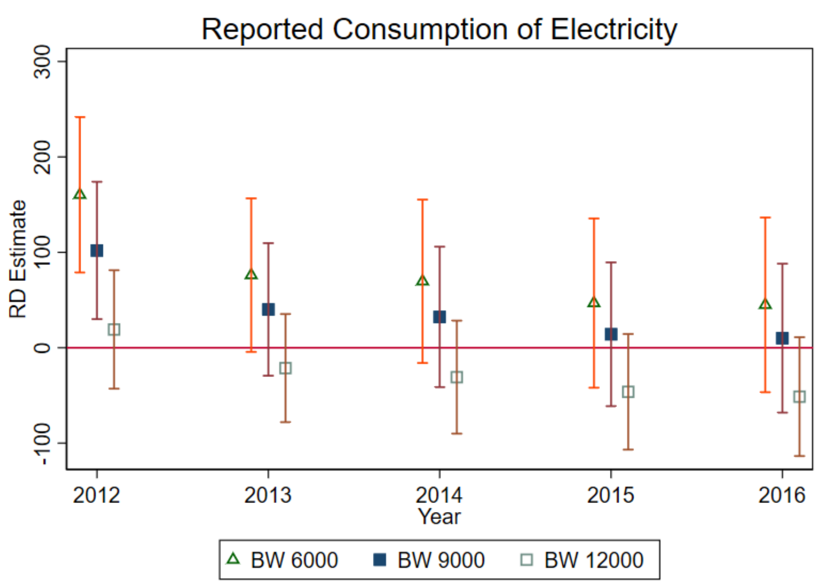
The chief minister had accused WBSEDCL of "callousness" and questioned the efficacy of such [anti-theft] drives." The Telegraph, July 31st 2014: *Power Theft Test for Mamata - State Utility to Seek CM's nod to Relaunch Crackdown.*

win over voters in losing assemblies. 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 just lost (Figure 1.6 & Figure 1.7). Together, the evidence points towards ex-post changes made in the reported consumption data, given that reported consumption is discontinuously lower in areas where the governing party narrowly won. Lastly, favoring voters in losing assemblies is unlikely to win new votes if the beneficiaries credit the Members of Legislative Assembly (MLAs) from the losing party, who is in office in those areas with providing better electricity access.

1.5.4 Falsification Tests and Robustness Checks

In order to test for the robustness of the results, I employ two main techniques. The first is to use multiple bandwidths in the RD analysis to check whether the results are consistent across bandwidths. I present these figures in Section 1.5.2 for the RD results on reported consumption, distance from the chi-squared distribution and bill items which are consistent across different bandwidths.

Figure 1.10: Placebo test: studying discontinuities in reported consumption using the winning and losing constituencies from the 2006 election

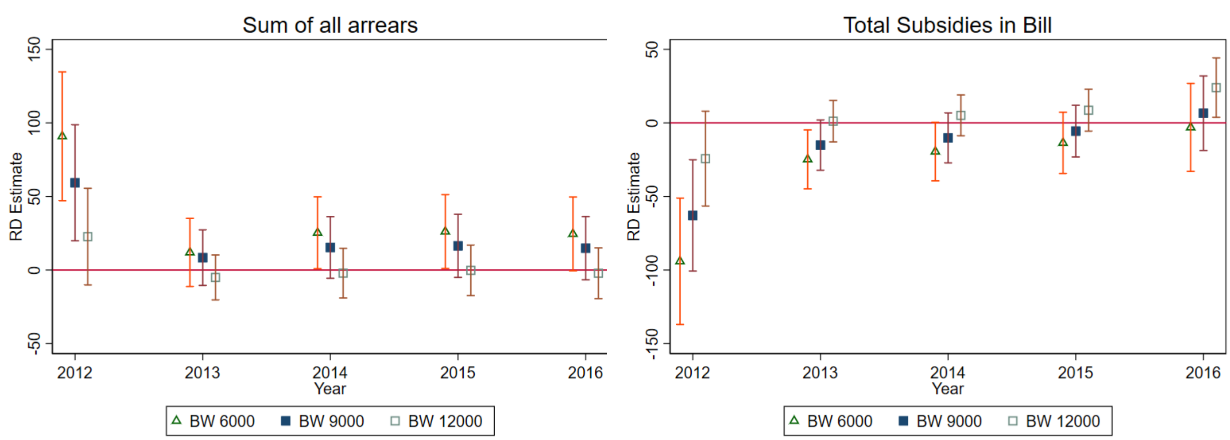


Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot RD coefficients for the reported consumption. The winning margin here is defined on the basis of legislative assemblies from the 2006 election, where the CPI(M) party won. This provides a falsification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

Then 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 1.10, I show the RD results analogous to those in Section 1.5.2. 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 only perhaps due to some persistence in corruption and manipulation that was occurring between 2006-11 under the previous government. After 2012, however, there is no detectable discontinuity.

Figure 1.11: Placebo test: studying discontinuities in bill items (arrears and subsidies) using the winning and losing constituencies from the 2006 election



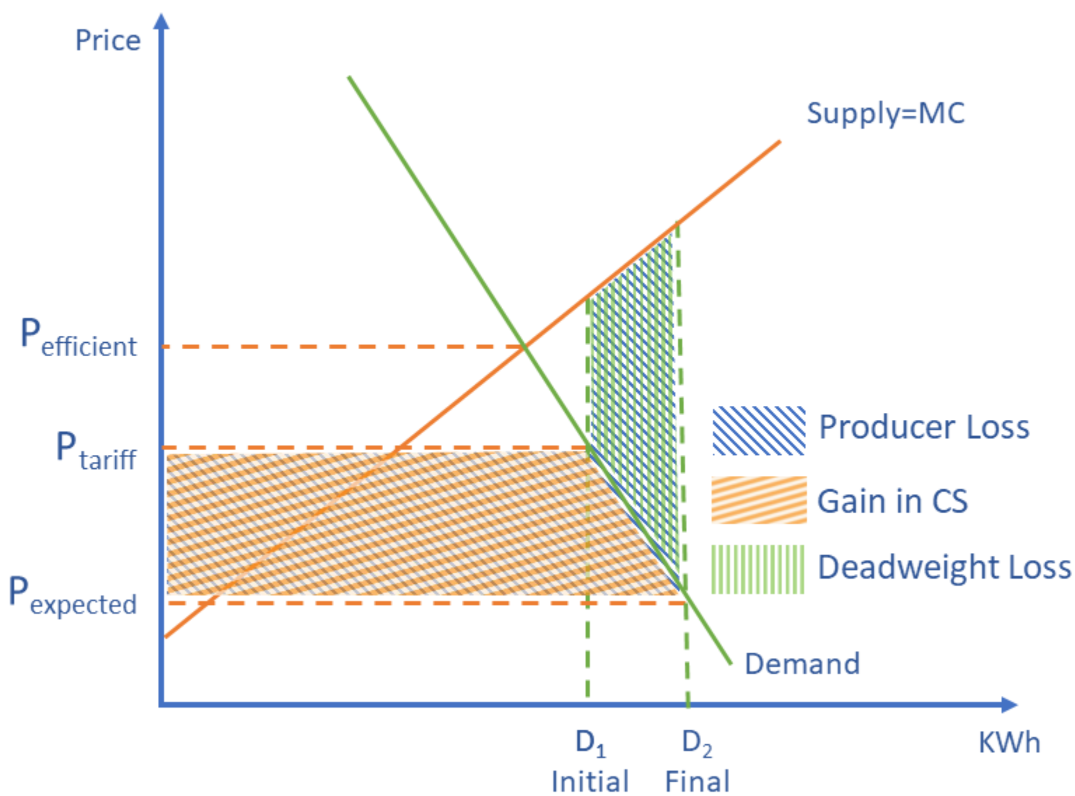
Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot RD coefficients for billing items including arrears and subsidies. The winning margin here is defined on the basis of legislative assemblies from the 2006 election, where the CPI(M) party won. This provides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

In Figure 1.11, I observe a similar pattern of no discontinuities. Again, there is weak evidence of a discontinuity in 2012, immediately post the 2011 elections, suggesting possible persistence in manipulation from the previous winning governing party. This suggestively 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 all politicians in government engage in actions to favor their constituents in terms of electricity access and price.

1.6 Welfare Consequences of Political Patronage

In order to design policies that can 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. This paper has 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.

Figure 1.12: 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.

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 1.12, 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 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 1.17 in Appendix 1.9 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 1.12 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 regression discontinuity estimates of the shortfall in consumption reporting at the cutoff, for each consumer group (Table 1.2) 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 1.12, 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.[9]

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 1.17 in Appendix 1.9 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.

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.[10] 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 end 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

[9]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).

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

in prior work. This analysis is described in Appendix 1.8.2.

1.6.1 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 1.5.2. The micro-level billing data allows me to observe the distribution of consumption for each assembly and I separate these assemblies into those where there is evidence of data manipulation, and those where there is no detectable evidence. This results in a dataset with 35 assemblies, for which I reject the hypothesis of data manipulation. For each assembly, I estimate the price elasticity of demand for each of the four consumer categories. The following specification, at the individual i and consumer category a level, is the simplest method of estimating elasticity but produces biased elasticities.

$$\log\left(Consumption\right)_{ia} = \delta_a \log\left(MarginalPrice\right)_{ia} + \epsilon_{ia} \tag{1.4}$$

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 endogeneity problem arising from an OLS specification, I use an instrumental variable strategy, exploiting 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 1.17 in Appendix 1.9).

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 elasticity. My specification is similar to Ito (2014), but leverages heterogeneity across individuals, and differential changes across price tiers, instead of creating a simulated IV.[11] 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. Con-

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

ditional on individual fixed effects, tier-by-month fixed effects, and consumer-category-by-month fixed effects, I instrument the marginal price $\log(MP)$ with the change in tariffs $\Delta \log(Tariff)$ across years. The first and second stage are respectively:

$$\log(MP)_{iamtcy} = \sum_{a} \gamma_{ac} \Delta \log(Tariff)_{amtcy} + \nu_{mta} + \zeta_{mac} + \eta_i + \varepsilon_{iamtcy} \ \ \forall \ \ a \in A \tag{1.5}$$

$$\log(Cons)_{iamtcy} = \sum_{a} \beta_{ac} \widehat{\log(MP)}_{iamtcy} + \tau_{mta} + \mu_{mac} + \omega_i + \epsilon_{iamtcy} \ \ \forall \ \ a \in A \tag{1.6}$$

I estimate β_{ac} separately for all constituencies a that lie in the set A of assemblies for which I reject the hypothesis of data manipulation. The four consumer categories c are RR (Residential Rural), RU (Residential Urban), CR (Commercial Rural) and CU (Commercial Urban). The regressions include individual fixed effects ω_i, 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 1.3 presents results by running the specification in Equations (1.5) and (1.6) for all assemblies with unmanipulated data. This table serves only to provide consolidated elasticities for the assemblies, but I use assembly level regressions 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 first stage F-stat is high, demonstrating instrument validity.

1.6.2 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.

Table 1.3: 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.

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 heterogenous means for various 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, the double-selection process whittles down the number of variables needed for predictive power, and the OLS regression is then run (separately for each consumer category) to predict elasticities for all assemblies. Table 1.8 in Appendix 1.9 shows the final model used in the prediction step.

1.6.3 Step 3: Predicting Elasticities for all Constituencies

Following the PDS OLS method, I predict elasticities for constituencies that showed evidence of data manipulation. Table 1.4 shows the mean values of the resulting elasticities. These differ from Table 1.3 because they represent the mean elaticity for each consumer category taking into account *all* assemblies, those with unmanipulated as well as manipulated data.

Table 1.4: 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 1.3 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 1.4 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 1.9 Table 1.6. The results in column 1 of Table 1.6 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 1.4). 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.

1.6.3.1 Targeting Inelastic Consumers

My model (Appendix Section 1.8.1) 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 would target areas with greater access to infrastructure follows from the fact that urban areas 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.[12]

When studying the effects of manipulation by consumer category in Table 1.2, I find no statistically significant discontinuity in reported consumption for commercial users in rural areas. Commercial rural areas have the most elastic demand, 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 1.4), 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 1.2 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 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.

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

1.6.4 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 1.2, 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 1.9), 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 1.7 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 Figure 1.5. 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 1.18 in Appendix 1.9 shows this relationship in graphical form. 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

Table 1.5: Utility Loss Calculations

Consumer Class	Producer loss (Million Rs./year)	Gain in surplus (Million Rs./year)
Residential (Rural)	Rs. 295.84	Rs. 101.27
Residential (Urban)	Rs. 323.77	Rs. 177.80
Commercial (Urban)	Rs. 111.41	Rs. 11.76
Total (Million Rs./year)	Rs. 731.01	Rs. 290.83
Total (Million Rs./year)	Rs. 3660.05	Rs. 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 1.9 Table 1.7. 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 1.9 Table 1.7. 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 1.4. $\Delta\log(prices)$ is $\Delta\log(consumption)$ in Appendix Section 1.9 Table 1.7 divided by elasticity ϵ.

in other parts of the world.

Based on the elasticity between night-lights and actual consumption, and the increase in night-lights at the cutoff, I obtain a value for $\Delta\log(Consumption)$ at the cutoff, and my results indicate 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 (1.21). I aggregate this figure based on the consumer base of the selected assemblies in Table 1.9. 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 sub-set 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.

1.7 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 cannot help identify 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 government 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. This reflects the fact that large numbers of them own generators and are able to substitute away from the grid. 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.

1.8 Model, Counterfactual Exercises, and Additional Tables and Figures

1.8.1 Model of Consumer Utility and Political Patronage

I present a simple model to solidify my hypotheses, derive the necessary estimation equations, and motivate the later welfare analysis. 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. The model contains decisions made by consumers or voters, and their political parties. It produces testable implications and generates estimation equations that I investigate empirically, with implications for consumer welfare.

1.8.1.1 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:

$$U_{ia} \;=\; v(z_{ia}) + c \;\equiv\; \frac{exp^{\beta x_{ia}}}{1-\epsilon} z_{ia}^{1-\epsilon} + c \tag{1.7}$$

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:

$$log\ z_{ia} = \frac{\beta}{\epsilon} x_{ia} - \frac{1}{\epsilon} log\ p_{ia} \tag{1.8}$$

In Equation (1.8), $\frac{1}{\epsilon}$ 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.

1.8.1.2 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:

$$vote = \begin{cases} 1 & \text{if } exp^{\gamma D_{ia}} v(z^*_{ia}) > \eta_{ija} \\ 0 & \text{otherwise} \end{cases} \tag{1.9}$$

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:

$$e_{ia} = p_{ia}^{-\alpha} \quad , \tag{1.10}$$

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):

$$v_{ia} = \frac{\left(exp^{\beta x_{ia}}\right)^{1+\frac{1}{\epsilon}}}{1-\epsilon} \, p_{ia}^{\frac{-(1-\epsilon)}{\epsilon}} = \frac{\left(exp^{\beta x_{ia}}\right)^{1+\frac{1}{\epsilon}}}{1-\epsilon} \, e_{ia}^{\frac{\alpha(1-\epsilon)}{\epsilon}} \tag{1.11}$$

Equation (1.11) 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:

$$\max_{e_{i1},\dots e_{1A}} \sum_a Pr\left(exp^{\gamma D_{ia}} v(z^*_{ia}) > \eta_{ija}\right) \quad s.t. \ \sum_a e_{ia} \leq E_i \tag{1.12}$$

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:

$$\max_{e_{i1},\dots e_{1A}} \sum_a \phi_a \left(exp^{\gamma D_{ia}} v(z^*_{ia}) - \mu_a\right) \quad s.t. \ \sum_a e_{ia} \leq E_i \tag{1.13}$$

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

$$\frac{\alpha\phi_a}{\epsilon} exp^{\left(\beta\frac{\epsilon+1}{\epsilon}x_{ia}\right)} exp^{\gamma D_{ia}} e_{ia}^{\frac{\alpha(1-\epsilon)-\epsilon}{\epsilon}} = \lambda \quad \forall\ a \tag{1.14}$$

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:

$$log\ e_{ia} = \frac{\epsilon}{\epsilon - \alpha(1-\epsilon)} \left[log\frac{\alpha}{\lambda\epsilon} + log\ \phi_a + \beta x_{ia} + \gamma D_{ia}\right] \tag{1.15}$$

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:

$$log\ p_{ia} = \frac{-\epsilon}{\alpha\left(\epsilon - \alpha(1-\epsilon)\right)} \left[log\frac{\alpha}{\lambda\epsilon} + log\ \phi_a + \beta x_{ia} + \gamma D_{ia}\right] \tag{1.16}$$

$$log\ z_{ia} = \frac{1}{\alpha\left(\epsilon - \alpha(1-\epsilon)\right)} \left[log\frac{\alpha}{\lambda\epsilon} + log\ \phi_a + \beta x_{ia} + \gamma D_{ia}\right] + \frac{\beta}{\epsilon} x_{ia} \tag{1.17}$$

1.8.1.3 Comparative Statics and Estimation Equations

Equations (1.15) through (1.17) 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 (1.15), with the help of a standard regression discontinuity equation:

$$\log\ e_{ia} = \delta_0 + f(\text{vote share of i in a}) + \tau_0 D_{ia} + \varepsilon_{ia} \tag{1.18}$$

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 (1.16) and (1.17) motivate regression equations on the form below, where we would expect $\widehat{\tau_1} < 0$ and $\widehat{\tau_2} > 0$:

$$\log\ p_{ia} = \delta_1 + f(\text{vote share of i in a}) + \tau_1 D_{ia} + \omega_{ia} \tag{1.19}$$

$$\log\ z_{ia} = \delta_2 + f(\text{vote share of i in a}) + \tau_2 D_{ia} + \xi_{ia} \tag{1.20}$$

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 (1.8), the demand equation, which I employ to estimate the price elasticity of demand $\frac{1}{\epsilon}$. Combining the measure of ϵ, with Equations (1.16) and (1.17), 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 (captured by x_{ia} that shifts out the taste for electricity). As these would most likely be urban areas, a testable implication is that urban areas are targeted more than rural areas.

The third equation of interest is Equation (1.11), 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:

$$\Delta log\ v_{ia} = (1-\epsilon)\Delta\ log\ z_{ia} = -\frac{1-\epsilon}{\epsilon}\Delta\ log\ p_{ia} \tag{1.21}$$

This measure of welfare, however, does not capture increases in utility losses, and perhaps the corresponding increases in taxes used to bail out the utility. To measure the extent of utility 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.

1.8.2 Estimating Elasticities - Counterfactual Exercise

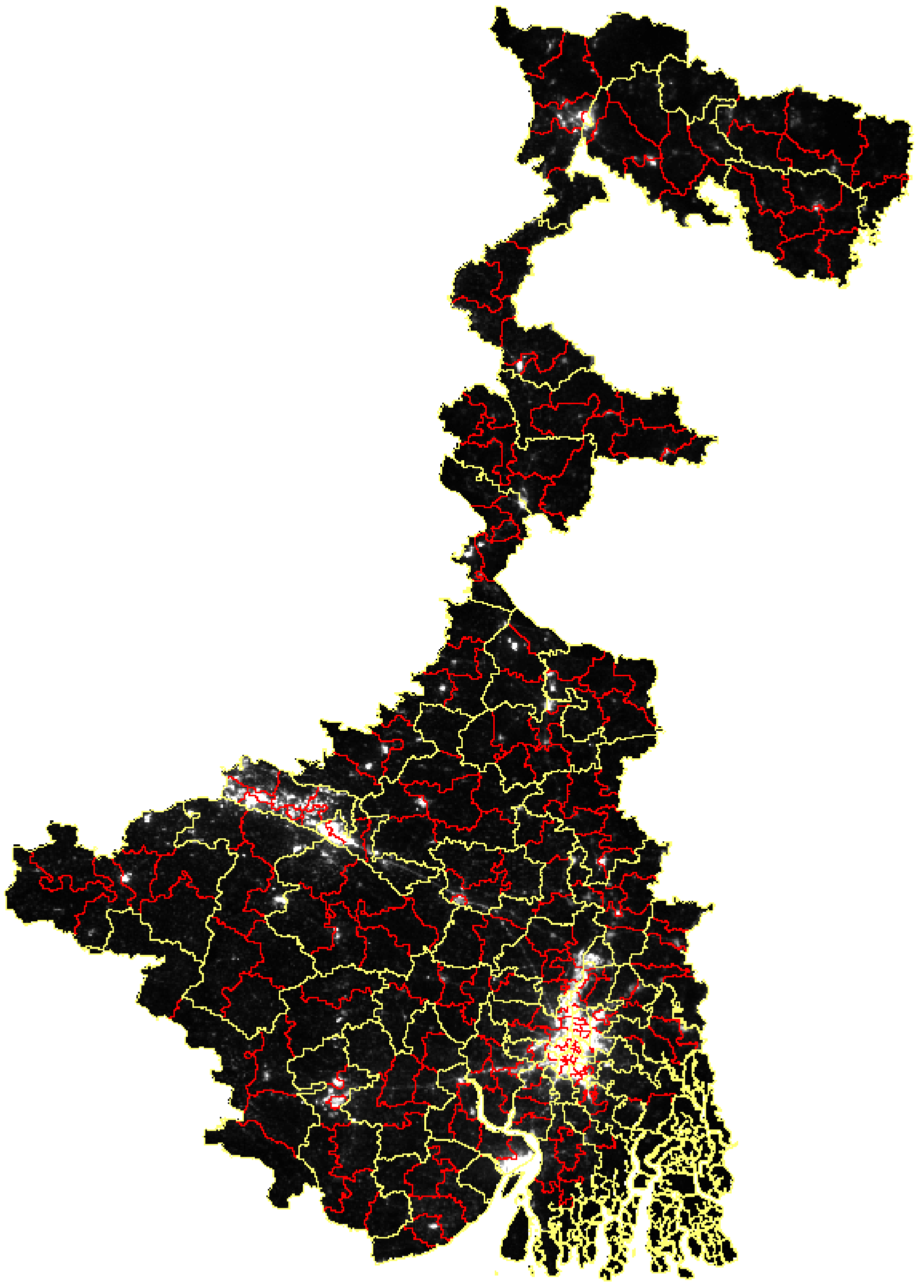
Table 1.6: Alternative Ways of Calculating Price Elasticities

	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* (0.228)	1.609*** (0.0596)	-0.240 (0.293)	-0.137 (0.0972)
Log Marginal Price Residential Urban	0.175 (0.220)	1.395*** (0.0574)	-0.666** (0.310)	-0.019 (0.0916)
Log Marginal Price Commercial Rural	-1.364** (0.535)	0.583*** (0.130)	-3.158*** (0.585)	0.0628 (0.155)
Log Marginal Price Commercial Urban	-1.800*** (0.460)	0.595*** (0.111)	-3.490*** (0.588)	-0.206 (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 1.6. Col 1 shows the elasticity estimates from the running the IV strategy in Table 1.3 on the manipulated sub-sample (Section 1.5.2). Col 3 follows Table 1.3, 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.

1.9 Additional Tables and Figures

Figure 1.13: Lights density mapped with assembly boundaries



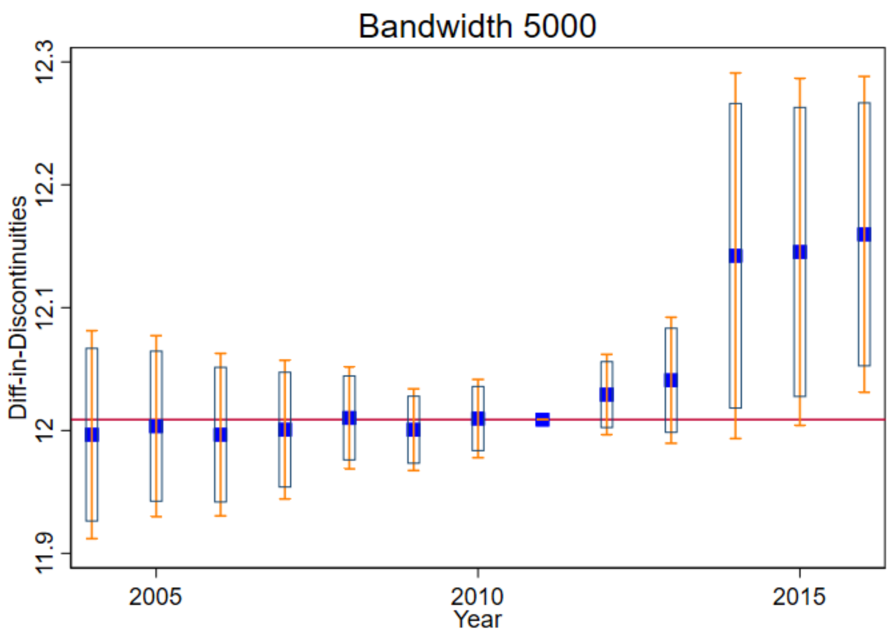
Notes: The figure shows boundaries of state legislative assemblies (in red), national-level parliamentary constituencies (in yellow), and data on nighttime lights density. For each legislative assembly, I calculate the mean value of light density to provide a measure of overall electricity consumption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

Table 1.7: 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

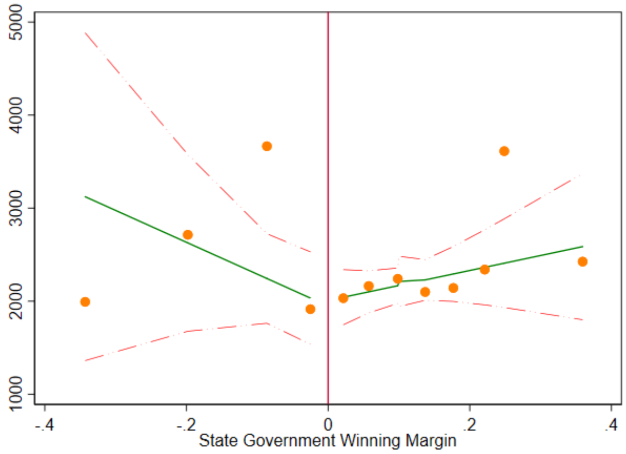
Figure 1.14: Levels of Nighttime Light Density: Difference-in-discontinuities Analysis



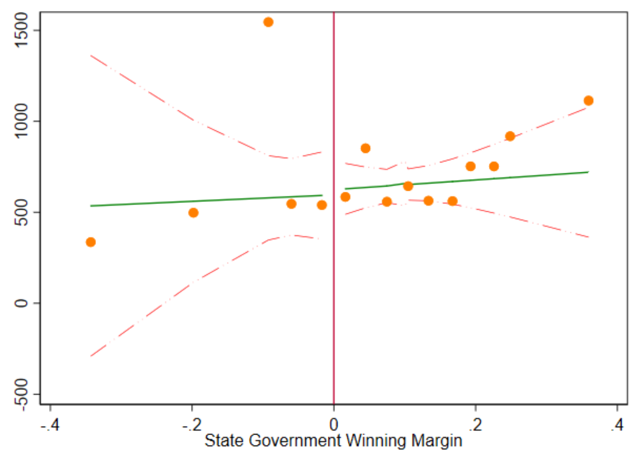
Notes: Using the optimal bandwidth and binning procedure described in Calonico et al. (2015), I plot the RD coefficients, and confidence intervals of errors clustered at the assembly level. The dependent variable is Log(light density). I plot coefficients over time and find a trend break after the 2011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

Figure 1.15: Balance Across RD Cutoff - Census Village-level Characteristics I

(a) Average Population

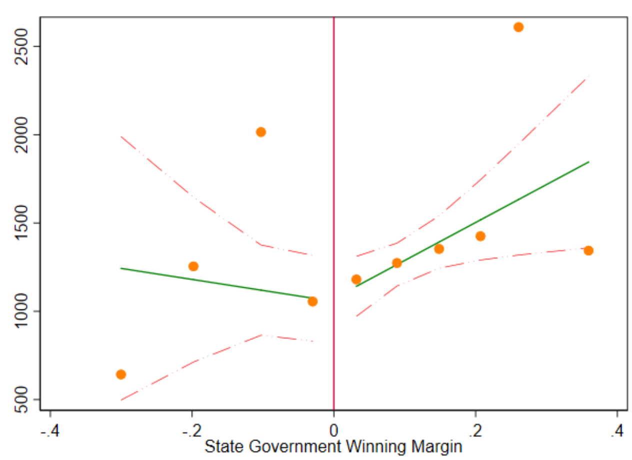


(b) Avg. no. from Scheduled Castes

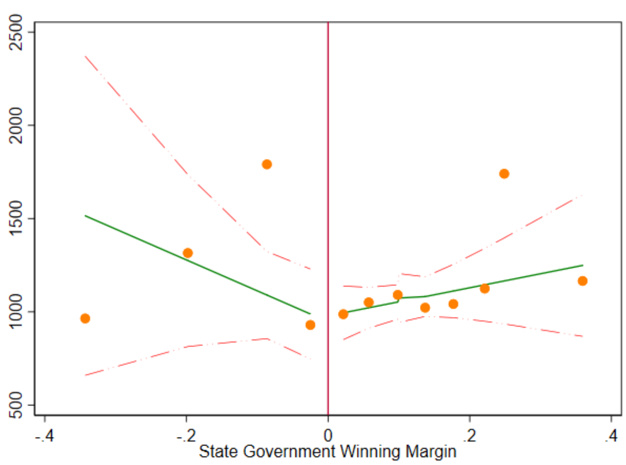


(c) Avg. no. from Scheduled Tribes

(d) Extent of Literacy



(e) Avg. no. of Females



(f) Avg. Population Under 6

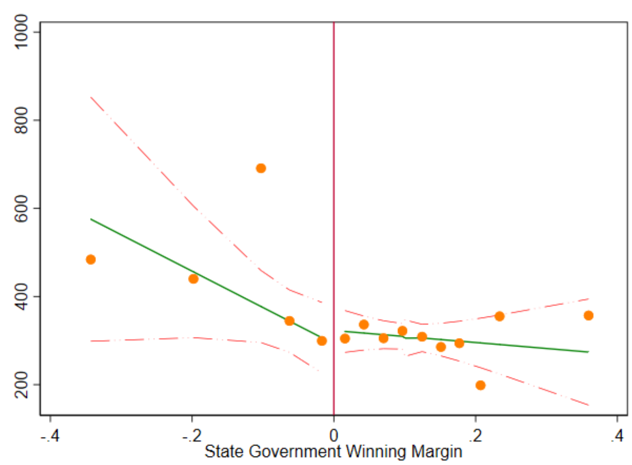
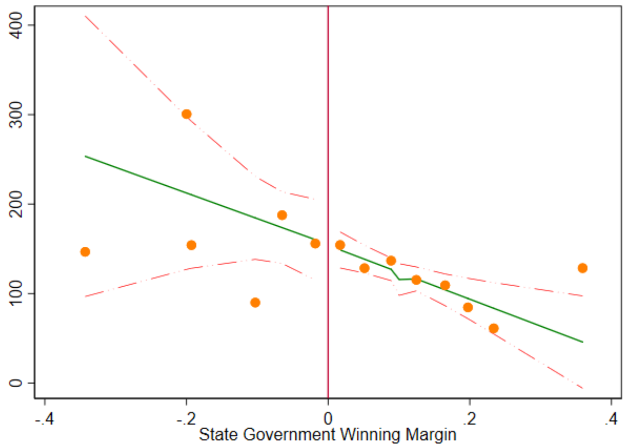
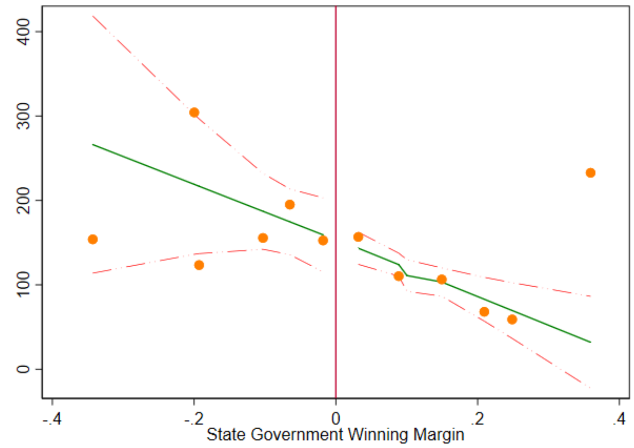


Figure 1.16: Balance Across RD Cutoff - Census Village-level Characteristics II

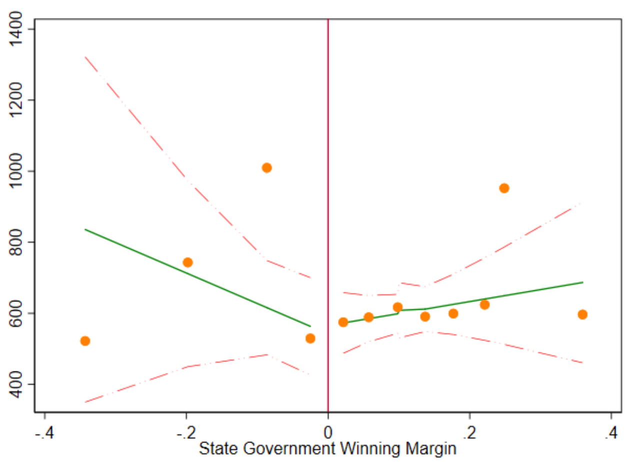
(a) Avg. No. of Agri. Workers



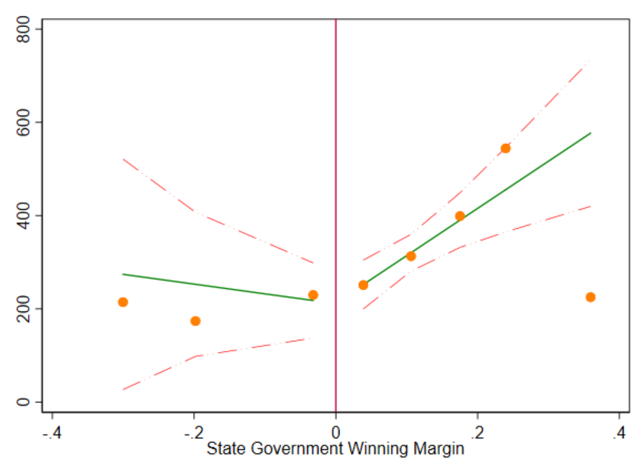
(b) Avg. No. of Cultivators



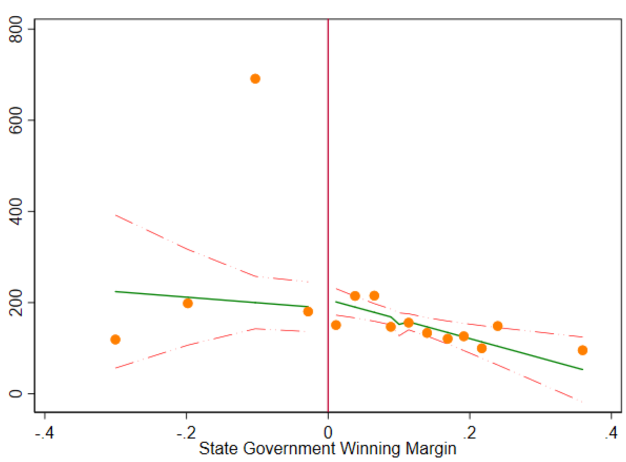
(c) Avg. No. of Manual Laborers



(d) Avg. No. of Other Workers



(e) Avg. No. of Female Workers



(f) Avg. No. of Marginal Workers

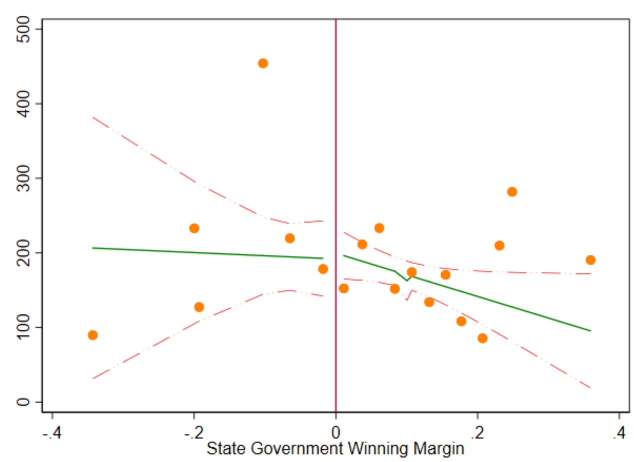
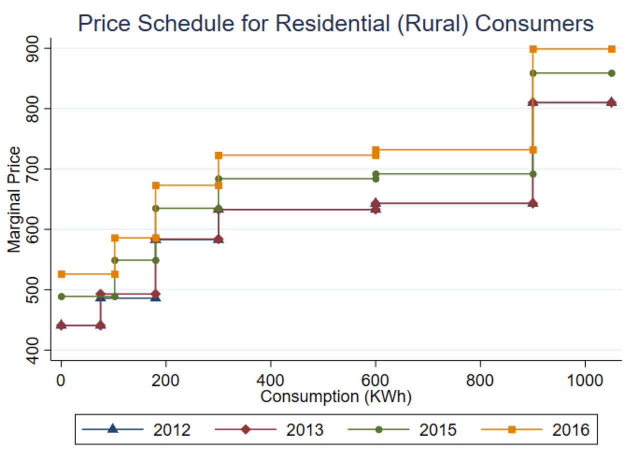
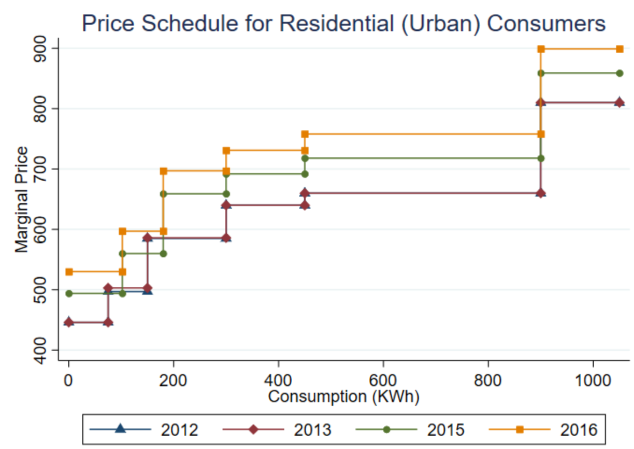


Figure 1.17: Change in Price Schedule Over Time

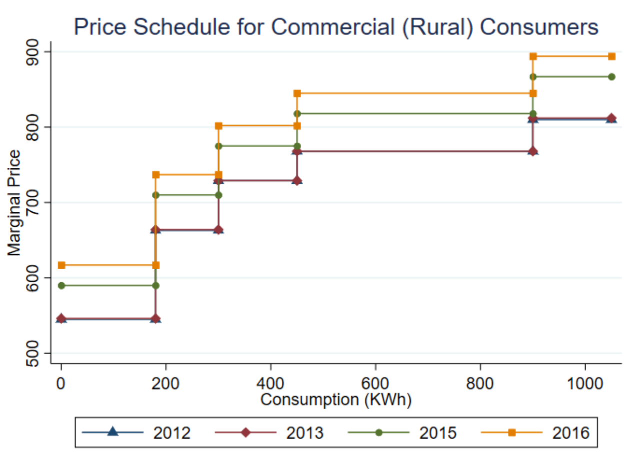
(a)



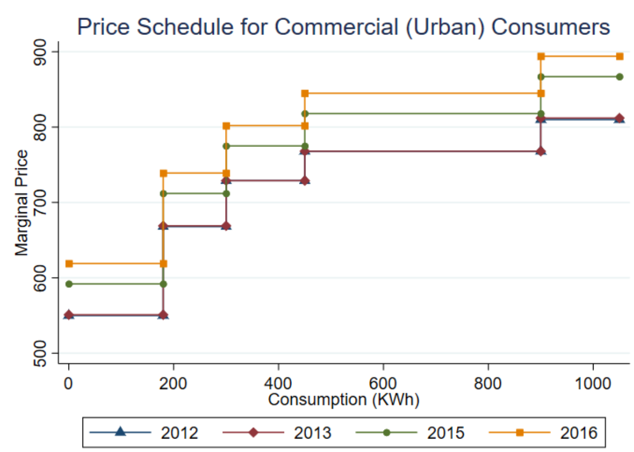
(b)



(c)



(d)



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 1.8: 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

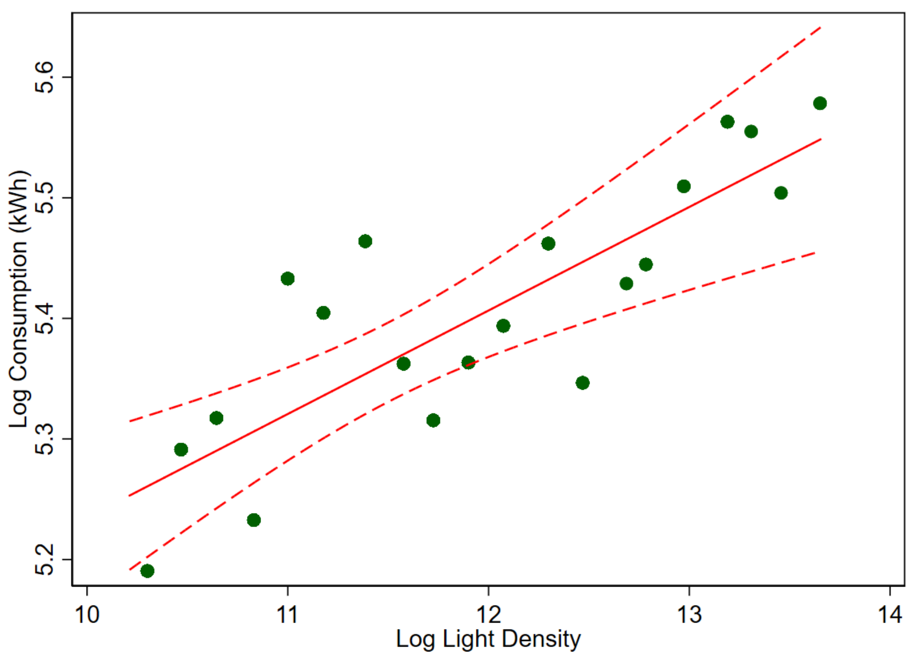
Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. This table shows results of the post-double OLS (Belloni et al., 2016) discussed in Section 1.6 Sub-section 1.6.3. Census data provides several village-level demographic characteristics which I use to build a model in order to predict 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 1.9: 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 1.2 and Appendix Section 1.9 Table 1.7 .

Figure 1.18: Regression of consumption (KWh) on nighttime lights density



Notes: This regression, with year fixed effects, yields a coefficient of 0.08. From Figure 1.5, 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.

CHAPTER 2

Behavioral Dynamics in Transitions from College to the Workforce

with Catalina Franco

We show the first estimates of changes in preferences and economic decision-making over major, expected life transitions. We elicit preference measures and cognition from Colombian students on the job market and those in lower years, and follow them across three stages: job search, receiving a job offer, and receiving a paycheck . Using a difference-in-differences setup, we find that job-market students become more altruistic, less impatient and perceive greater liquidity after receiving a job offer, despite this being an expected outcome. These effects dissipate by the paycheck stage, creating "windows of clarity", which may improve long-term decisions about healthcare, pensions or insurance. Furthermore, the results are driven by poorer students.

2.1 Introduction

Transitions from college to the labor force are both widespread and an important event for many across the world. Starting a first job involves making choices that impact an individual's life quality along the entire life cycle. Standard economic theory predicts that during the transition students' behavior, and their ability to make sound decisions, should not change in the absence of liquidity constraints. Yet, empirical tests of standard consumption-smoothing hypotheses in other contexts find that it not to hold (Shapiro and Slemrod, 1995; Stephens, 2003). How decision-making changes around the time that agents make important life choices can have long lasting impacts on lives. Importantly, standard models assume preferences to be constant Andersen et al. (2008); Dasgupta et al. (2017); Harrison (2005); Roszkowski and Cordell (2009) except when individuals experience an unexpected shock or in response to events Cho et al. (2018); Necker and Ziegelmeyer (2016). Yet, even though the transition from college to the labor force may involve predictable outcomes for individuals, it may change the decision-making processes and underlying preferences of economic agents. We are among the first to document changes in students' behavior along key stages of this transition with substantial implications for policy.

Given that several high-stakes decisions such as on health, retirement benefits and life insurance are made after joining a first job, it is clear that understanding the changes in behavior and decision making over this period is important. Despite this, there is very little work on understanding the dynamics involved. To empirically assess whether there are behavioral changes in the transition from college to the workforce, we compare experimentally-measured tasks for students in their last semester at college and on the job market, to students in lower years at a university in Colombia. This comparison takes place across three main stages: job search (baseline), receiving and accepting a job offer, and receiving the first paycheck. A unique feature of our empirical design is the fact that we can measure the effect of receiving a job offer alone, without conflating it with receiving a paycheck. We distinguish between these two stages to separate the effects of an increase in liquidity from a paycheck and higher responsibilities from a job, with the resolution of uncertainty that a job offer would entail. The effects we observe from the job offer stage therefore reflect the resolution of uncertainty surrounding the *details* of the job, even as there is little uncertainty about *receiving* a well-paying job, as 84.3% of engineering students graduating in 2014 from this university were in formal jobs a year later.[1]

Our empirical strategy uses the fact that students on the job market are similar to students in

[1] Source: Ministry of Education, Colombia.

previous years except in characteristics (such as age) that are unrelated to their performance on economic decision-making tasks. We pick similar students in previous semesters (i.e. those who are involved in day-to-day college life) to match the gender, major-choice and economic background of last-semester students to most accurately mimic what would have happened to last-semester students had they not finished college. With this comparison group, we perform difference-in-differences (DID) analyses of our outcomes of interest. We compare the job-offer and first-paycheck stages to the baseline to provide evidence on changes in decision-making along these two important stages of the shift to the labor market.

The experimental measures we elicit include social preferences (measured via dictator and ultimatum games), time preferences Andreoni and Sprenger (2012), ambiguity aversion Tanaka et al. (2014), risk aversion Eckel and Grossman (2002); Tanaka et al. (2010), and cognitive measures (IQ test, cognitive reflection test, numerical Stroop task, Flanker task). We also measure perceived financial situations, psychological factors, and credit and degree of independence from parents through survey questions.

We first demonstrate how resolving uncertainty around the details of a highly likely event can affect decision-making. Our main results show that subjects graduating from college, relative to similar students in earlier years, become more altruistic, less impatient, and perceive greater financial liquidity and psychological well-being after merely receiving a job offer, despite it being an almost certain outcome of graduating from a prestigious university in Colombia. These effects are observed before subjects receive their first paycheck, ruling out certain mechanisms, like the easing of liquidity constraints. Moreover, these effects dissipate by the time subjects receive their first paychecks, creating brief "windows of clarity", which are ideally suited to make long term decisions such as about healthcare, pensions, and life insurance.

Once these students receive a paycheck, they no longer give more in dictator and ultimatum games, and perform worse on cognitive tasks relative to students in the comparison group. After being paid, these students might have to undertake greater responsibilities and may have more factors to juggle, leading to a greater cognitive load Mullainathan and Shafir (2013). Indeed, they are more likely to report paying for most of their expenses in the after paycheck stage lending credence to their added responsibilities. The feelings of less worry, tiredness, depression and frustration, as well as the observed changes in economic preferences, dissipate by the time they start getting paid.

Given the variation we have across our subjects in terms of socioeconomic status, we also provide suggestive evidence on how poverty correlates with the results we find. Students who come from relatively poorer backgrounds drive the results we observe during the after-offer stage - specifically experiencing less depression and greater perception of liquidity. Comparable students in lower years who do not have the benefit of a job offer and also come from poorer backgrounds are found to be far more depressed and perceive less liquidity. Richer students appear to have similar preferences whether or not they have a job offer. Interestingly, giving in dictator and ultimatum games, and cognitive performance is not associated with socieconomic status.

Our findings open up avenues for future work focusing on these "windows of clarity" for people from poorer financial backgrounds. Our findings on the correlation between poverty and behavior are consistent with studies such as Spears (2011), who shows how poor participants found decision making depleting, Mani et al. (2013), who show that poor participants perform worse in a cognitive task, Haushofer et al. (2013), who find that experiencing a positive shock makes people more patient. This result also fits in with the literature on the behavioral effects of positive affect such as Erez and Isen (2003); Ifcher and Zarghamee (2011); Kirchsteiger et al. (2006); Oswald et al. (2015). Our study is able to strengthen this analysis by using a real instance of a an expected positive affect as opposed to a lab induced mechanism, providing a realistic application of these results.

Rational models of economic agents treat preferences and decision-making ability as constant. Yet, we highlight how external context substantially affects these factors. Knowing this can have significant implications for policies that encourage the optimal timing of when decisions are made. For instance, in our context, decisions on pension and retirement plans should be encouraged to be made before the work actually begins.

Our results also highlight the implications of psychological factors as mediators in important economic decision-making and in experimentally-elicited preferences. Experimental measures of time and social preferences are used extensively in economics and are increasingly being added to large-scale national surveys. Self-perceived emotions could act as important mediators to the experimentally measured outcomes and not accounting for them could lead to erroneous conclusions about some of their economic decision-making behavior. For instance, last-semester students appear to become less risk averse when they get a job offer, but this effect disappears after controlling for psychological factors. What is important to note is that while observed changes in risk behavior were simply reflections of changes in these emotions, changes in social and time preferences stand even after accounting for the change in these variables.

We see three main contributions of our work. First, even though the transition from college to the work force is an important and ubiquitous one, this paper is one of the few studying decision-making and preferences along this path, and showing that they change substantial over the period. Second, we speak to the few studies showing how risk, time and social preferences measured through experimental games and survey measures can change in response to idiosyncratic shocks (Chuang and Schechter, 2015), and show that such changes can occur even over expected life transitions. Finally, we highlight the importance of psychological factors as mediators in economic decision-making.

The rest of this paper is organized as follows: Section 2.2 presents the background of our setting and the research design. Section 2.3 presents details about data collection and the experimental measures we use. Section 2.4 discusses the difference-in-differences results. Section 2.6 concludes.

2.2 Background and Research Design

We study economic decision making along the transition from college to the labor market with students from a large public university in Colombia, recruited primarily from the Engineering School. In general, students from this university, and engineering students specifically, have very good prospects in the Colombian labor market, given the academic quality and prestige of the university they attend. In other words, obtaining a job is a very likely outcome. According to Colombia's Ministry of Education Statistics, about 84% of engineering students from this university are in formal jobs one year after graduation.

This context is unique in that we have similar subjects in terms of education level along with natural variation in socioeconomic status. This university admits students on a "need-blind" basis, solely determined by an admissions exam, and admitted students are at the very top of the admission score distribution. This guarantees that our subjects will have similar cognitive ability, giving us results that will presumably not be affected by big differences in baseline cognition or level of education. Furthermore, students are charged tuition rates that are monotonically proportional to their family income. We use self-reported tuition rates to determine the financial status of students.

In order to study the transition from school to the labor market, we use the fact that students in their last semester of college will experience certain changes as they go on the job market, while comparison students not on the job market will not, but still serve to control for other factors related to being in college. We differentiate between the two groups by calling them "last-semester students" and "comparison students", respectively. In order to construct an appropriate comparison group to isolate the differential effects on students on the job market, we use students from lower years. Besides age and variables that are naturally different when one is farther along in college, we expect that students about to graduate and in previous semesters are similar in most observed and unobserved characteristics. More importantly, for our identification purposes, as we will show, these students have similar baseline trends in preferences and measures of economic decision making. Our group of comparison students is selected to closely match the last-semester students on gender, major and economic background at baseline. In Table 2.1 we provide descriptive statistics across these groups.

We invited engineering students at this university to participate in a research study about economic decision making. Students signed up in April, 2016 using an online form with questions about demographics, major, current semester in the major, GPA, tuition, socio-economic measures at the household level, whether they work, and perceived probability of finding a job between April and October, 2016 for those who plan to graduate in August. We selected students in lower semesters to equal the number of students in their last semester who signed up. We did so by stratifying on gender, major, and tuition above or below the median.[2] Our resulting number of observations at baseline is 363 of which 178 (or 49.1 percent of) students were in their last semester of college.

The different stages in the research design are summarized in Table 2.2. We collect data from five waves of surveys taking place at the recruitment stage and at three points in time according to Table 2.2: (i) Sign-up survey; (ii) Two surveys at baseline (during the last semester before graduation of last-semester students); (ii) One survey each approximately after receiving and accepting a job offer[3]. We refer to this as "Round 2"; (iii) One survey approximately after starting in the new job and receiving at least one paycheck. We refer to this as "Round 3".[4] Participants responded to all surveys online on roughly the same dates between April, and December, 2016. All surveys

[2] The median tuition per semester in our sample is COP 600,000 which is equivalent to around US$200.

[3] Notice that the timing of this survey does not coincide with graduation in August. Because all surveys were conducted either well before or well after graduation, we do not think that our results are driven by the volatility in emotions associated with graduation.

[4] Some last-semester students either received a paycheck by the time we administered the survey for Round 2 or never reported a paycheck despite indicating having got a job offer. We move their survey responses from Round 2 to Round 3 in the former case, and from Round 3 to Round 2 in the latter.

except the sign-up questionnaire contained the same tasks, although in cognitive tests we varied the questions or worded them differently every time to reduce the role of memory. For other tasks, remembering would be harder because each task involved many choices.

A particular feature of this context is that, in general, students cannot perfectly smooth consumption by taking loans that will help them keep their standard of living constant before and after graduating from college. In our sample, only 6 percent of students have credit cards with a credit limit above $1,000 (the equivalent of about 1.5 times the expected monthly salary in their first job after graduation) and about 10 percent have loans over $5,000 at baseline. Not only is it difficult to obtain a credit card, there is also a cultural lack of comfort with the use of credit cards. Therefore, there are both administrative as well as cultural barriers to borrowing for consumption smoothing. The effect of receiving a paycheck somewhat alleviates this credit constraint. The job-offer stage, on the other hand, would not, and we investigate whether merely receiving an offer affects decision-making performance.

2.3 Experimental tasks

2.3.1 Tasks and incentives

In order to measure economic decision-making, we administer a battery of tasks for students to complete. Our online surveys contained three types of questions: economic decision-making tasks, social preferences, cognitive tests, and questionnaires about socioeconomic situation, consumption of durable goods, debt and credit, stress, and salary expectations. In addition, we ask last-semester students about job offer and paycheck dates. We follow previous studies in the design of these tasks and provide details below. In order to incentivize students, a task was picked at random, and a payment was assigned as per the students' choices or performance. For instance, if the task chosen was risk aversion, the gamble picked by the student is played and they are paid accordingly. If the chosen task is the Raven's Matrices, students are rewarded based on the number of correct answers. Such a payment structure incentivizes students to reveal their choices thruthfully and to exert effort to maximize performance. Further details on the tasks can be found in the appendix.

The order in which tasks appeared to participants was random although they always came before the questionnaire about psychological and stress measures, expenditures, salary expectations, and relevant dates of job offer and paycheck. No feedback about performance after each survey was given to participants. At the end of the survey, one of the tasks was selected for payment. The computer followed the instructions communicated to the participants regarding the selection of task for the final payment. In each survey excluding the recruitment survey, prizes ranged from the equivalent of US$7 to US$57. The mean prize across all three rounds of surveys was $30. To have an idea of how much this represents to students, $30 is worth about 40 bus tickets or two weeks of restaurant meals.

Social Preferences: We administer a set of dictator and ultimatum games, where subjects choose to allocate a portion of an endowment of 20,000 pesos ($7) to other participants or to a foundation that helps children in Bogota. In the ultimatum game, the recipient student can then choose to accept or reject. See appendix 2.8.1 for more details.

Time Preferences: We adapt the elicitation task presented in Andreoni and Sprenger (2012). Subjects are given a pre-specified monetary amount and are required to allocate it between two dates: earlier and later (see appendix 2.8.2.1). Unlike Andreoni and Sprenger (2012), subjects allocate the endowment of 50,000 pesos ($17) in increments of 1,000 pesos. We study the number of non-monotonic decisions made by a student [5], the "impatience" of students [6] and the present-biasedness.[7] See appendix 2.8.2 for more details.

Risk Preferences: We elicit risk aversion using the Eckel and Grossman (2002) measure. Students pick one gamble to play from a list of 6 gambles, and their choice indicates their degree of risk-aversion. To analyze risk choices we use the risk lotteries from Tanaka et al. (2010) (see example and details in Appendix 2.8.3.1). We also use the risk lotteries task in order to understand consistency in student choices and obtain Prospect Theory parameters. Commonly, in tasks designed as a list of lotteries, subjects pick a particular column and then switch permanently to the other, and the row where they switch provides information about their risk preferences. However, we allow them to switch between columns more than once to understand if they make inconsistent risk choices. See appendix 2.8.3 for more details.

Ambiguity Aversion: This is the preference for known risks relative to unknown risks Camerer

[5]Non-monotonocity in this case refers to allocating an increasing amount of money to an earlier period as the interest rate for the later period payout goes up.

[6]We examine how many times out of 16 allocation choices students allocate the entire initial amount to the earlier period.

[7]We calculate the probability of assigning a greater amount of money to the earlier period, and weighting it by the interest rate in that particular choice.

and Weber (1992); Ellsberg (1961), and is measured using a task based on Tanaka et al. (2014) in which subjects must choose between a gamble whose outcome objective probabilities are known relative to one in which they are unknown. See appendix 2.8.4 for more details.

Cognitive Tests: In terms of cognition, the bandwidth theory proposed by Mullainathan, Shafir and coauthors implies that scarcity (of time or resources) affects cognitive functioning which may compromise decision making (Mani et al., 2013; Mullainathan and Shafir, 2013; Shah et al., 2012). To measure different dimensions of cognition we use tasks such as a Raven's matrices-type IQ test, the Cognitive Reflection Test (CRT), the Flanker's task, and the numerical Stroop test. See appendix 2.8.5 for more details.

2.3.2 Trends and Levels for Outcomes of Interest

For our difference-in-differences analysis it is necessary for both our last-semester students and students in lower years to have similar trends in these measures; an assumption we provide empirical support for in the form of parallel pre-trends. However, we also show that the *levels* at baseline are also similar, except in variables such as age and degree of independence. Table 2.1 presents the means of variables collected at sign-up for the two groups. As expected, we see that last-semester students are older, and more likely to be employed and to have accumulated more work experience at the time of the survey (during their last semester of college). The two groups have very similar means in their demographic and schooling variables, as well as in their salary expectations upon graduation.

To construct our difference-in-differences (DID), we collected information from students over a total three rounds. From the two surveys in the baseline period (round 1), we are able to establish parallel trends. In Table 2.3 we show, for each variable of interest, the change in the means between survey 1 and 2 of the baseline, and the p-value of the point estimate from a DID regression involving the two baseline surveys. Given that virtually all variables exhibit parallel trends in the period before the changes associated with the transition to the labor market take place, we believe that the difference-in-differences analysis is a valid method to analyze our data.

Round 2 refers to the survey administered to last-semester students after they obtain a job and this may be interpreted as having one's uncertainty about the job being resolved. Students in lower years were also administered a survey to coincide with time, to generate a comparison group.

Round 3 refers to the surveys administered to last-semester students after they receive their first paycheck and this round captures the easing of the liquidity constraint because these students would now be paid a salary, among other changes associated with starting a new job. Similar to round 2, comparison students are also asked to fill in the surveys and solve the experimental tasks to coinside with this timeline. In what follows and for clarity in what we are measuring, we refer to these stages as: Baseline, after offer, and after paycheck.

2.4 Difference-in-Differences Results

In order to examine whether the economic and social preferences, cognitive performance and survey responses of these students change along the transition from college to the labor market, we employ a difference-in-differences (DID) strategy. The two main periods of interest are Round 2, when last-semester students receive a job offer, and Round 3, when they finally start working and receive at least one paycheck. An important contribution of this paper is to separate out these two periods in order to understand whether there are any changes in behavior along the transtition and whether the changes in decision making accompany a mere resolution of uncertainty after getting a job offer or whether there needs to be a real increase in their incomes. In order to tease out these various changes, we run the following DID regression specification:

$$\begin{aligned} y_{it} = \alpha_1 + \alpha_2 \text{Baseline 2} + \alpha_3 \text{After offer} + \alpha_4 \text{After paycheck} + \beta_1 \text{Last sem.} + \beta_2 \text{Baseline 2} \times \text{Last sem.} \\ + \beta_3 \text{After offer} \times \text{Last sem.} + \beta_4 \text{After paycheck} \times \text{Last sem.} + \varepsilon_{it} \quad (2.1) \end{aligned}$$

In the above specification, the dependent variables include social preferences, time preferences, risk aversion, cognitive performance, personal finances and psychological factors. We collect four measurements of these variables so the index t goes from 1 to 4. The constant in this regression provides the mean of the comparison group in Baseline 1. The rest of the α coefficients provide the difference in the means for the comparison groups across rounds. The coefficient β_1 reflects the difference in means between last-semester students and comparison group in baseline 1, β_2 provides a test for the parallel trends assumption, and β_3 and β_4 are the DID coefficients in the after offer and after paycheck stages, respectively. The standard errors are clustered at individual level.

Our main hypothesis was that even though these students attend one of the most pretigious universities in Colombia, they still have considerable uncertainty about when and where they will get their jobs. This uncertainty may be enough to affect their decision making behavior, in addition to when they receive their first paycheck and resolve a potential liquidity constraint.

2.4.1 Psychological Measures

One of the contributions of this paper is to highlight the role of psychological measures as mediators for economic decision making. For all of our regression specifications, we also control for psychological measures in order to check if the observed effect on the decision making task is simply a result of the subject's current psychological state. There are some important patterns within these psychological measures in the transition we analyze. As shown in Figure 2.1, last-semester students who receive a job offer report being differentially less tired, worried, depressed and frustrated. Further results are presented in Table 2.4. What is clear is that while last-semester students report being less worried or tired in Round 2 after the resolution of job uncertainty, by Round 3 after receiving a paycheck, these effects disappear, that is, there is not difference in psychological measures between last-semester and comparison. Part of this may be related to the additional responsibilities they have to take care of as they become more independent and pay for most of their expenses (Table 2.12). In Appendix 2.8.7, we report specifications adding controls for psychological factors and degree of independence.

2.4.2 Social Preferences

Dictator and ultimatum games have long been used in economics to provide important insight about the social preferences of people, more specifically, the importance of fairness, equity and reciprocity in their lives Henrich et al. (2004). We use these games to measure our subjects' social preferences. For the dictator games, we ask students to donate to both, another student as well as a foundation, from their endowment of 20,000 pesos. Unlike some of the versions of this task in earlier papers, where students were asked to make a hypothetical choice Kahneman et al. (1986), the game we set has real consequences for subjects' payoffs.

We show that under changes in people's conditions like when transitioning from college to the workforce, there may be temporary changes in social preferences. A number of studies have shown the remarkable stability of social preferences Carlsson et al. (2014), and our finding provides further nuance to that view. Comparing the first and last periods, there are no significant changes to the level of altruism towards other students and to a charitable foundation. However, on receiving a job offer, there is a temporary rise in altruism. This shows how a period that is also marked by a positive effect on psychological measures, such as perceptions of happiness, can have a statistically significant effect on altruism.

In Table 2.5, we see that students' generosity towards other individuals and towards the charitable foundation after job uncertainty is resolved is higher compared to lower-year students. On average, students allocate a far higher share to a foundation than to a student in the dictator game, donating more than half of their endowment to a foundation versus around 30 percent to a student in a dictator game. This is a very high share compared to the 20 percent allocation to other students observed in the literature List (2007). The increase in altruism observed in the post-job-offer stage ranges from 12 to 17 percentage points. In the dictator games, this translates to donating more than half one's endowment to other students, and about two-thirds of one's endowment to a foundation. The ultimatum game, where the recipient has a chance to respond to a donation, has an even higher level of average donation to another student, almost half of one's endowment. This further climbs to almost 60 percent for students who have received a job-offer.

Interestingly, after receiving a paycheck, subjects differentially give less in the dictator and ultimatum games relative to students in the comparison group. However, their donations to a foundation do not differ from those made by comparison students. We hypothesize that this behavior could respond to the fact that they already graduated and do not care as much for fellow students as when they were students themselves. Another explanation could be that their increased responsibilities as they are now more independent makes them want to keep more money for themselves. In fact, we see that after controlling for a dummy indicating that they pay for most of they expenses, the negative coefficient in dictator and ultimatum games dissapear, lending some support to the second hypothesis.

What is also striking about these results is how stable these preferences are among comparison students (Figure 2.2), and how there is a stark change in this pattern when last semester students receive job offers. The lack of differential pre-trends helps validate our research design in terms of reasonable identification. Furthermore, we find that similar effects persist even after controlling for emotions (Appendix Table 2.13). In sum, compared to the literature on social preferences, we observe very high levels of altruism. Our results show that resolution of uncertainty of a life event may change individuals' willingness to give.

2.4.3 Time Preferences

A growing body of research in economics studies how time preferences may be different between the poor and the non-poor (Carvalho et al., 2016; Haushofer et al., 2013; Lawrance, 1991), and whether time preferences are stable over time (Chuang and Schechter, 2015). Most of this literature is based on experimental measures of time preferences. An important critique to measuring time preferences parameters with experimental tasks comes from Dean and Sautmann (2016). Their 'soft credit constraints' model suggests that experimental tasks eliciting time preferences measure the marginal rate of intertemporal substitution (MRS) but do not identify the time preferences parameters because subjects' choices are affected by external factors outside of the experiment. In this model, for example, income shocks decrease the MRS while expenditure shocks increase it. In our setting, we measure how external factors (including but not limited to income and expenditure) associated with the transition from college to the labor force affect experimental choices in a time preferences task.

We look at three main measures calculated from a task adapted from Andreoni and Sprenger (2012). We find that students on the job market are less impatient after receiving a job offer relative to their counterparts who are not on the job market. Students who receive a job offer allocate the full initial endowment to the earlier period in 2 fewer intertemporal choices than comparison students who allocate the full amount in more than 5 of the 16 intertemporal choices (see middle column of Table 2.6). While we see a DID coefficient in the after paycheck stage of -0.8, indicating less impatience when students have already started working, it is not statistically significant unlike the after-offer coefficient which is significant at the 1% level.

Table 2.6 also shows results for a measure of present bias (column 3) and inconsistency in intertemporal choices (column 1). The measure of present bias enumerates the instances where a subject allocated a greater amount to the earlier period when the earlier period was a week from now rather than 5 weeks from now for the same delay length until the later period, i.e., week 1 vs. week 5 and week 5 vs. week 9. This is then weighted by the interest rate for the later period payoff in each intertemporal choice, to end up with a percentage of present bias. By the time last-semester students start working and receive a paycheck, we find an increase in present bias of 12 percentage points that is significant at the 10% level. Our subjects do not exhibit differences in the number of non-monotonic choices they make (i.e., allocating more to the sooner period when the interest rate for delaying increases).

To contextualize our result, Carvalho et al. (2016) find that before payday, poor individuals in the U.S. are more present biased when making choices about monetary rewards. We observe the opposite behavior for the period before individuals receive their income. The period immediately before being paid is characterized by the resolution of uncertainty regarding their future income and the outcome of their college education investment. Our results support the idea that individuals become less focused on the present once they resolve the uncertainty about their near future. Their choices, however, could also be explained by changes in psychological factors or credit constraints. As we reported earlier, individuals feel less tired, depressed, frustrated and worried after receiving and accepting a job offer (Table 2.4). Although we cannot see their credit constraints directly because they may have access to credit cards or loans but decide not to take them up, we see that last-semester students report lower loan-takeup than students in the comparison group in the after-offer period (Table 2.12).

In Table 2.15, we see that the impatience result holds up after controlling for psychological factors in the DID regression so changes in emotions are unlikely to be behind the observed decline in impatience. In terms of credit, we find evidence pointing to a real change in behavior. Despite not differing from comparison-groups students in the baseline period, results from a triple difference (DDD) regression (available upon request) show that students who become more impatient are also those who take up credit cards or loans in the after-offer and after-paycheck periods. For students who do not adopt these credit instruments, we continue to observe a fall in impatience after they receive a job-offer.

2.4.4 Risk and Ambiguity Preferences

Conceptually, risk and ambiguity preferences may change as individuals age, in response to shocks such as economic crises or natural disasters, or by temporary variations in self-control, emotions or stress Schildberg-Hörisch (2018). Given the evidence we present regarding changes in psychological factors among students transitioning to the labor market (subsection 2.4.1) and that the first two mechanisms are unlikely to take place in the timeframe of our study, we focus on how the third mechanism may affect risk and ambiguity aversion measures along the transition from college to the workforce.[8]

[8] Carvalho et al. (2016) propose liquidity constraints and scarcity as potential factors affecting risk aversion before and after receiving a paycheck. We are unable to test these mechanisms because in our "After paycheck" stage, our subjects do not only receive a paycheck but are still adapting to their new role as a working person.

To measure risk aversion, we elicit the preferred gamble from a choice-set of the six proposed in the Eckel and Grossman (2002) task. We code subjects as risk averse when they pick the first four or the first three of the six gambles proposed in this task. Inconsistency in risk choices is coded as an indicator of whether the subject switches more than once in the multiple price list by Tanaka et al. (2010). The ambiguity aversion variable counts the number of times the ambiguous urn is chosen out of nine possible choices ranging from the fully visible and the ambiguous urn. For details on the tasks or the definition of the variables see Section 2.3 and Appendix 2.8.3.

We find that students on the job market appear differentially less risk-averse after receiving a job offer. Table 2.7 demonstrates that there is a decrease in risk aversion among all students across rounds, but statistically significantly more so for students who receive a job offer.[9] Therefore, students who have their job uncertainty resolved appear to have a higher propensity to pick riskier gambles than lower-year students by about 11.1 percentage points. It is worth pointing out that the majority of students were risk averse at baseline with about 70 percent of students choosing one of the three least risky gambles to play. There is a statistically significant reduction in risk aversion among all students by the job offer stage in which the proportion of risk-averse subjects is reduced to about 50 percent and 38 percent among comparison group and last-semester students, respectively. The DID coefficient has a positive sign in the after paycheck period but it is not significant. In terms of inconsistencies of risk choices, we see no differences between the comparison group and last-semester students.[10]

The ambiguity aversion results show that, at baseline, students chose the ambiguous urn very few times (less than 4 times on average out of a total of 9 choices) independent of their last-semester or comparison group status. These preferences remain remarkably stable over time, with there being a slight trend towards less ambiguity aversion with every period. We observe no differential impact of being a last-semester student in any period.

To test whether self-control, stress and emotions may drive the results in Appendix Table 2.16 we check how our data conforms with existing theories. The work by Fudenberg, Levine and

[9]The standard measure of risk aversion in the Eckel and Grossman (2002) task is choosing one of the four most risk averse gambles which implies a CRRA greater than 0.5. With this measure we do not see parallel trends so we check an indicator for whether the subject chooses one of the first 3 gambles. The results are qualitatively similar and we cannot reject parallel trends in the second case.

[10]The 1132010Tanaka et al. () task allows to compute the Prospect Theory parameters (curvature of the value function, probability weighting and loss aversion) for students who switch at most once in the multiple price lists. We do not find any differential effects for last-semester students in these three measures (for details see Appendix Table 2.17.

coauthors Fudenberg and Levine (2006,0); Fudenberg et al. (2014) posits that when individual self-control resources are low, the risk-averse short-term self may prevail over the deliberative risk-neutral long-run self. The Cognitive Reflection Test (CRT) we administered measures exactly these “dual system” responses. As will be discussed in subsection 2.4.6, we do not find evidence of any changes in the CRT number of correct answers or time spent in this test, suggesting that self-control may not be driving the reduction of risk aversion among last-semester students accepting a job offer.

In the case of stress and emotions, the hypothesis that has received most support in psychology is the affect-infusion model Forgas (1995) which implies that individuals in a happy mood may be willing to take more risks than those in a sad mood. In subsection 2.4.1 we documented positive changes in last-semester students’ worry, depression, frustration, and tiredness at the time of accepting a job offer. We control for these psychological factors in Table 2.16. With the controls in place, the after-offer result for last-semester students being differentially less risk averse than their counterparts in lower years vanishes, giving support to the affect infusion model and underlining the importance of sujective measures of wellbeing and emotions as mediators in economic decision making.

2.4.5 Self-Perceptions of Financial Status

We find statistically significant effects of receiving a job offer on the self-reported financial health of last-semester students. In Table 2.8 we show the result of regressions on outcomes such as whether it is hard to come up with money for an emergency, whether it is hard to cover next week’s expenses with the money they have today, and whether they are stressed about personal finances. In the first case, receiving a job offer has a significant and positive effect for last-semester students. To elaborate, they report finding it hard to come up with money less frequently (20 percentage points) than the baseline, and when compared to students in lower years. Therefore, in terms of perception of own wealth, there is a clearly positive effect of merely receiving a job offer, without having yet been paid. There is no statistically significant differential effect for last-semester students after receivng a paycheck, which is telling of the immense effect that the resolution of uncertainty alone has on perception of one’s coping ability.

The results for the self-reported measure on it being hard to come up with money in an emergency holds up to the regression specification with different controls - i.e. last-semester students less frequently report this in Round 2 after receiving a job offer. Indeed, after controlling for a dummy indicating that the subject pays for most of own expenses, there is an additional effect of about -13 percentage points for last-semester students after receiving their paycheck as well.

2.4.6 Cognitive Performance

Corroborating the above pattern of reports of better psychological well-being dissipating by Round 3, we have further evidence that after receiving their paycheck, there are increased responsibilities as subjects now pay most of their expenses themselves (Table 2.12). We study changes in cognitive performance by looking at how students perform in tasks such as the Raven's Matrices, Cognitive Reflection Test - CRT, Flanker task and the Stroop test (Table 2.9). These increased responsibilities may be contributing to an increasing cognitive load and we find that after receiving a paycheck, last-semester students perform differentially worse than lower-year students in the Raven's Matrices test. We find no differences in performance in any of the other cognitive tasks nor in time to respond the CRT.

Performance for all students in the Raven's Matrices task improves over time and this may be attributable to learning effects. On average, they respond correctly to 4 questions (column 1 in Table 2.9) at baseline, and lower year students improve their score by 2.5 correct questions, while last-semester students lag slightly behind at a score below 6 after receiving a paycheck. This is consistent with the additional responsibilities and changes associated with starting a new job. It is possible that these changes impose a cognitive load on last-semester students and impairs their performance in these cognitive tasks.

It is interesting to note that Mani et al. (2013) find changes in Raven's test performance among sugarcane farmers before or after receiving payment from harvest while Carvalho et al. (2016) do not find changes in the cognitive tasks they administer (which do not include the Raven's test) when comparing before and after receiving a paycheck. Because it is the only test in which we find significant differences, it would be interesting to know whether there is something particular about the Raven's test that Mani et al. (2013)'s paper and ours find results when aministering it.

2.5 Effect heterogeneity by socioeconomic background

In this section, we take advantage of the natural variation in socio-economic backgrounds of students in our sample. Because their background is not randomized, we cannot make causal inferences in this section, but instead provide suggestive evidence for the correlates of poverty.

We find that while in college, poorer students are likely to experience a larger burden in terms of psychological well-being and perception of financial stress than richer students. However, we do not observe differences by socio-economic status (SES) in social preferences, impatience, cognitive measures, or risk preferences. In addition, once students transition to the labor market, any previously observed variations across students with different SES fall or become negligible, suggesting that obtaining a job equalizes the differences previously observed in college.

Exploring this dimension of heterogeneity adds to the recent literature on poverty, cognition and economic decision-making by quantifying how more or less affluent students perform in the tasks, as well as how their performance changes during the college to labor market transition. The current evidence shows that poor people may respond differently to cognitive and decision-making tasks than richer people. Mani et al. (2013) show that the poor perform worse in cognitive tasks when primed for poverty with a hypothetical situation. They also confirm this result with poor farmers performing the tasks before and after harvest. Similarly, Carvalho et al. (2016) find that subjects are more present-biased before receiving a paycheck. This effect on time preferences is similar to when experimental subjects receive a negative income shock in the lab Haushofer et al. (2013).

In each of the graphs in this section we plot the local polynomial smooth of the mean outcome by tuition level for the group of last-semester students and the comparison group. The plots shown side by side correspond to the after-offer and after-paycheck stages of the outcome shown in the figure caption. To have a sense of the statistical significance of the differences between groups and across tuition levels within the same group we include 83% confidence intervals.[11] To have a reference point, a vertical line highlights the median tuition (600,000 COP or US$200).

Figures 2.3 and 2.5 show one of the measures of psychological well-being (depression) and perceived financial wellbeing as measured by the question asking whether it would be hard to come

[11] 83% confidence intervals are commonly used when comparing two means (http://chris-said.io/2014/12/01/independent-t-tests-and-the-83-confidence-interval-a-useful-trick-for-eyeballing-your-data/)

up with a sum of money for an emergency within a week. In general, more affluent students in the comparison group are better off in terms of psychological well-being and perceived financial stress than students around the median tuition level. Around 60 to 80 percent of the poorer comparison-group students report that it would be hard to come up with the equivalent of US$ 1,000 for an emergency relative to less than 40 percent of the richer students (Figure 2.5). Although the confidence intervals are wider in Figure 2.3 panels, the pattern suggests that depression is highest among students around the median tuition and lowest for richer students.

Figures 2.3 and 2.5 also show that when students receive and accept an offer (left panel), the mean of these outcomes goes down substantially relative to the comparison group and there is very little variation by tuition level. In the hard-to-come-up-with-money measure, the new level for all last-semester students is similar to the richest students in the comparison group. In the after paycheck stage, the levels go up again and are not statistically different from the comparison group at most tuition levels.

To summarize the findings in the social preferences measures, we show the fraction of the 20,000 COP endowment allocated to another student in the dictator game in Figure 2.4. There is no variation in generosity by student SES and the increase in donations in the after-offer period is homogeneous across SES levels.

Regarding risk aversion, there is almost no variation by tuition level (see Figure 2.6). Between 40 and 50 percent of poorer and richer students in the comparison group are risk averse. The level of this measure is slightly lower for last-semester students, specifically for the richest students in the sample whose risk aversion is lower by about 20 percentage points relative to the poorer students. That is, students who started out as richer become less risk averse once they receive and accept a job offer. In the after paycheck period, we see a similar but nuanced pattern.

Performance in the Raven's test is fairly constant between 5 and 5.5 questions answered correctly out of 9 for all students in the after offer stage. There is no indication that performance correlates with SES or with receiving a job offer. After paycheck, performance levels go up to about 6.5 correct answers for students in the comparison group and around 6 correct answers for last-semester students but the confidence intervals overlap across most of the support of the tuition variables.

Overall, the findings in this section indicate that most of the results in section 2.4 arise from bigger changes in the poorer students' performance in the experimental tasks. In other words, the average poor student in college is more likely to experience more financial and psychological stress. However, by the time students receive and accept a job offer, there is no longer outcome variation by SES. Perhaps surprisingly, other outcomes such as generosity in the dictator game, are not sensitive to socioeconomic background but change along the different stages of the transition.

2.6 Conclusion

Our paper highlights significant, yet temporary spikes in behavior and preferences that result while transitioning from being a college student to a working member of society. We are among the first to highlight "windows of clarity" that occur between receiving a job-offer and getting the first paycheck. These "windows of clarity" are associated with a fall in impatience, a rise in altruistic behavior, and a perception of greater liquidity despite not having received any paycheck. These findings point towards a policy opportunity with regard to the timing of crucial long-term decisions such as health care choices, pension and life insurance decisions. These choices that are often made after people begin working and after they have been paid for the first time at their new jobs, so this evidence suggests a departure from that norm by indicating that they should be made during these windows of clarity.

Our results are surprising since graduates of this prestigious university, consistently ranked amongst the top two in the country, are almost certain to find a job and have financial security. Therefore, in expectation, we should not have noted changes in social or time preferences, cognitive performance and other related tasks and decisions when these students received their job offer. However, our results imply that uncertainty surrounding specificities of the job, such as when they will receive a job offer and how much it pays, can have substantial behavioral effects. These uncertainties may be large enough to cause changes in decision-making merely in response to receiving a job offer, even before being paid for the first time.

We use a difference-in-differences design to study the effect of first transitioning from being a (last-semester) college student to receiving a job offer, and then the effect of receiving a paycheck, on decision-making. By having matched lower-year students in the comparison group answer the same questions as the last-semester students at roughly the same times, we can effectively compare their answers across rounds to determine differential trends among these final semester students. Yet, as we are unable to randomly assign the status of being a last semester student, we cannot make strong causal claims about the results. However, we do observe parallel trends among last semester and lower year students, adding credibility to the DID setup. And the patterns we observe are strongly suggestive of the effects that transitioning to the job market can have on decision-making behavior measured through experimental tasks.

We find that there is indeed a change in time preferences, altruism, perceptions of financial health and psychological measures because of merely receiving a job offer. The finding that last-semester students become differentially more altruistic and less impatient solely in response to a job offer demonstrates what a strong effect the resolution of this job uncertainty can have. These students report it now being less hard to come up with money for an emergency even though they have not been paid in their new jobs just yet. This contradicts the perception that students at a top university would have no uncertainty about getting a job since without any change in their earnings, their perception of their status changes significantly. Furthermore, these students report that they are less worried, tired, depressed and frustrated when they receive their job offer. While this is not surprising, what is striking is the large effect these emotions have on their decision-making during the transition, acting as mediators. Without accounting for these feelings, students appear to become less risk-averse on receiving their job offer. However, once we control for these emotions, these associations become statistically indistinguishable from zero. In other cases, the effect of the transition is made stronger, like in the case of becoming less impatient when the job uncertainty is resolved. Often when studying decision-making behavior, such self-reported measures are not taken into account and this could be affecting the interpretation of results.

After receiving at least one paycheck from their new jobs, all the effects we observed in the after-job-offer period dissipate and are no longer statistically or substantively significant. Furthermore, these students perform differentially worse on the Raven's Matrices-type cognitive test. Finally, after receiving a paycheck, students report being more frustrated, worried and tired. These results are consistent with the hypothesis that after actually receiving some income, these students have to take on many more responsibilities relating to becoming more independent. They may also have to take care of other family members, adding to their stress levels and generating a decrease in the bandwidth available to solve problems Mullainathan and Shafir (2013).

It appears that the resolution of uncertainty regarding the details of their job is the crucial factor that induces changes in the decision-making of students who transition to the labor market. Their perceptions of their financial health also change positively. However, after starting to work and being paid, there may be greater cognitive load that comes with having a lot more responsibilities that lead to changes in cognitive performance and feelings of worry and tiredness.

2.7 Tables and Figures

Table 2.1: Differences in baseline characteristics

Variable	Comparison	Last-sem. students	Observations
Fraction with tuition <600,000 pesos (US$200)	0.54	0.59	363
Female	0.25	0.25	363
Age	22.98	25.07	352
Undergraduate	0.87	0.88	363
Semester in college	6.18	10.44	355
GPA (from 0 to 5)	3.80	3.81	356
Poor	0.36	0.38	363
Residential stratum (1=lowest, 6=highest)	2.85	2.93	362
Employed	0.43	0.65	363
Expected first salary after graduation (pesos)	2,052,162	1,957,584	363
Expected salary five years after graduation (pesos)	4,508,649	4,819,663	363
No. of semesters working full time from start of college	0.32	0.52	363
No. of semesters working part time from start of college	1.97	2.84	363
No. of semesters in an intership from start of college	0.09	0.53	363
How hard to find a job after graduation (1=very hard, 5=very easy)	2.83	2.57	363
Parents live in different household	0.05	0.07	363
Parents pay most expenses	0.72	0.57	363

Notes: Last-semester students are in their final semester in college. Comparison students are in lower years and selected to match last-semester students in gender, major and economic background.

Table 2.2: Summary of research design

	Round 1 **Job search**	Round 2 **After job offer**	Round 3 **After pay**
Last-semester students	Send resumes, job interviews	Receive and accept offer	Cash on hand
Other students	Normal student life		
Timeline	April - May, 2016	October, 2016	December, 2016

Table 2.3: Mean levels and differences between baseline surveys

	Baseline 1 level (y_{11})		$y_{12} - y_{11}$		**p-value**
	Comparison	*Last sem.*	*Comparison*	*Last sem.*	**DID**
Social preferences					
Fraction given to student - dictator	0.331	0.367	-0.022	-0.059	0.394
Fraction given to foundation - dictator	0.553	0.522	-0.084	-0.053	0.537
Fraction given to student - ultimatum	0.438	0.463	-0.002	-0.027	0.965
Financial situation					
Hard to come up with money for emergency	0.592	0.539	0.005	0.058	0.123
Hard to cover expenses	0.234	0.180	-0.036	0.017	0.200
Stressed about personal finances	0.391	0.427	-0.032	-0.068	0.538
Risk and ambiguity aversion					
Risk averse	0.832	0.843	-0.034	-0.045	0.035
Extremely risk averse	0.696	0.685	-0.013	-0.003	0.387
Fraction of inconsistent risk choices	0.261	0.230	-0.110	-0.080	0.289
No. times choosing ambiguous urn (out of 9)	3.793	3.932	-0.018	-0.156	0.722
Time preferences					
No. of non-monotonic choices (out of 16)	2.258	2.305	-0.474	-0.521	0.039
No. times allocating full amount to sooner period (out of 16)	2.577	2.831	1.165	0.912	0.265
% present biasedness (weighted by interest rate)	30.073	28.800	-1.214	0.059	0.719
Cognitive tests					
Cognitive Reflection Test (CRT) (correct out of 3)	1.142	1.023	0.180	0.300	0.155
Raven's matrices (correct out of 9)	4.038	4.130	2.359	2.267	0.612
Stroop test (correct out of 45)	16.837	16.438	1.199	1.599	0.250
Flanker's test (correct out of 45)	28.453	27.744	1.709	2.417	0.895
Psychological measures					
Tired	0.598	0.539	-0.047	0.011	0.550
Frustrated	0.196	0.180	0.100	0.116	0.271
Worried	0.402	0.331	0.069	0.140	0.363
Depressed	0.141	0.157	0.083	0.067	0.428

Notes: Columns 2 and 3 show survey 1 means of the variables in the rows. Columns 4 and 5 show the first difference of the means in the two baseline periods. Column 6 shows the p-value of the DID estimate (Col. 5 - Col. 4).

Table 2.4: Psychological measures

	Tired	**Frustrated**	**Worried**	**Depressed**
Baseline 1	0.598***	0.196***	0.402***	0.141***
	(0.036)	(0.029)	(0.036)	(0.026)
Baseline 2	0.005	0.141***	0.136***	0.098***
	(0.045)	(0.040)	(0.045)	(0.037)
After offer	-0.163***	0.087**	0.022	0.098**
	(0.052)	(0.043)	(0.048)	(0.039)
After paycheck	-0.375***	0.043	-0.136***	0.060
	(0.047)	(0.041)	(0.049)	(0.041)
Baseline 1 * Last-sem.	-0.059	-0.016	-0.071	0.016
	(0.052)	(0.041)	(0.051)	(0.038)
Baseline 2 * Last-sem.	-0.039	-0.063	-0.057	-0.042
	(0.065)	(0.057)	(0.063)	(0.053)
After offer * Last-sem.	-0.174**†	-0.132**††	-0.130**†	-0.167***†††
	(0.069)	(0.059)	(0.066)	(0.051)
After paycheck * Last-sem.	0.175**††	0.081	0.178**††	0.000
	(0.070)	(0.066)	(0.073)	(0.061)
Observations	1355	1355	1355	1355
No. subjects	362	362	362	362
Controls	No	No	No	No

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1. Multiple Inference (Benjamini-Hochberg on after-offer and after-paycheck DID coefficients): ††† p_m<0.01, †† p_m<0.05, † p_m<0.1

Table 2.5: Social preferences

	Fraction given to student - dictator	**Fraction given to foundation - dictator**	**Fraction given to student - ultimatum**
Baseline 1	0.331***	0.553***	0.438***
	(0.018)	(0.023)	(0.014)
Baseline 2	-0.052***	-0.076***	-0.015
	(0.016)	(0.017)	(0.014)
After offer	0.012	-0.085***	0.027
	(0.027)	(0.030)	(0.023)
After paycheck	0.008	-0.076**	0.029
	(0.030)	(0.030)	(0.026)
Baseline 1 * Last-sem.	0.036	-0.031	0.025
	(0.025)	(0.034)	(0.019)
Baseline 2 * Last-sem.	0.021	0.015	0.001
	(0.024)	(0.024)	(0.020)
After offer * Last-sem.	0.173***†††	0.163***†††	0.124***†††
	(0.044)	(0.044)	(0.037)
After paycheck * Last-sem.	-0.135***†††	-0.054	-0.121***†††
	(0.040)	(0.044)	(0.034)
Observations	1355	1355	1355
No. subjects	362	362	362
Controls	No	No	No

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1. Multiple Inference (Benjamini-Hochberg on after-offer and after-paycheck DID coefficients): ††† p_m<0.01, †† p_m<0.05, † p_m<0.1

Table 2.6: Time preferences (based on 16 choices)

	No. of non-monotonic choices	No. choices allocating full amount to sooner period	% present biased-ness (weighted by interest rate)
Baseline 1	1.740***	2.981***	30.847***
	(0.257)	(0.318)	(3.812)
Baseline 2	-0.418*	1.358***	-9.176**
	(0.251)	(0.302)	(4.517)
After offer	-0.894***	2.500***	-2.664
	(0.279)	(0.382)	(4.831)
After paycheck	-0.997***	2.117***	-7.596
	(0.262)	(0.392)	(4.854)
Baseline 1 * Last-sem.	-0.022	0.718	-4.675
	(0.410)	(0.473)	(5.147)
Baseline 2 * Last-sem.	-0.478	-0.132	9.098
	(0.431)	(0.497)	(6.299)
After offer * Last-sem.	0.033	-1.897***††	-1.857
	(0.471)	(0.624)	(6.857)
After paycheck * Last-sem.	-0.135	-0.802	12.489*
	(0.424)	(0.634)	(7.187)
Observations	787	787	787
No. subjects	314	314	314
Controls	No	No	No
Cond. on correct answer to understading question	Yes	Yes	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1. Multiple Inference (Benjamini-Hochberg on after-offer and after-paycheck DID coefficients): ††† p_m<0.01, †† p_m<0.05, † p_m<0.1

Table 2.7: Risk and ambiguity aversion

	Picked one of 4 most risk averse gambles (of 6)	**Picked one of 3 most risk averse gambles (of 6)**	**Inconsistent in risk choices**	**No. ambiguous urns chosen (out of 9)**
Baseline 1	0.832***	0.696***	0.261***	3.793***
	(0.028)	(0.034)	(0.033)	(0.142)
Baseline 2	0.005	0.011	-0.071**	-0.053
	(0.030)	(0.033)	(0.031)	(0.150)
After offer	-0.234***	-0.196***	-0.125***	0.433**
	(0.041)	(0.043)	(0.031)	(0.178)
After paycheck	-0.277***	-0.272***	-0.168***	0.583***
	(0.043)	(0.045)	(0.032)	(0.168)
Baseline 1 * Last-sem.	0.011	-0.010	-0.031	0.138
	(0.039)	(0.049)	(0.045)	(0.203)
Baseline 2 * Last-sem.	-0.095**	-0.045	-0.047	-0.087
	(0.045)	(0.052)	(0.045)	(0.244)
After offer * Last-sem.	-0.143**	-0.111*	-0.031	-0.202
	(0.064)	(0.066)	(0.046)	(0.286)
After paycheck * Last-sem.	0.078	0.056	0.016	-0.285
	(0.064)	(0.069)	(0.047)	(0.284)
Observations	1355	1355	1355	1232
No. subjects	362	362	362	362
Controls	No	No	No	No

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Multiple Inference (Benjamini-Hochberg on after-offer and after-paycheck DID coefficients): ††† $p_m<0.01$, †† $p_m<0.05$, † $p_m<0.1$

Table 2.8: Perceived financial status

	Hard to come up with money	**Hard to cover expenses**	**Stressed about personal finances**
Baseline 1	0.592***	0.234***	0.391***
	(0.036)	(0.031)	(0.036)
Baseline 2	0.071**	-0.038	-0.065*
	(0.032)	(0.035)	(0.037)
After offer	-0.022	0.011	-0.054
	(0.041)	(0.042)	(0.043)
After paycheck	-0.049	-0.022	-0.054
	(0.040)	(0.038)	(0.043)
Baseline 1 * Last-sem.	-0.053	-0.054	0.036
	(0.052)	(0.043)	(0.052)
Baseline 2 * Last-sem.	-0.071	0.061	0.032
	(0.046)	(0.047)	(0.051)
After offer * Last-sem.	-0.200***†††	-0.049	-0.096
	(0.060)	(0.056)	(0.063)
After paycheck * Last-sem.	-0.038	0.042	0.045
	(0.064)	(0.057)	(0.068)
Observations	1355	1355	1355
No. subjects	362	362	362
Controls	No	No	No

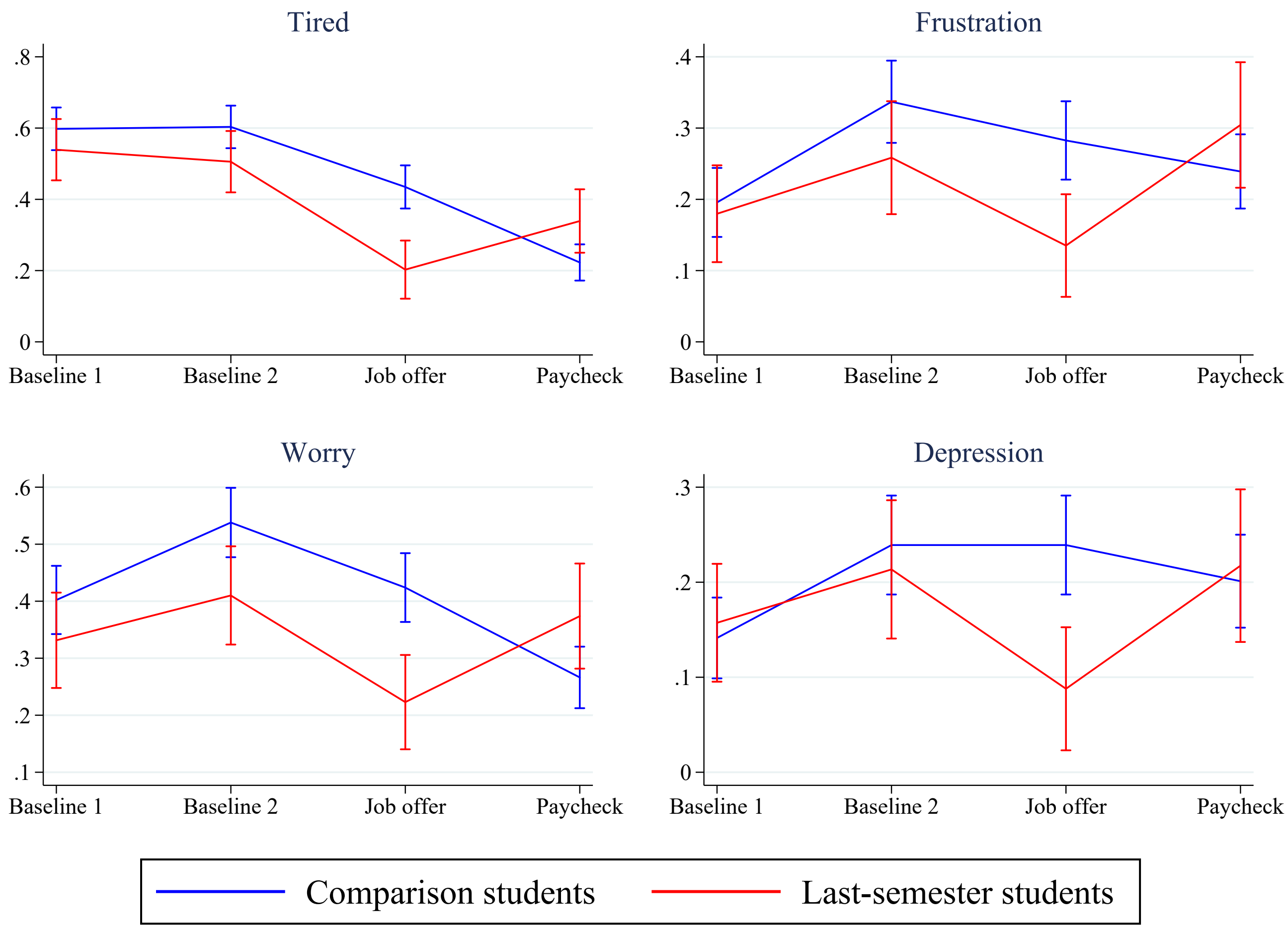
Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1. Multiple Inference (Benjamini-Hochberg on after-offer and after-paycheck DID coefficients): ††† p_m<0.01, †† p_m<0.05, † p_m<0.1

Table 2.9: Performance in cognitive tests

	Raven's Test	**Numerical Stroop Test**	**Flanker Task**	**CRT No. correct**	**CRT min. to answer**
Baseline 1	4.038***	16.837***	28.453***	1.142***	2.564***
	(0.118)	(0.463)	(0.735)	(0.061)	(0.188)
Baseline 2	2.370***	1.002*	2.125***	0.175***	0.902***
	(0.138)	(0.568)	(0.773)	(0.062)	(0.265)
After offer	1.116***	1.402**	1.216	0.045	-0.425*
	(0.134)	(0.608)	(0.988)	(0.070)	(0.227)
After paycheck	2.500***	2.943***	3.240***	0.303***	-0.058
	(0.146)	(0.587)	(0.932)	(0.069)	(0.225)
Baseline 1 * Last-sem.	0.092	-0.399	-0.708	-0.119	0.229
	(0.171)	(0.714)	(1.046)	(0.086)	(0.324)
Baseline 2 * Last-sem.	-0.102	0.930	-0.148	0.130	-0.472
	(0.200)	(0.807)	(1.119)	(0.091)	(0.378)
After offer * Last-sem.	0.049	0.714	1.556	0.038	-0.344
	(0.230)	(0.958)	(1.557)	(0.107)	(0.363)
After paycheck * Last-sem.	-0.604**	0.216	0.311	-0.091	-0.378
	(0.234)	(0.912)	(1.384)	(0.104)	(0.347)
Observations	1249	1228	1229	1243	1301
No. subjects	362	362	361	362	362
Controls	No	No	No	No	No

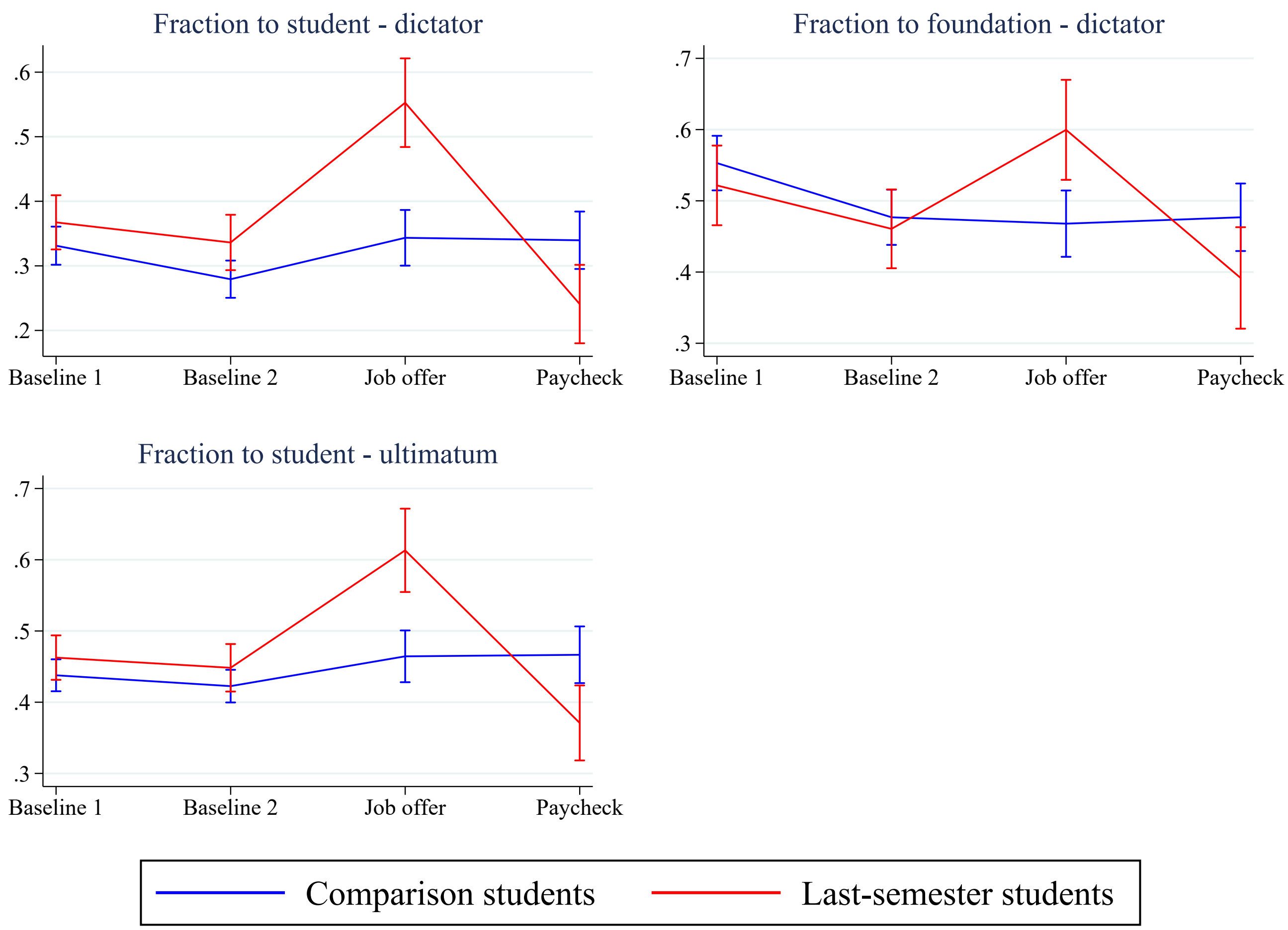
Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1. Multiple Inference (Benjamini-Hochberg on after-offer and after-paycheck DID coefficients): ††† p_m<0.01, †† p_m<0.05, † p_m<0.1

Figure 2.1: Patterns in self-reported psychological factors



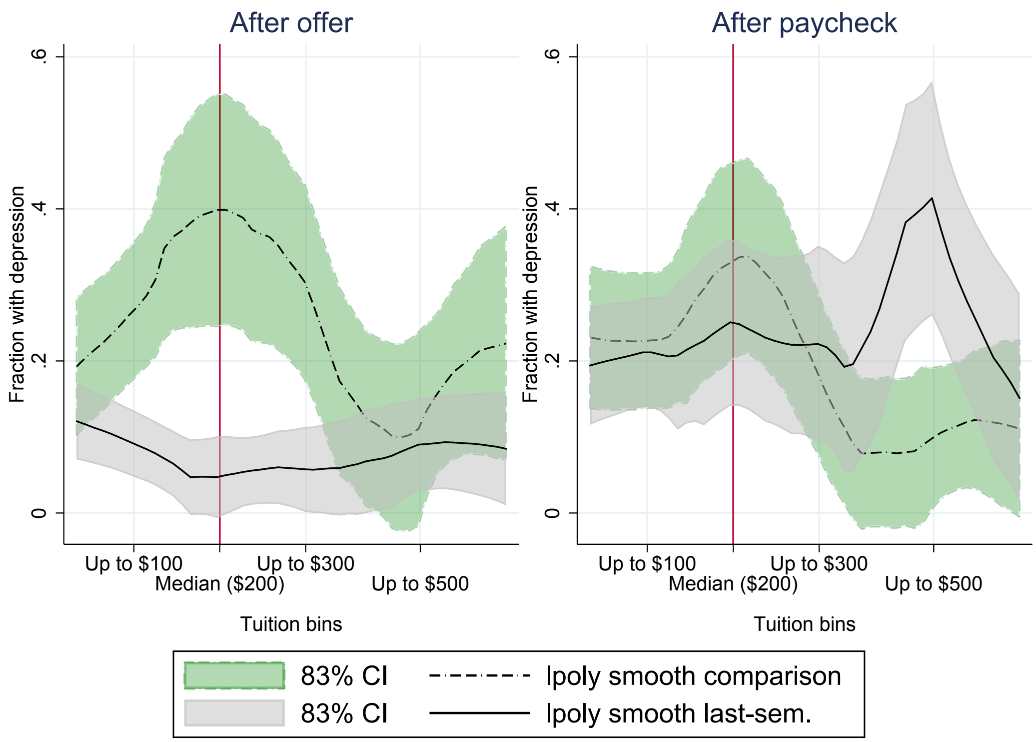
Notes: For each outcome we report means and 95% confidence intervals for last-semester (red) and comparison students (blue). These means are calculated from DID regressions. For the after-offer and after-paycheck stages, comparison students were asked to complete the same tasks as last-semester students during the same period.

Figure 2.2: Social Preferences (dictator and ultimatum games)



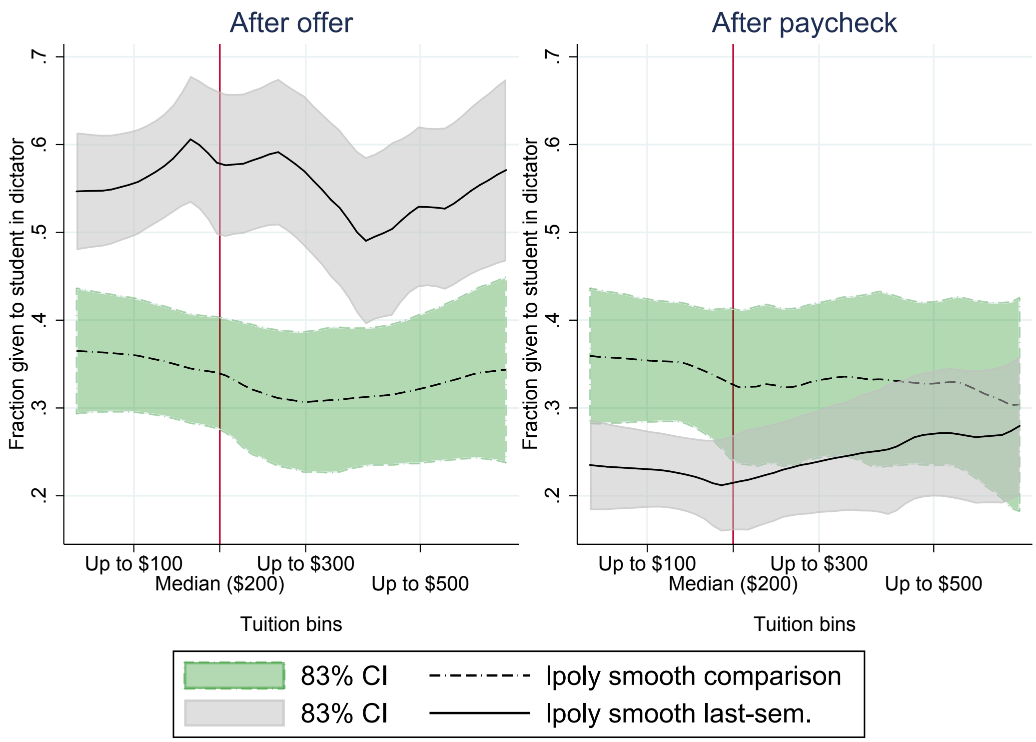
Notes: For each outcome we report means and 95% confidence intervals for last-semester (red) and comparison students (blue). These means are calculated from DID regressions. For the after-offer and after-paycheck stages, comparison students were asked to complete the same tasks as last-semester students during the same period.

Figure 2.3: Psychological well-being (depression) by tuition level



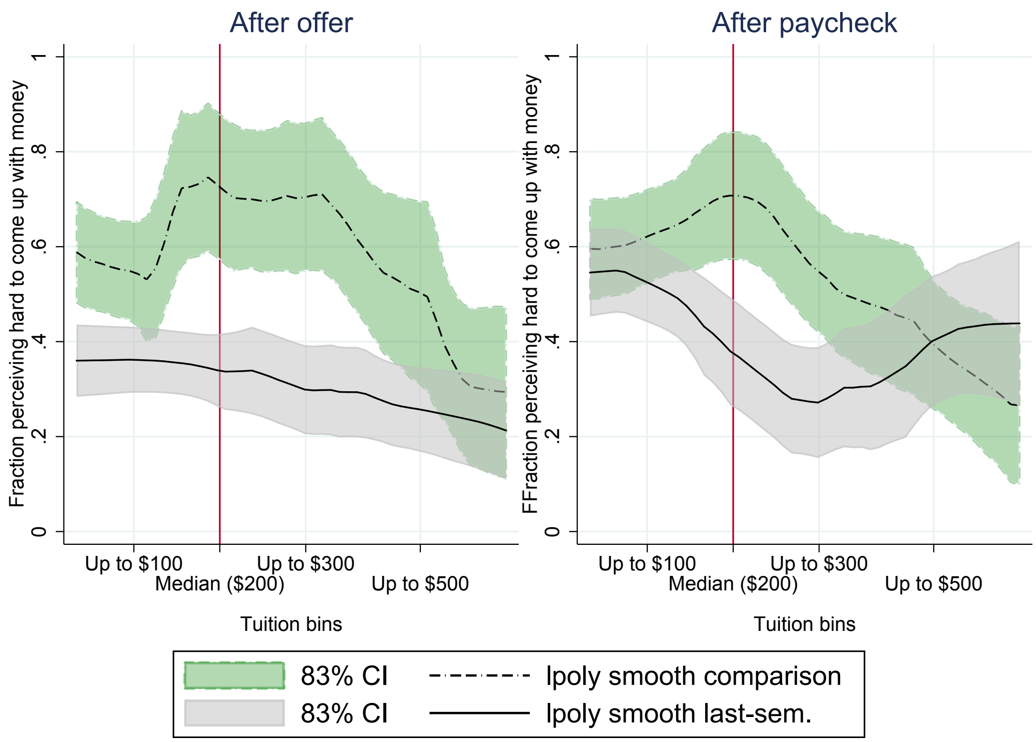
Notes: The lines in both panels are smoothed local means of the outcome, along tuition bins. The shaded areas are 83% confidence intervals to check whether two parameter means are different from each other at the 95% level. The left panel shows means for last-semester and comparison students in the after-offer stage while the right panel shows means in the after-paycheck stage.

Figure 2.4: Dictator game by tuition level



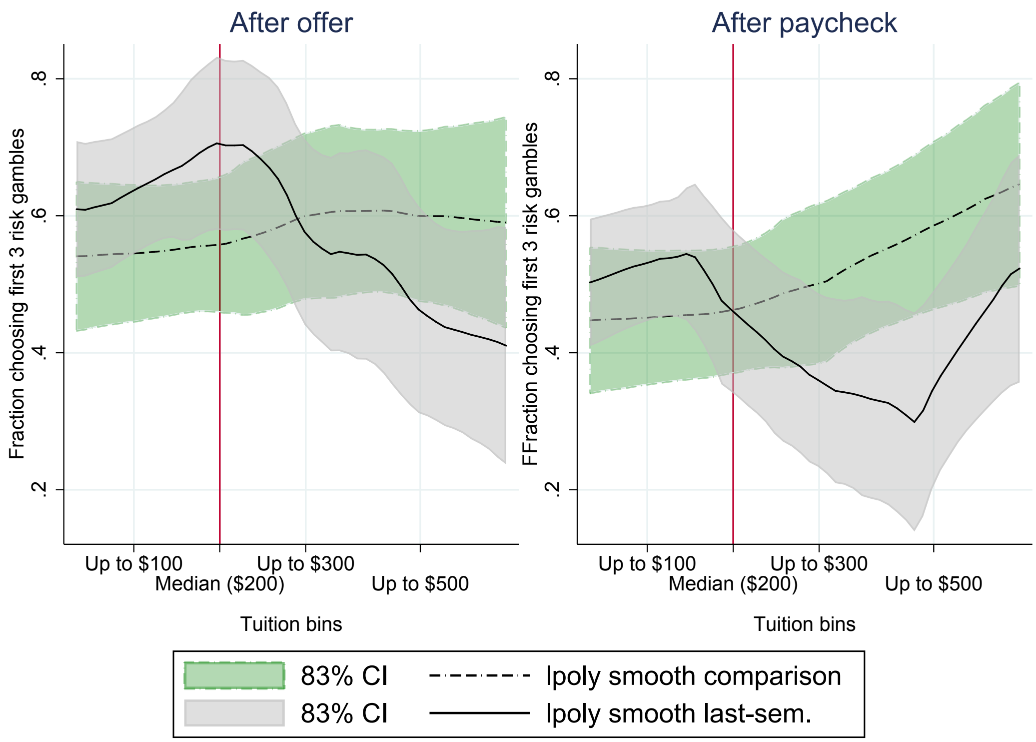
Notes: The lines in both panels are smoothed local means of the outcome, along tuition bins. The shaded areas are 83% confidence intervals to check whether two parameter means are different from each other at the 95% level. The left panel shows means for last-semester and comparison students in the after-offer stage while the right panel shows means in the after-paycheck stage.

Figure 2.5: Hard to come up with money for an emergency by tuition level



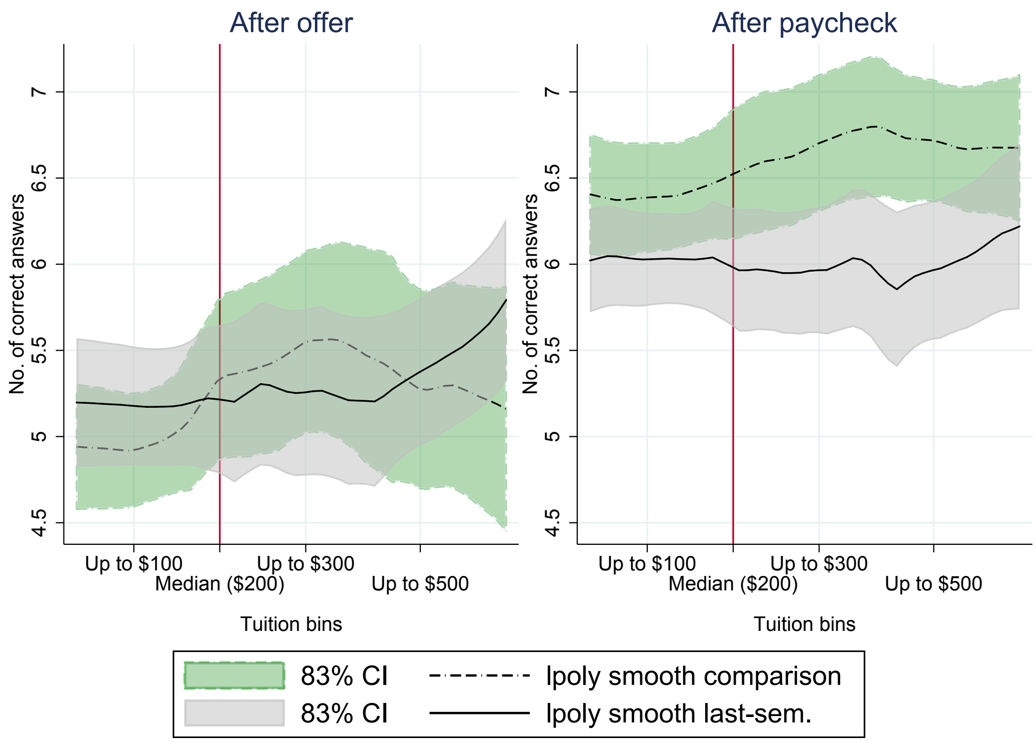
Notes: The lines in both panels are smoothed local means of the outcome, along tuition bins. The shaded areas are 83% confidence intervals to check whether two parameter means are different from each other at the 95% level. The left panel shows means for last-semester and comparison students in the after-offer stage while the right panel shows means in the after-paycheck stage.

Figure 2.6: Risk aversion by tuition level



Notes: The lines in both panels are smoothed local means of the outcome, along tuition bins. The shaded areas are 83% confidence intervals to check whether two parameter means are different from each other at the 95% level. The left panel shows means for last-semester and comparison students in the after-offer stage while the right panel shows means in the after-paycheck stage.

Figure 2.7: Raven's test by tuition level



Notes: The lines in both panels are smoothed local means of the outcome, along tuition bins. The shaded areas are 83% confidence intervals to check whether two parameter means are different from each other at the 95% level. The left panel shows means for last-semester and comparison students in the after-offer stage while the right panel shows means in the after-paycheck stage.

2.8 Preference Measures, Attrition, and Mechanisms

2.8.1 Social Preferences

For social preferences we use the dictator and ultimatum games. By introducing these games, we were interested in seeing whether altruistic behavior changes across the different stages. In the dictator game, participants were told that they will receive 20,000 pesos for sure. Then, they had to choose whether to give part of their allocation to another randomly chosen student participating in the study. If this task would be chosen for payment, the allocation chosen by the student would be implemented. A second question of this game changes the recipient of the gift from a randomly chosen student to a foundation that helps kids in need in Bogota. In the ultimatum game, the setup is the same except that now the subject proposes an allocation to the recipient student which can be rejected or accepted by the recipient.[12]

2.8.2 Time Preferences

Time preference is an important dimension of economic decision making that we measure in this study by adapting the elicitation task presented in Andreoni and Sprenger (2012). The idea of the task is that subjects are given a pre-specified monetary amount and are required to allocate it between two dates: earlier and later (see appendix 2.8.2.1). A difference between our implementation of the task relative to Andreoni and Sprenger (2012) is that the allocations are not continuous but discretized to increase in 1,000 pesos (about 33 dollar cents) intervals. Participants are instructed to allocate 50,000 pesos (about US$17) between two dates separated by 4 or 9 weeks. The earliest payment they could receive is one week from the date they respond to the survey because participants were responding to the surveys online and that made it impossible to pay them the same day they finished it. The trade-offs they face are between weeks 1 and 5, 1 and 9, 5 and 9, and 5 and 13. For each of these trade-offs, they made 4 decisions with varying interest rates if money is allocated to the later date (1, 10, 50, and 100 percent interest rates). When they choose a value in the earlier date, the amount to be received in the later date was automatically calculated in the "later" column including interest. In total, in each round they made 16 choices allocating money to earlier and later dates. We analyze the monetary values assigned to early dates in each of the four time comparisons. The first decision we study is the number of non-monotonic decisions made by

[12] The response from the recipient in this game was actually not implemented because participants were not responding to the survey simultaneously. In practice, whatever amount the participant donated was assigned to a randomly chosen student.

a student. Non-monotonocity in this case refers to allocating an increasing amount of money to an earlier period as the interest rate for the later period payout goes up. We also examine the "impatience" of students by examining how many times out of the 16 choices students allocate the entire 50,000 pesos to the earlier period. We also study a measure of present-biasedness by examining at each of the four interest rate levels, what the probability is of assigning a greater amount of money to week 1 vs. week 5 for a delay of 4 weeks and 8 weeks, respectively. These probabilities are then weighted proportional to the interest rates in order to derive a percentage of "present-biasedness".

2.8.2.1 Time preferences based on Andreoni and Sprenger (2012)

	EARLIER			LATER	
Allocate		to be received next week	AND		to be received in five weeks with a 1% interest
Allocate		to be received next week	AND		to be received in five weeks with a 10% interest
Allocate		to be received next week	AND		to be received in five weeks with a 50% interest
Allocate		to be received next week	AND		to be received in five weeks with a 100% interest
Allocate		to be received next week	AND		to be received in nine weeks with a 1% interest
Allocate		to be received next week	AND		to be received in nine weeks with a 10% interest
Allocate		to be received next week	AND		to be received in nine weeks with a 50% interest
Allocate		to be received next week	AND		to be received in nine weeks with a 100% interest

2.8.3 Risk Preferences

We elicit risk aversion using the Eckel and Grossman (2002) measure (see example and details in Appendix 2.8.3.1). This method consists of presenting six different gambles, varying the expected return, the standard deviation, and the implied CRRA range. Subjects are instructed to select one of the gambles to play. Each gamble has a 50 percent probability of receiving a low payoff and 50 percent probability of receiving a high payoff, except the first one in which both payoffs are the same. If this task is selected for payment at the end of the survey, the gamble they choose will actually be played. The expected payoff in gambles 1 to 5 increases linearly with risk. For gambles 5 and 6, the expected payoff is the same but the risk is bigger in gamble 6 as reflected by the higher standard deviation. Risk-averse subjects are expected to choose gambles with a lower standard deviation, while risk neutral subjects should choose the gamble with higher expected return (gamble 5) and risk-seeking subjects should choose gamble 6 Charness et al. (2013).

To analyze risk choices we use the risk lotteries from Tanaka et al. (2010). This method is intended to capture Prospect Theory parameters through a series of three lotteries which are much more complex than the Eckel and Grossman (2002) measure.[13] The lotteries consist of a given number of rows and two columns designated A and B. In each row, columns A and B contain two values each that represent payoffs and their probabilities appear at the top of the table. For each row subjects have to choose whether they prefer column A or B. In the instance that this task is randomly chosen for payment at the end of the survey, students are instructed that if one of the rows is chosen at random, the payment they receive will depend on the probability stated at the top of the column. The lotteries are designed such that a risk neutral person will choose column A up to row 6 and column B starting in row 7 (see appendix 2.8.3.2). This is because the expected payoff of choosing column A is higher for rows 1 to 6 and higher for column B in rows 7 to 14. Ideally, subjects will switch columns only once but it has been found that if monotonic switching is not enforced, subjects often switch multiple times especially in populations with low education Tanaka et al. (2010). Hence, most papers using this method only ask for the row in which the subject would switch. Because we want to study inconsistencies in choices, we do not enforce monotonic switching but rather ask for choices in every row. We are interested in seeing whether making mistakes (switching back and forth from Column A to Column B) changes across the three stages differentially for those who will find jobs while we control for learning or understanding the task better with the performance of the group of students in the comparison group.

[13]The lotteries elicit the three Prospect Theory parameters: risk aversion, loss aversion, and non-linear probability weighting. Prospect Theory provides a different and more general characterization of risk preferences than Expected Utility Theory.

2.8.3.1 Risk lottery based on Eckel and Grossman (2002)

Row no.	**Column A** (if heads comes out)	**Column B** (if tails comes out)
1	28,000 pesos	28,000 pesos
2	24,000 pesos	36,000 pesos
3	20,000 pesos	44,000 pesos
4	16,000 pesos	52,000 pesos
5	12,000 pesos	60,000 pesos
6	2,000 pesos	70,000 pesos

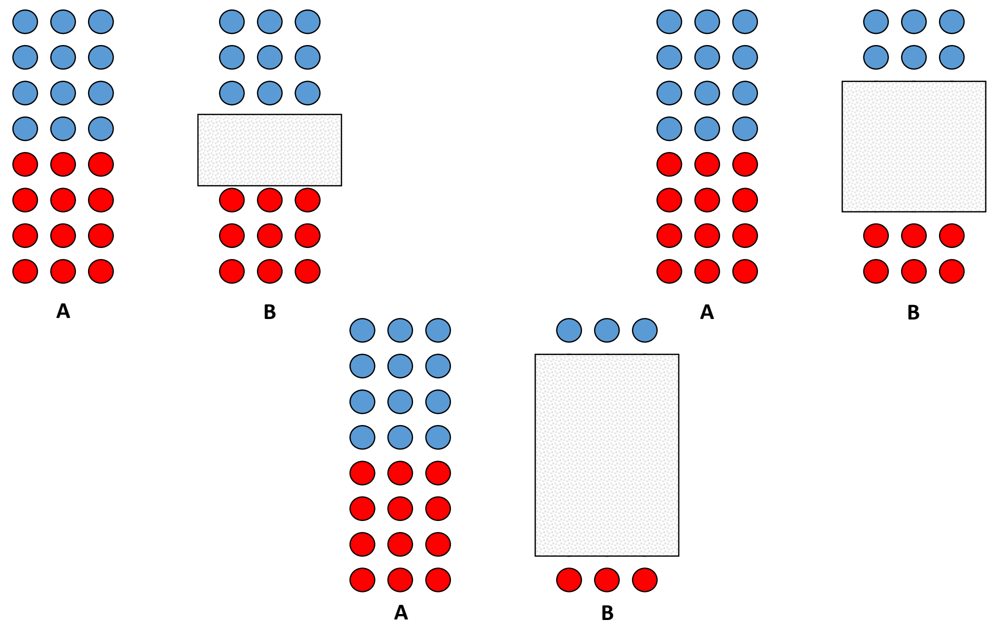
2.8.3.2 Risk lotteries based on Tanaka, Camerer and Nguyen (2010)

	Column A		Column B		Exp. payoff diff.
Row no.	**If 1 to 3 comes out**	**If 4 to 10 comes out**	**If 1 comes out**	**If 2 to 10 comes out**	(A - B)
1	4,000 pesos	1,000 pesos	6,800 pesos	500 pesos	770 pesos
2	4,000 pesos	1,000 pesos	7,500 pesos	500 pesos	700 pesos
3	4,000 pesos	1,000 pesos	8,300 pesos	500 pesos	620 pesos
4	4,000 pesos	1,000 pesos	9,300 pesos	500 pesos	520 pesos
5	4,000 pesos	1,000 pesos	10,600 pesos	500 pesos	390 pesos
6	4,000 pesos	1,000 pesos	12,500 pesos	500 pesos	200 pesos
7	4,000 pesos	1,000 pesos	15,000 pesos	500 pesos	-50 pesos
8	4,000 pesos	1,000 pesos	18500 pesos	500 pesos	-400 pesos
9	4,000 pesos	1,000 pesos	22,000 pesos	500 pesos	-750 pesos
10	4,000 pesos	1,000 pesos	30,000 pesos	500 pesos	-1,550 pesos
11	4,000 pesos	1,000 pesos	40,000 pesos	500 pesos	-2,550 pesos
12	4,000 pesos	1,000 pesos	60,000 pesos	500 pesos	-4,550 pesos
13	4,000 pesos	1,000 pesos	100,000 pesos	500 pesos	-8,550 pesos
14	4,000 pesos	1,000 pesos	170,000 pesos	500 pesos	-15,550 pesos

2.8.4 Ambiguity Aversion

Ambiguity aversion, the preference for known risks relative to unknown risks Camerer and Weber (1992); Ellsberg (1961), is another measure of economic decision making that we analyze. To implement this measure we use a task based on Tanaka et al. (2014) in which subjects must choose between a gamble whose outcome objective probabilities are known relative to one in which they are unknown. In practice, participants are presented with a series of comparisons as in appendix 2.8.4.1. In each of 9 choices, they see two urns filled with 24 blue and red balls and are instructed that they will receive the monetary reward associated to the urn they choose to play if a red ball is drawn from that urn. In urn A (left-hand side), there are always 12 red and 12 blue balls completely visible to participants and the payment in case of drawing a red ball from that urn is 20,000 pesos (about US$7) in each of the 9 choices. Urn B (right-hand side) is partially covered so that it is impossible to know the mix of red and blue balls, hence urn B is the ambiguous urn. The occluder covers 1/4, 1/2 and 3/4 of the urn depending on the choice. In the first three choices, the value of urn A and B is the same (20,000 pesos) but in subsequent choices, the value of urn B increases to 30,000 (in choices 4-6) and to 40,000 pesos (in choices 7-9) if a red ball is randomly selected. To analyze ambiguity aversion we create a variable counting the number of times the students choose the ambiguous urn from a total of 9.

2.8.4.1 Ambiguity aversion based on Tanaka et al. (2014)



2.8.5 Cognitive Performance

In terms of cognition, the bandwidth theory proposed by Mullainathan, Shafir and coauthors implies that scarcity (of time or resources) affects cognitive functioning which may compromise decision making Mani et al. (2013); Mullainathan and Shafir (2013); Shah et al. (2012). To measure different dimensions of cognition we use tasks such as a Raven's matrices-type IQ test, the Cognitive Reflection Test (CRT), the Flanker's test, and the numerical Stroop test.

The IQ test is a version of the Raven's test in which a pattern must be completed by the participant by choosing one of the choices given. This test provides a non-verbal measure of fluid intelligence which, as discussed in Mani et al. (2013), proxies the capacity to solve problems without prior knowledge. There were 9 questions in total and a time limit of 3 minutes to solve them. The test is difficult enough that very few people are able to correctly answer all questions in 3 minutes. Upon completion of the 3 minutes, participants were automatically directed to the next task. The same questions were given in Baseline 1 and after the job offer, and in Baseline 2 and after their first paycheck so the participants did not see the questions in at least 5 months.

The Numerical Stroop Test requires the subject to enter the number of digits displayed to them without getting distracted by the digit itself. For example if they see “3 3” they must respond “2” which is the number of objects displayed and not “3” which is the number that may come first to mind. This test has been used by Mani et al. (2013) and Carvalho et al. (2016) as a measure of cognitive control which is related to inhibiting inappropriate responses and selecting the appropriate information for processing. Because our surveys are taken online, our version of the Numerical Stroop Task involves using the keyboard to select the correct number of objects displayed out of 45 in total in 30 seconds. Participants receive 1,000 pesos for each correct answer if this task is selected for payment at the end of the survey.

In the Flanker Test subjects see a sequence of five arrows pointing to the left or to the right. They have to press the arrow in the keyboard that corresponds to the direction that the middle arrow in the sequence is pointing to. This test measures the ability to ignore distracting information and supress inappropriate responses. Again they have 30 seconds to correctly respond as many questions as possible.

The last cognitive measure we study is the Cognitive Reflection Test (CRT). This test measures the ability to supress incorrect intuitive and spontaneous answers to give the reflective correct answer Frederick (2005). The test usually consists of 3 questions (see appendix 2.8.5.1) but we add three more from Sinayev and Peters (2015) or change the wording of the original three questions so that it is harder for participants to recognize them from previous rounds.

2.8.5.1 Cognitive Reflection Test (CRT)

The questions that were asked in Spanish are a translation or adaptation of the following questions:

- A bat and a ball cost $1.10 total. The bat costs $1.00 more than the ball. How much does the ball cost? (Intuitive error: 10; correct: 5)
- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Intuitive error: 100; correct: 5).
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (Intuitive error: 24; correct: 47)

- Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? (Intuitive error: 15, 30; correct: 29)
- A man buys a pig for \$60, sells it for \$70, buys it back for \$80, and sells it finally for \$90. How much has he made? (Intuitive error: 10; correct: 20)
- Simon decided to invest \$8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point Simon has (a) broken even in the stock market, (b) is ahead of where he began, (c) has lost money. (Intuitive error: b; correct: c).

2.8.6 Attrition

In this section we assess the extent of attrition across different rounds and whether it can be predicted from baseline covariates or baseline outcomes. If attrition happened differentially for students with certain characteristics, some of our results in the next section could be driven by selection into staying in the sample.

For the baseline, surveys 1 and 2, we collected data from 363 participants. The aggregate attrition rate after receiving a job offer is 16.7 percent and after receiving the first paycheck, it is 21.5 percent. Effective attrition, after making the adjustments in the definition of stages discussed in the previous section, is 28.9 percent for the after offer period (Round 2) and 24.8 percent for the after paycheck period (Round 3). We consider these rates to be low among longitudinal studies. For the econometric analyses we use the sample that contains students who answered all surveys.

As expected, attrition is higher among last-semester students who eventally graduate and find jobs. Further, students who stay in the sample are more likely to be undergraduates although the statistical significance of this variable disappears when adjusting for multiple inference testing. The evidence shows that attrition is not related to baseline covariates except for the variable indicating whether the student is a last-semester student.

Because comparison group students are more likely to respond to all surveys, we examine whether outcomes measured at baseline are related to staying in the sample for the comparison and last-semester students separately. Tables 2.10 and 2.11 show the attrition test results. The dependent variable in the two tables is an indicator equal to one if the student responds all surveys. We regress that indicator on all the variables in the rows. Given the large number of regressors, we present these regressions in two tables. Appendix Table 2.10 shows the risk, time and social preferences outcomes at baseline. Similarly, Appendix Table 2.11 shows cognitive tests, perceptions on personal finances and psychological measures. The three columns of results correspond to one of three samples (all, comparison group, and last-semester students). For statistical significance, we report the usual tests without adjusting for multiple hypothesis testing (stars) and the tests adjusting using the Benjamini-Hochberg method within column (daggers).

Column 1 of Table 2.10 shows that students who remain in the sample are more likely to make non-monotonic switches in the risk lottery and less likely to be present-biased than students who leave the sample. Only the result for the significant correlation in inconsistencies in risk choices survives the multiple testing adjustment. Moreover, it is clear that this significant relationship is driven by last-semester students as can be seen in column 3. In Table 2.11 we do not see any statistically significant differences in the baseline outcomes between those who stay and those who attrit (Columns 1 to 3).

Overall, we find that last-semester students are more likely to attrit. Students who remain in the sample are more likely to make inconsistent choices in the risk lottery and, to some extent, to be less present biased. However, these correlations with baseline characteristics are not crucial in the statistical sense.

Table 2.10: Tests for sample attrition (Part 1)

	Levels in Baseline 2			**Change from Baseline 1 to 2**		
	All	Comparison	Last-sem. Students	All	Comparison	Last-sem. Students
Risk averse	0.095	0.049	-0.005	0.006	-0.042	-0.242**
	(0.086)	(0.094)	(0.131)	(0.075)	(0.088)	(0.110)
CRRA	-0.033	-0.021	-0.020	-0.015	-0.028	0.007
	(0.024)	(0.026)	(0.038)	(0.020)	(0.021)	(0.030)
Inconsistent risk lottery	0.205***†††	-0.001	0.388***††	0.017	-0.069	0.092
	(0.065)	(0.085)	(0.109)	(0.066)	(0.059)	(0.103)
Ambiguity averse	0.030	0.075	0.033	-0.031	0.023	-0.023
	(0.052)	(0.057)	(0.087)	(0.051)	(0.052)	(0.081)
Present biased	-0.002**	-0.002	-0.001	-0.002***	-0.002**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Impatient	-0.003	-0.002	0.006	-0.005	-0.011	0.003
	(0.008)	(0.011)	(0.011)	(0.010)	(0.014)	(0.012)
Non-monotonic choices	0.022**	0.008	0.026	0.014	0.001	0.005
	(0.009)	(0.010)	(0.017)	(0.009)	(0.014)	(0.013)
Fraction to student - dictator	-0.027	0.049	0.131	-0.056	0.186	-0.077
	(0.134)	(0.158)	(0.181)	(0.142)	(0.201)	(0.204)
Fraction to foundation - dictator	-0.110	-0.128	-0.192	-0.097	-0.422***††	0.159
	(0.097)	(0.120)	(0.136)	(0.129)	(0.137)	(0.173)
Fraction to student - ultimatum	0.090	0.157	-0.046	0.144	0.359	-0.052
	(0.175)	(0.253)	(0.235)	(0.169)	(0.247)	(0.242)
Constant	0.651***††	0.812***	0.533	0.583***††	0.827***††	0.306***††
	(0.198)	(0.251)	(0.337)	(0.047)	(0.062)	(0.071)
N	352	178	171	317	157	157

Notes: Standard errors clustered at the individual level in parentheses
*** p<0.01, **p<0.05, * p<0.1
Multiple testing: ††† p<0.01, †† p<0.05, † p<0.1

Table 2.11: Tests for sample attrition (Part 2)

	Levels in Baseline 2			Change from Baseline 1 to 2		
	All	Comparison	Last-sem. Students	All	Comparison	Last-sem. Students
IQ test (Raven's)	0.012	0.001	0.026	0.035***	0.020	0.052**
	(0.018)	(0.019)	(0.027)	(0.013)	(0.015)	(0.021)
CRT test	0.008	-0.024	-0.017	-0.023	-0.005	0.008
	(0.041)	(0.045)	(0.059)	(0.031)	(0.034)	(0.053)
Stroop test	-0.001	0.002	0.004	0.004	-0.003	0.011**
	(0.005)	(0.004)	(0.009)	(0.004)	(0.004)	(0.006)
Flanker test	-0.003	0.000	-0.006	-0.004	-0.000	-0.008**
	(0.003)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)
Hard to come up with money	-0.008	-0.018	-0.117	0.022	-0.039	0.025
	(0.056)	(0.066)	(0.085)	(0.065)	(0.069)	(0.108)
Hard to cover expenses	-0.020	0.064	0.027	-0.040	0.089	-0.097
	(0.074)	(0.085)	(0.108)	(0.064)	(0.072)	(0.103)
Insatisfied HH finances	-0.026	0.002	0.001	-0.035	0.040	-0.050
	(0.066)	(0.068)	(0.097)	(0.058)	(0.059)	(0.089)
Stressed personal finances	-0.033	0.047	-0.071	0.025	0.002	-0.035
	(0.064)	(0.079)	(0.095)	(0.059)	(0.049)	(0.090)
Inconsistent in the value of money	-0.031	-0.031	-0.048	-0.007	-0.039	-0.089
	(0.055)	(0.065)	(0.096)	(0.060)	(0.066)	(0.102)
Happy	0.117*	0.087	0.131	0.094	0.042	0.229***
	(0.071)	(0.092)	(0.101)	(0.058)	(0.067)	(0.086)
Frustrated	0.028	-0.054	0.035	0.017	-0.023	0.089
	(0.070)	(0.073)	(0.133)	(0.054)	(0.057)	(0.085)
Depressed	-0.032	-0.063	0.015	-0.041	-0.110*	-0.044
	(0.074)	(0.082)	(0.144)	(0.057)	(0.058)	(0.097)
Worried	0.005	-0.039	-0.021	-0.011	0.023	-0.086
	(0.061)	(0.071)	(0.088)	(0.050)	(0.055)	(0.075)
Enjoying myself	-0.109*	-0.022	-0.215**	-0.126**	-0.013	-0.182**
	(0.066)	(0.084)	(0.089)	(0.052)	(0.068)	(0.076)
Tired	0.017	-0.001	-0.013	-0.029	0.024	-0.071
	(0.058)	(0.066)	(0.084)	(0.050)	(0.052)	(0.074)
Constant	0.651***††	0.812***††	0.533	0.583***††	0.827***††	0.306***††
	(0.198)	(0.251)	(0.337)	(0.047)	(0.062)	(0.071)
N	352	178	171	317	157	157

Notes: Standard errors clustered at the individual level in parentheses
*** p<0.01, **p<0.05, * p<0.1
Multiple testing: ††† p<0.01, †† p<0.05, † p<0.1

Table 2.12: Credit and expenses

	Has a credit card	**Has a loan**	**Pays most of own expenses**
Baseline 1	0.158***	0.266***	0.207***
	(0.027)	(0.033)	(0.030)
Baseline 2	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)
After offer	0.049*	0.058*	0.100***
	(0.027)	(0.033)	(0.029)
After paycheck	0.055**	-0.005	0.116***
	(0.026)	(0.028)	(0.034)
Baseline 1 * Last-sem.	0.045	0.104**	0.187***
	(0.041)	(0.049)	(0.047)
Baseline 2 * Last-sem.	0.000	0.001	0.001
	(0.001)	(0.003)	(0.003)
After offer * Last-sem.	-0.017	-0.222*** †††	0.081
	(0.049)	(0.052)	(0.057)
After paycheck * Last-sem.	-0.005	0.089	0.108**
	(0.045)	(0.054)	(0.053)
Observations	1246	1315	1246
No. subjects	362	362	362
Controls	No	No	No

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1. Multiple Inference (Benjamini-Hochberg on after-offer and after-paycheck DID coefficients): ††† p_m<0.01, †† p_m<0.05, † p_m<0.1

2.8.7 Other difference-in-differences results adding psychological and independence controls

Table 2.13: Social preferences with controls

	Fraction given to student - dictator		Fraction given to foundation - dictator		Fraction given to student - ultimatum	
Baseline 1	0.389***	0.328***	0.602***	0.556***	0.486***	0.435***
	(0.022)	(0.018)	(0.027)	(0.023)	(0.017)	(0.014)
Baseline 2	-0.042**	-0.056***	-0.068***	-0.079***	-0.005	-0.018
	(0.017)	(0.016)	(0.018)	(0.016)	(0.014)	(0.013)
After offer	0.005	-0.088***	-0.092***	-0.163***	0.024	-0.055***
	(0.027)	(0.022)	(0.030)	(0.028)	(0.023)	(0.019)
After paycheck	-0.018	-0.117***	-0.100***	-0.172***	0.010	-0.072***
	(0.029)	(0.023)	(0.030)	(0.028)	(0.026)	(0.022)
Baseline 1 * Last-sem.	0.030	0.033	-0.038	-0.029	0.020	0.023
	(0.026)	(0.026)	(0.034)	(0.035)	(0.019)	(0.020)
Baseline 2 * Last-sem.	0.014	0.021	0.010	0.015	-0.005	0.001
	(0.025)	(0.024)	(0.025)	(0.025)	(0.021)	(0.020)
After offer * Last-sem.	0.153***	0.024	0.148***	0.025	0.105***	-0.008
	(0.044)	(0.037)	(0.045)	(0.044)	(0.037)	(0.030)
After paycheck * Last-sem.	-0.115***	-0.013	-0.036	0.046	-0.105***	-0.022
	(0.039)	(0.035)	(0.044)	(0.043)	(0.033)	(0.031)
Observations	1355	1246	1355	1246	1355	1246
No. subjects	362	362	362	362	362	362
Controls:						
Psychological factors	Yes	No	Yes	No	Yes	No
Self pays own expenses	No	Yes	No	Yes	No	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1.

Table 2.14: Perceived financial status with controls

	Hard to come up with money		Hard to cover expenses		Stressed about personal finances	
Baseline 1	0.459***	0.616***	0.141***	0.233***	0.255***	0.391***
	(0.041)	(0.037)	(0.031)	(0.032)	(0.038)	(0.037)
Baseline 2	0.036	0.074**	-0.066*	-0.037	-0.113***	-0.063*
	(0.032)	(0.031)	(0.037)	(0.035)	(0.038)	(0.037)
After offer	-0.020	0.075**	0.007	0.047	-0.068	-0.004
	(0.041)	(0.038)	(0.042)	(0.044)	(0.042)	(0.045)
After paycheck	0.000	0.066*	0.004	0.017	-0.021	0.009
	(0.041)	(0.039)	(0.038)	(0.041)	(0.043)	(0.046)
Baseline 1 * Last-sem.	-0.036	-0.031	-0.046	-0.055	0.051	0.036
	(0.051)	(0.053)	(0.041)	(0.042)	(0.051)	(0.053)
Baseline 2 * Last-sem.	-0.053	-0.071	0.075	0.061	0.054	0.032
	(0.046)	(0.046)	(0.048)	(0.047)	(0.053)	(0.051)
After offer * Last-sem.	-0.150**	-0.104*	-0.005	-0.026	-0.033	0.013
	(0.060)	(0.063)	(0.055)	(0.064)	(0.061)	(0.071)
After paycheck * Last-sem.	-0.087	-0.127**	0.012	0.002	-0.004	-0.018
	(0.063)	(0.063)	(0.056)	(0.059)	(0.066)	(0.070)
Observations	1355	1246	1355	1246	1355	1246
No. subjects	362	362	362	362	362	362
Controls:						
Psychological factors	Yes	No	Yes	No	Yes	No
Self pays own expenses	No	Yes	No	Yes	No	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.15: Time preferences (based on 16 choices)

	No. non-mono-tonic choices		**No. choices allocating full amount to sooner period**		**% present biased-ness (weighted by interest rate)**	
Baseline 1	1.671***	1.753***	3.357***	2.986***	28.598***	30.722***
	(0.287)	(0.259)	(0.368)	(0.323)	(4.091)	(3.800)
Baseline 2	-0.412	-0.416*	1.400***	1.359***	-9.059**	-9.198**
	(0.255)	(0.251)	(0.301)	(0.302)	(4.520)	(4.526)
After offer	-0.869***	-0.870***	2.489***	2.444***	-2.010	-4.145
	(0.286)	(0.284)	(0.388)	(0.382)	(4.926)	(4.730)
After paycheck	-0.946***	-0.986***	1.980***	2.121***	-6.051	-7.701
	(0.277)	(0.258)	(0.401)	(0.399)	(4.902)	(4.928)
Baseline 1 * Last-sem.	-0.013	-0.005	0.713	0.725	-4.296	-4.845
	(0.411)	(0.416)	(0.470)	(0.485)	(5.172)	(5.244)
Baseline 2 * Last-sem.	-0.474	-0.483	-0.190	-0.134	9.172	9.151
	(0.434)	(0.430)	(0.496)	(0.495)	(6.298)	(6.306)
After offer * Last-sem.	0.032	0.035	-2.037***	-1.928***	-1.942	-0.262
	(0.472)	(0.474)	(0.627)	(0.621)	(6.915)	(6.807)
After paycheck * Last-sem.	-0.162	-0.131	-0.747	-0.801	11.451	12.455*
	(0.430)	(0.424)	(0.618)	(0.634)	(7.124)	(7.190)
Observations	787	784	787	784	787	784
No. subjects	314	314	314	314	314	314
Controls:						
Psychological factors	Yes	No	Yes	No	Yes	No
Self pays own expenses	No	Yes	No	Yes	No	Yes
Cond. on correct answer to understading question	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1.

Table 2.16: Risk and ambiguity aversion with controls

	Risk averse		**Extremely risk averse**		**No. of ambiguous urns chosen (out of 9)**	
Baseline 1	0.788***	0.842***	0.661***	0.697***	3.855***	3.834***
	(0.033)	(0.028)	(0.040)	(0.035)	(0.163)	(0.143)
Baseline 2	-0.009	-0.000	-0.000	0.005	-0.040	-0.050
	(0.031)	(0.030)	(0.033)	(0.033)	(0.153)	(0.151)
After offer	-0.238***	-0.148***	-0.199***	-0.128***	0.440**	0.480***
	(0.041)	(0.041)	(0.043)	(0.043)	(0.180)	(0.179)
After paycheck	-0.266***	-0.177***	-0.261***	-0.201***	0.572***	0.615***
	(0.044)	(0.043)	(0.047)	(0.048)	(0.178)	(0.172)
Baseline 1 * Last-sem.	0.016	0.017	-0.006	-0.013	0.137	0.175
	(0.039)	(0.040)	(0.049)	(0.050)	(0.203)	(0.206)
Baseline 2 * Last-sem.	-0.088*	-0.090**	-0.040	-0.039	-0.097	-0.087
	(0.046)	(0.045)	(0.052)	(0.051)	(0.245)	(0.245)
After offer * Last-sem.	-0.123*	0.026	-0.097	0.017	-0.229	-0.212
	(0.064)	(0.067)	(0.066)	(0.071)	(0.291)	(0.289)
After paycheck * Last-sem.	0.063	-0.020	0.043	-0.022	-0.279	-0.271
	(0.064)	(0.064)	(0.070)	(0.071)	(0.286)	(0.285)
Observations	1355	1241	1355	1241	1232	1226
No. subjects	362	362	362	362	362	362
Controls:						
Psychological factors	Yes	No	Yes	No	Yes	No
Self pays own expenses	No	Yes	No	Yes	No	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1.

Table 2.17: Prospect theory parameters with controls

	Curvature of the value function		**Probability weighting**		**Loss aversion**	
Baseline 1	0.320***	0.317***	0.739***	0.739***	2.353***	2.373***
	(0.030)	(0.032)	(0.022)	(0.023)	(0.179)	(0.183)
Baseline 2	-0.004	-0.004	-0.003	-0.003	-0.355*	-0.361*
	(0.033)	(0.033)	(0.028)	(0.028)	(0.192)	(0.191)
After offer	-0.096**	-0.102***	0.022	0.025	-0.577***	-0.572***
	(0.039)	(0.038)	(0.028)	(0.029)	(0.210)	(0.210)
After paycheck	-0.179***	-0.180***	0.010	0.010	-0.795***	-0.789***
	(0.040)	(0.039)	(0.032)	(0.032)	(0.197)	(0.196)
Baseline 1 * Last-sem.	0.040	0.038	-0.007	-0.007	-0.101	-0.085
	(0.041)	(0.040)	(0.029)	(0.029)	(0.241)	(0.243)
Baseline 2 * Last-sem.	-0.032	-0.031	0.010	0.010	0.014	0.014
	(0.047)	(0.047)	(0.037)	(0.037)	(0.256)	(0.256)
After offer * Last-sem.	0.029	-0.002	-0.046	-0.044	0.129	0.038
	(0.055)	(0.058)	(0.043)	(0.045)	(0.272)	(0.276)
After paycheck * Last-sem.	-0.002	0.021	-0.046	-0.048	0.234	0.385
	(0.059)	(0.056)	(0.046)	(0.046)	(0.277)	(0.298)
Observations	857	846	857	846	743	736
No. subjects	326	324	326	324	306	303
Controls:						
Psychological factors	Yes	No	Yes	No	Yes	No
Self pays own expenses	No	Yes	No	Yes	No	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.18: Performance in cognitive tests with controls

	Raven's Test		Numerical Stroop Test		Flanker Task	
Baseline 1	0.320***	0.309***	0.739***	0.740***	2.353***	2.219***
	(0.030)	(0.035)	(0.022)	(0.023)	(0.179)	(0.177)
Baseline 2	-0.004	-0.013	-0.003	-0.007	-0.355*	-0.374*
	(0.033)	(0.034)	(0.028)	(0.028)	(0.192)	(0.200)
After offer	-0.096**	-0.103***	0.022	0.019	-0.577***	-0.566***
	(0.039)	(0.040)	(0.028)	(0.029)	(0.210)	(0.217)
After paycheck	-0.179***	-0.181***	0.010	0.007	-0.795***	-0.705***
	(0.040)	(0.041)	(0.032)	(0.032)	(0.197)	(0.198)
Baseline 1 * Last-sem.	0.040	0.043	-0.007	-0.006	-0.101	-0.109
	(0.041)	(0.041)	(0.029)	(0.029)	(0.241)	(0.239)
Baseline 2 * Last-sem.	-0.032	-0.029	0.010	0.011	0.014	0.058
	(0.047)	(0.048)	(0.037)	(0.037)	(0.256)	(0.258)
After offer * Last-sem.	0.029	0.035	-0.046	-0.048	0.129	0.194
	(0.055)	(0.056)	(0.043)	(0.043)	(0.272)	(0.277)
After paycheck * Last-sem.	-0.002	-0.006	-0.046	-0.048	0.234	0.214
	(0.059)	(0.058)	(0.046)	(0.046)	(0.277)	(0.278)
Observations	857	857	857	857	743	743
No. subjects	326	326	326	326	306	306
Controls:						
Psychological factors	Yes	No	Yes	No	Yes	No
Self pays own expenses	No	Yes	No	Yes	No	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1.

Table 2.19: Performance in Cognitive Reflection Test (CRT) with controls

	No. Correct		Minutes to answer	
Baseline 1	1.135***	1.158***	2.567***	2.592***
	(0.071)	(0.062)	(0.196)	(0.194)
Baseline 2	0.163***	0.175***	0.914***	0.902***
	(0.063)	(0.062)	(0.258)	(0.265)
After offer	0.033	0.053	-0.401*	-0.411*
	(0.072)	(0.070)	(0.225)	(0.226)
After paycheck	0.292***	0.312***	-0.040	-0.043
	(0.073)	(0.070)	(0.222)	(0.223)
Baseline 1 * Last-sem.	-0.118	-0.105	0.231	0.255
	(0.086)	(0.087)	(0.324)	(0.333)
Baseline 2 * Last-sem.	0.134	0.130	-0.475	-0.473
	(0.092)	(0.091)	(0.375)	(0.378)
After offer * Last-sem.	0.043	0.058	-0.375	-0.348
	(0.108)	(0.107)	(0.384)	(0.381)
After paycheck * Last-sem.	-0.093	-0.083	-0.471	-0.449
	(0.103)	(0.104)	(0.349)	(0.345)
Observations	1243	1241	1243	1241
No. subjects	362	362	362	362
Controls:				
Psychological factors	Yes	No	Yes	No
Self pays own expenses	No	Yes	No	Yes

Notes: Notes: Each column shows the coefficients of a DID regression for the outcomes in the top row. The coefficients on the interaction terms demonstrate the additional effect of being a last-semester student compared to a comparison student. The coefficients on the four baseline terms help establish pre-trends, while the coefficients on the after offer and after paycheck stages are the DID estimates for the corresponding stages relative to baseline 1. *** p<0.01, **p<0.05, * p<0.1.

CHAPTER 3

Unbundling the Load: How Electricity Restructuring Affects the Manufacturing Sector

The Indian power sector was largely comprised of state-owned, vertically-integrated firms that, for years, accrued large losses. Between 1998 and 2003, a number of electricity reforms, under the umbrella of the Electricity Act, were phased in, which sought to make the electricity sector more competitive and efficient. I use an event-study framework to study the effect of these reforms on the manufacturing sector, one of the biggest consumers of electricity in India. The reforms affected manufacturing firms via two main channels: the price per unit of electricity and the quality of electricity. While many expected states with multiple distributors to benefit consumers, I find that instead states that ended up with a single distributor supplied more reliable access to energy. Single distributors found it easier to raise prices, reflecting political realities. In such states, the reforms increased the amount of electricity consumed by manufacturing firms. Consistent with moving to a higher price tier, the average price paid by firms increases, and firms re-optimize their production decisions. Despite a price increase, firms respond to decreased blackouts (greater reliability) by increasing purchased electricity, worker hours and worker productivity. I find corroborating evidence of greater electrification using luminosity data (satellite night lights) that shows an increase in light density in these particular states.

3.1 Introduction

An oft quoted statistic in the case of India is that 240 million people do not have access to the electrical grid (International Energy Agency, 2015). However, what is arguably more striking is the fact that even consumers who do have access to the grid, do not receive uninterrupted electricity, with 'scheduled' power cuts being a regular feature of life in India. The 2014 World Bank Enterprise Survey found that a substantial fraction of surveyed Indian firms reported electricity outages as their biggest obstacle to production, second only to corruption. These power cuts occur not only because the generation of electricity in India is low, but because of losses in electricity due to inefficiencies in bill collections, transmission and distribution. In order to mitigate these leakages, a number of reforms were instituted in the late 90s to transition away from the state-owned, vertically integrated enterprises that characterized the electricity sector.

In 2003, the Electricity Act sought to unbundle generation from the transmission and distribution sectors. Additionally, the reforms mandated the establishment of central and state-level regulatory commissions in a move away from the traditional State Electricity Boards, which lacked transparency in their actions. This Act was followed by a number of reforms that aimed to overhaul the sector in the long run, but in this paper I focus on the Electricity Acts of 1998 and 2003, and in particular on the state-wise variation in unbundling of generation from transmission and distribution, and the establishment of regulatory commissions.

I study the effect of these reforms on manufacturing firms in India, using data from the Annual Survey of Industries between 1998 and 2012, spanning a major portion of the enactment period of these reforms. Cropper et al. (2011) show that these reforms were not successful in increasing thermal power plant efficiency, but they do find evidence of reduced power outages. I use the outcomes of these manufacturing firms as a litmus test in order to evaluate the efficacy of the reforms, given that electricity is a major input into manufacturing, and one that is considered a major constraint to production. The reforms may affect the manufacturing sector via the retail price of electricity and the quality of the electricity supply. Using the Annual Survey of Industries, I study the effect of the reforms on the average price-per-unit of electricity, total unit of electricity purchased, total output, worker productivity and own-generated electricity. I obtain satellite data on night-time lights as a proxy for electrification (Burlig and Preonas, 2017; Mann et al., 2016; Min and Golden, 2014), and use it to study the effect on power outages.

I use an event study analysis to estimate the causal effects of the reforms on manufacturing sector outcomes. I exploit the variation across states in the timing of reforms, and also the terms of reform implementation, that led to different electricity sector configurations across regions. The reforms were phased in over a six year period, and the event study framework allows me to identify impacts with fewer restrictive assumptions than a traditional difference-in-differences design. Unlike a traditional difference-in-difference where the ‘treatment’ occurs one specific year, the staggered implementation of reforms allows me to study the impacts free of coincident changes in one particular year. Year and district fixed effects control for coincident shocks common across all districts, and time-invariant unobserved characteristics. Indeed, I can test for whether these outcomes had pre-trends by studying semi-parametric estimates in a graphical format.

I study the effects across different manners of implementation as well. Importantly, different states split up the three sectors – generation, transmission and distribution – in different manners. Some partially unbundled the components, whereas others completely unbundled all electricity firms. Inadvertently, the reforms had a differential impact on manufacturing firms in different states, based on the way they were implemented. As a result of the reforms, a few states ended up with a single state-wide electricity distributor, while others ended up with multiple.

I find that in states with one unified distribution network, there was higher consumption of grid-generated electricity by establishments, as well as lower blackouts and greater reliability of supply. Manufacturing firms with access to a single distribution utility benefited in terms of increased worker hours and productivity. It may sound counter-intuitive that less competition from just a single distributor is good for consumers. Yet, with a single distribution entity for the state, it is politically and logistically feasible to raise prices and improve electricity reliability, in contrast with coordinating tariff increases across multiple non-competing utilities. This is also confirmed by Cropper et al. (2011), who find that the reforms did indeed lead to a reduction in power outages, and firms were consequently able to expand production. Such unexpected results demonstrate the unintended effects of reform interpretation and implementation, and highlights the importance of understanding these varied implications when designing such policies.

This study highlights an understudied aspect of the consequences of electricity restructuring. Without separating states by the method of reform implementation, it may appear that the effects on manufacturing outcomes are very low or non-existent. However, by exploiting descriptive administrative data which details the variation across states, this paper is able to identify and highlight an important aspect of the implementation process. State electricity utilities in India widely underprice their electricity, due to numerous subsidies. Such subsidies can sometimes lead to perverse incentives, and reduce infrastructure investment and electricity reliability (McRae, 2015). In order to sustainably supply electricity in the long-run, distribution companies often appeal to regulatory commissions to raise their prices and increase their revenues. The commercial losses due to unpaid bills are particularly high in India, making this an important step. However, electricity being an important issue to voters (Chhibber et al., 2004), there is political pressure to keep prices low (Chatterjee, 2018).

In states with multiple distribution companies, the coordination problem arising from political pressure and the need to keep electricity prices approximately equal across the state, makes it difficult to get tariff revisions approved by regulatory commissions. On the other hand, a single distribution would find it easier to negotiate such increases, by avoiding the coordination failures from multiple distributors. This paper shows that in the Indian context, these tariff increases were especially beneficial to manufacturing firms, particularly because of the improvement in electricity reliability. This result was an inadvertent consequence of how state institutions led to a particular execution of the reforms and underlines the importance of implementation in policy design.

3.2 The Power Sector in India

The Indian power sector was historically known to be characterized by insufficient generating capacity, inefficient transmission and distribution infrastructure and frequent power blackouts. The tariff structure that was in place did not allow the costs of these power plants to be covered and therefore the electricity sector was highly subsidized to keep it operational. The low prices charged to households and agricultural users of electricity were partially compensated for by charging higher prices for industrial users. However, often, these industrial consumers would generate their own electricity to avoid paying these higher tariffs. The future reforms were further precipitated by the fact that the distribution networks suffered from huge losses in power due to technical difficulties and theft (Tongia (2003)).

Even though the economic liberalization reforms in 1991 addressed restructuring in the electricity sector, they did not lead to many positive changes. The aim of these reforms had been to encourage the entry of independent power producers (IPPs) into the electricity market and make its operations more competitive and efficient. However, generation, transmission and distribution sectors largely remained vertically integrated and state owned, and continued to be considered highly inefficient. Despite these reforms, about a third of the country's population remained without any access to the electricity grid, and those who did have access, faced massive power outages. On the supply side, the power sector was still making losses in the late 1990s, well after the reforms, to the tune of Rs. 250 billion (almost $6 billion), with cost-recovery continuously declining (Pargal and Banerjee (2014)). In fact, in 2001, the Planning Commission of India oversaw a large scale bailout of the power sector to uphold their credit worthiness.

Given the unsustainability of these mounting losses, in 1998 and 2003, the Indian government passed the Electricity Acts in order to consolidate the reforms enacted in the past decade and address the issues facing the sector. Eighty-five percent of the state-owned coal-based power plants were unbundled from the transmission and distribution sectors. The main aim was to expand generation capacity and encourage entry of IPPs to reduce costs. While the generation companies remained state owned, they were now independent from the distribution and transmission companies and therefore had greater autonomy in their decisions. The EA 2003 was followed by a series of reforms which aimed to emphasize competition in the power sector, promote rural electrification and renewable energy, introduce licensed power trading and move towards a multi-year tariff framework. Restructuring also involved the setting up of a wholesale market for electricity with hourly spot markets. Independent System Operators (ISOs) were assigned to the management of regional transmission grids.

As a part of the EA 2003, apart from unbundling generation companies from the vertically integrated power firms, independent regulators were made mandatory at the central and state level (Singh (2006)). This was done to increase the internal accountability of the power utilities and separate them from the lack of transparency and politically motivated decisions that marked the sector when it was under the State Electricity Boards (SEBs) that monitored them before the reform. At the national level, a central electricity regulatory commission (CERC) was set up, while at the state level, these independent regulators, also called the state electricity regulatory committees (SERCs) were set up in at least 28 states as of 2013.

Another major goal of these reforms was to modify the tariff structure for end users and generators. Steps were taken to try and compensate for the subsidized rates offered to agricultural and household users by charging industry users a higher tariff. The EA also made provisions for choice at the consumer end by allowing multiple retail licenses (often indistinguishable from distribution licenses in the Indian context, where these terms are used interchangeably), and even for parallel distribution lines. This lowers the barrier to entry at the retail end. Private retailers (or distributors) were also allowed to use the state utility's distribution lines by paying a wheeling charge, but in practice retail competition did not really take off except in a few isolated cities.

3.3 Prior Work on Electricity Reforms

In the US, like in India, restructuring occurred at state level, with there being variation in the timing and extent of various policies. However, in most states that it did occur, it included the vertical disintegration of the generation, transmission and distribution sectors. This resulted in either complete separation of these sectors or at the very least limited upward and downward flow of information.

There is mixed evidence from the work has been done to examine the effect of restructuring in the

US, particularly on power plant outcomes. A lot of these papers identify the effects of restructuring by exploiting the variation in the roll-out of reforms across states, or by comparing neighboring states where one enacted the restructuring and the other did not. Fabrizio et al. (2007) use variation in the timing and outcomes of state-level restructuring efforts to show there are large medium-run gains in thermal plant-level efficiency and that market-based incentives and competition play an important role in promoting technical efficiency. Knittel (2002) suggests that power plants facing compensation schemes with performance incentives were more efficient than plants compensated on a traditional cost-plus basis. Davis and Wolfram (2012) find that the selling of nuclear reactors to independent power producers led to a decrease in forced outages at nuclear power plants and a corresponding increase in electricity production. Cicala (2015) compares generation firms that completed divestiture and nearby firms that did not, in order to show that doing away with cost-of-service regulation led to generating firms making an effort to lower their input costs. Similarly, Apt (2005) focuses on the effect of encouraging retail competition on prices, and finds no evidence of lower industry prices in states that restructured the electricity sector. Mansur (2007) also uses variation in the extent of restructuring across states to show how restructuring of the electricity wholesale market in Pennsylvania, Maryland and Maryland led to anti-competitive behavior among large sellers.

In India, Cropper et al. (2011) study the impact of the unbundling of generation on the efficiency of state-owned power plants in India. Similar to my empirical strategy, they also take advantage of the variation in the timing of reforms and use a difference-in-differences methodology to estimate the effect on thermal power plant efficiency. They hypothesize that separating generation from transmission and distribution should lead to greater responsiveness to price incentives and greater efficiency in resource allocation within these thermal plants. They expect that would increase plant availability and reduce forced outages. They study the effect of unbundling using a panel dataset of thermal power plants from information collected by the Central electricity Authority (CEA) of India from 1994-2008. After controlling for state-level time trends, year and plant fixed effects, they find that unbundling led to increased annual plant availability by 4.6 percentage points and reduced forced power outages in generation firms by 25% in states that had unbundled before 2003. However, they find that restructuring did not improve thermal efficiency, and they attribute this to the fact that unbundling had not attracted IPPs into the market to the same extent as in the US. They state that despite the setting up of SERCs at state level, it appears that conditions are not yet optimal for IPPs to enter and there is a need for greater separation of the SERCs from state political processes.

Sen and Jamasb (2012) on the other hand, find that the average plant load factor (PLF) of power plants goes up substantially in states that unbundled their electricity sector. Cropper et al. (2011)

find however, that these large effects dissipate on controlling for state and year fixed effects. Sen and Jamasb (2012) do emphasize repeatedly that the political processes in each state play a major role in determining the success or failure of deregulation, and that the nature of unbundling in each state depends qualitatively on this factor. Bacon and John (2002) further stress that unbundling is not very effective if unaccompanied by a major overhaul of the tariff structure. In India, by 2006 only a third of the states that had unbundled were breaking even (The Energy and Resources Institute 2009).

Therefore, the evidence from India appears to suggest that while restructuring may have led to reductions in blackouts, the effect on thermal efficiency of power plants is mixed. In addition, the effect of restructuring on wholesale and retail price of electricity is unclear. This paper proposes to study the effect these reforms had on the manufacturing sector in India. The manufacturing sector is one of the largest paying consumer bases of the electricity sector and would be affected by these reforms through two main channels: the price of electricity and the quality of the electrical supply (power outages, voltage fluctuations). In this regard, looking at how the manufacturing sector reacted to changes caused by restructuring would provide a yardstick of the success or failure of these reforms. I am able to use self-reported data by manufacturing firms on their electricity consumption, average price paid, own-generated electricity, total output and labor inputs in order to assess the effect of the reforms. In the same style as the literature from the U.S., I use variation in the timing and the extent of restructuring reforms to identify their effect on manufacturing.

Power outages are a major constraint to manufacturing in India. Allcott et al. (2016) study the effect of electricity shortages on Indian manufacturing firms and find that power shortages cost the average plant 5 to 10 percent of its revenues. Furthermore, these shortages more severely affect firms without generators, likely the smaller plants. If the Indian reforms can continue to reduce power outages, this should affect firm production. With uninterrupted electricity supply it is also more likely that a firm would re-optimize production, and outcomes such as its total output, productivity and labor input should also be affected. Alam (2015) finds that power outages lead to losses equivalent to 6.4% of annual sales. However, firms may still be perceived to not be affected by blackouts if they are able to adjust their behavior to cope with it. Alam (2015) explains how rice mills, for instance, unlike steel mills adjust to blackouts by altering their input mix. Therefore, there is a large degree of heterogeneity across types of firms in how they can cope with these blackouts. To take this cross-industry heterogeneity into account, I use industry fixed effects in all my regressions. The limitation of this strategy, however, is that it cannot account for how different industries will adapt to changes stemming from the reforms.

3.4 The Possible Consequences of Power Sector Reform

One of the main channels through which electricity reforms may affect the manufacturing sector is the retail price of electricity for industrial or commercial users. However, the effect on this retail price is not obvious. There may be a direct effect of restructuring on the retail price, as well as an indirect effect, through the wholesale price of electricity.

In India, just like in the U.S., the establishment of a competitive wholesale market for electricity gives rise to the potential for market power. This may be caused by a number of factors such as the concentration of generation in isolated regions where the import of electricity is limited, geography-based transmission constraints and the low elasticity of demand for electricity (Borenstein et al. (2002); Joskow (1997)). There is evidence from various countries that market power in generation, following reforms that encouraged privatization and competition, led to huge increases in the wholesale price of electricity (Borenstein et al., 2002; Joskow and Kahn, 2002; Wolfram, 1999). These increases in wholesale price then carry forward into increases in retail prices to end-use consumers as well.

However, the reforms could also cause a decline in commercial prices for electricity. This could happen if the privatization and competition encouraged in the generation sector leads to a lowering of costs, which could then get passed on to consumers through lower prices (Bushnell and Wolfram, 2005; Cicala, 2015; Markiewicz et al., 2004).

Therefore the evidence on the effect of reforms on wholesale prices, in the Indian context is unclear. On the one hand, the Indian government delicensed thermal generation, opening up the sector to private participation and introduced wholesale power trading, after vertically disintegrating generating units from distribution. This could have the effect of increasing the market power of some of the generating units, leading to an increase in the wholesale price. Divested generation plants now have the freedom to sell power on the wholesale market, rather than being forced to sell downstream within their own firm. On the other hand, this delicensing and entry of private participation in generation may lead to greater efficiency in electricity production, which may get passed through to wholesale prices.

As far as retail prices are concerned, once again, the effect of the reforms is ambiguous. Some states executed the tenets of the reform by completely separating the generation, transmission and distribution units. Some of these states ended up with a single distribution company after executing their reform legislation, and some states ended up with multiple distribution companies. In India, distribution companies are, in most cases, also involved in retailing electricity. In fact, utilities in most states suffix their names with “discom” in reference to the fact that they also operate at the distribution end. The Electricity Act (EA) 2003 provided for consumer choice in terms of the utility that supplies power to them. Given that most utilities were also distribution companies, in order to encourage competition and choice, the Act mandated that any utility could use existing distribution infrastructure for a minor wheeling fee. The Act also provided for new distribution infrastructure to be set up in parallel, but there is very low take up of this.

However, this legislation has met with very limited success. One of the biggest reasons for the lack of separation between the distribution and retail supply of electricity in India is that it is almost impossible to accurately attribute distribution losses to the distribution and retail suppliers separately, because several states still have many unmetered consumers. Ideally, technical losses should be attributed to the distribution network, while commercial losses arising because of faulty meters and non-metering should be attributed to the retailer. In the absence of advanced metering, it is hard to encourage competition in the retail sector of electricity (Forum of Regulators (2013)). There is limited scope for retail prices to fluctuate too much, because they are still controlled by State Electricity Regulatory Commissions. After the Electricity Act was passed, concerted efforts were made to charge tariffs that could at least cover all costs, with limited success. Raising tariffs is politically very unpopular, and several states tariffs still remain too low to cover all costs.

The data that this paper uses on electricity price is derived by calculating the average price per KWh from self-reported data from manufacturing firms on the total electricity purchased by them and total amount spent on the purchase. Therefore, this price data, to some extent, masks the effect of the reforms on wholesale price, as well as the exact tariff set by the electricity retailer. Therefore, the results I obtain based on reduced-form effects of restructuring on this calculated price can be considered as suggestive, and further analysis is needed using data on actual wholesale prices and retail tariffs in order to further study the effect of the electricity reforms on the manufacturing sector.

However, price is not the only channel through which manufacturing firms may be affected. For instance, Cropper et al. (2011) find that the reforms reduced the amount of forced outages in the electricity sector. In 2006, three years after the EA was passed, 35% of establishments surveyed under the World Bank Enterprise Survey reported electricity as the biggest obstacle to production, second only to corruption. Furthermore, this survey also reported that the annual losses due to electrical outages was as much as 6.6% of annual sales. Therefore, if the reforms went on to reduce forced outages, given that all states had not restructured by 2006, this could increase the consumption of electricity at the firm level. Firms may scale down own generated electricity and consume more electricity from the grid. However, because I do not have data on the fall in power outages, I am unable to examine this mechanism.

3.5 Data and Variables

My main source of data on manufacturing firms is from the Annual Survey of Industries (ASI) for the years 1998-99 until 2011-12. This data therefore ranges from the period covering the 1998 Electricity Act, through the 2000s which saw the enforcement of several tenets of the 2003 Act.

This dataset consists of several modules, from which I extract the data that is relevant to this paper. I use data on the number of days a firm is operational, number of worker days, wages, output produced and inputs to the firms. Among the input data, I make use of the comprehensive information on electricity consumed.

To study the impact of the restructuring, one of the key variables I look at is the price paid by the firm for per unit electricity. Since there are a lot of missing values and discrepancies (the price is reported as zero a few times, even though the total purchase value is a large rupee amount) in the data across years, I calculate this using other available data. I calculate the rate per unit by dividing the total purchase amount of electricity by the number of units consumed. When either one of these variables is missing, I use the reported rate per unit when it is available. In several cases, quantity of electricity consumed is reported as 0, but the price of electricity is reported. In these cases, I report these observations as missing with respect to the price of electricity. I lose about 1% of observations through this step, because of the number of firms who report consuming no electricity.

According to the data, around a third of the firms surveyed had some electricity generation capacity and among those who did, they were able to self-generate about 20% of their total electricity consumption. Clearly, only the large firms have generating capacity, and in fact the mean amount of electricity generated by them is higher than the full sample mean of electricity consumed.

Table 1 presents summary statistics of the main variables I use for this paper. I define two measures of productivity of workers. I call one of them wage productivity, which is simply the total wages divided by the total workers days. Similarly, I define output productivity and the total value of output (sum of finished and semi-finished output value) divided by the total workers days. I compute worker days by summing across all the different employees across each firm. I perform the same calculation to arrive at firm level total wages.

Several variables such as the quantity of electricity consumed and generated, total worker days, value of output and wages have outliers. To deal with this in a uniform way, I winsorize the data for these variables.[1] The calculated per unit prices of electricity I calculate are consistent

[1] To elaborate, I assign the bottom most percentile of observations for each of these variables, the value of the 1st

Table 3.1: Summary statistics for outcome variables of manufacturing firms (in a year)

Restricted Sample			
Variable	**Obs.**	**Weighted mean**	**SD**
No. of manufacturing days	436151	219.87	123.74
No. of non-manufacturing days	436151	8.38	38.32
No. of total operational days	436151	245.16	108.96
Quantity of electricity consumed (KWh) (mill.)	399344	0.45	1.28
Price of electricy consumed (Rs. per KWh)	399344	4.85	1.36
Quantity of electricity generated (KWh) (mill.)	139927	0.24	0. 84
Total number of worker days	415134	16080.13	32747.04
Total wages (all employees) (Rs.) (mill.)	415134	4.05	11.6
Value of finished goods (Rs.) (mill.)	294132	6.39	20.5
Worker productivity (wages/worker days)	415134	186.09	368.3
Worker productivity (total output/worker days)	151760	526.57	1934.58

Note: SD - standard deviation of variable. Obs. - number of observations. Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations. Exchange rate approximately Rs. 66/$

with average annual tariff prices for manufacturing firms reported by the World Bank open data website.

I get my information on the unbundling status of various states from Pargal and Banerjee (2014). I have included the details of how various states acted on the reform in the appendix (table A1).

I restrict the dataset to only those firms that are privately owned. This is because there government-owned firms may get preferential rates, and would not accurately reflect the effect of these reforms. Furthermore, I restrict the data to include firms from only relatively electricity-intensive manufacturing industries (full list in table A2). And finally, I only include states for which the data has more that 1000 observations (firm-year level).

Finally, I also make use of luminosity data from the Defense Meteorological Satellite Program's Operational Linescan System. This data is constructed as an annual average of satellite images of the earth taken daily between 20:30 and 22:00 local time. The raw data is at a 30 second resolution, which implies that each pixel in the raw data is roughly one square kilometer. I average over pixels within an administrative boundary – the smallest of which in this context is the sub-distract. The raw luminosity data for each pixel is reported as a six-bit integer ranging from 0 to 63. To be

percentile, and do the same for the observations with values above the 99th percentile, by assigning them the value of the 99th percentile.

consistent with the literature (Henderson et al., 2012), I create a variable that is the log of the light-density per capita, where population data is taken from Landscan. Since the average within a district is never 0, the natural log is taken. This data is usually used in economics as a proxy for economic activity. In the absence of any data on quality of electricity and blackouts, I use this data merely as a check to see if it is consistent with my results.

3.6 Empirical Strategy

Using an identification strategy similar to Cicala (2015); Cropper et al. (2011); Markiewicz et al. (2004), I exploit the variation in timing and the exact nature of the restructuring in various Indian states in order to find their effect on electricity price and other manufacturing outcomes. I also use the fact that different states executed the reforms differently. I expand on these different kinds of restructuring in the data section, and use these variables in order to identify (i) the marginal effects of separating only generation and distribution from transmission, (ii) separating all three sectors and setting up a single distribution company, and finally, (iii) setting up multiple distribution companies.

Given the various ways in which reforms could affect the electricity sector, I use a reduced form equation to help estimate the net effect of unbundling and the setting up SERCs on various outcome variables. This would capture the effect on the manufacturing sector regardless of the channel through which restructuring could affect the manufacturing sector. I conduct an event study analysis in order to understand the effects of these reforms on the outcomes of interest. These outcomes include the unit price of electricity, the amount of electricity consumed by a firm, their own-generated electricity, days worked, total output of the firm, and measures of productivity. I use the following specification for my event study.

$$
\begin{aligned}
Y_{fist} = \gamma + & \sum_{z=1}^{3} \left[\sum_{y=-9}^{-1} \delta_{zy} Unbundled_{zs} \mathbb{1}(t - T_s = y) + \sum_{y=1}^{7} \delta_{zy} Unbundled_{zs} \mathbb{1}(t - T_s = y) \right] \\
& + \theta_s + \theta_t + \theta_i + \nu_{fist} \qquad (3.1)
\end{aligned}
$$

The outcome variable is referred to by Y_{fist} where f specifies the firm level, i specifies the industry group, s the state and t the year. The variable $Unbundled_{zs}$ refers to the nature of restructuring in a particular state s. When $z = 1$, it is an indicator variable that takes the value 1 for states where generation and distribution were separated from the transmission sector. This occurs in all states that acted on the reform. When $z = 2$, it is an indicator variable that takes the value 1 for states, that additional to the separation referenced by $z = 1$ also separate generation from distribution, i.e. separate all three sectors, and end up with one state-wide distribution company. When $z = 3$, it is an indicator variable that takes the value 1 for states that additional to the steps taken into account by $z = 1$ and $z = 2$ also set up multiple distribution companies, and zero for others. So, for example, the state of Maharashtra restructured in 2005 ($T_s = 2005$), and separated all three sectors (generation, transmission and distribution), resulting in a single state-wide distribution company ($Unbundled_1 = 1$ and $Unbundled_2 = 1$). Here, when I have a firm level observation at year 2000 ($t = 2000$) the expression within indicator function in Equation (3.1) is 1 when $t - T_s = -5$, Therefore, we would pick up effects of unbundling in the year 2000 and state of Maharashtra through $\delta_{1,-5}$ and $\delta_{2,-5}$.

Table 3.2: Explanation of indicator variables used in event study

	Variables		
Features	$Unbundled_1$	$Unbundled_2$	$Unbundled_3$
Generation and distribution separated from transmission	1	0	0
Generation, transmission and distribution separated, single distribution company	1	1	0
Generation, transmission and distribution separated, multiple distribution companies	1	1	1

This regression specification also includes state, year and industry fixed effects, as well as a linear time trend for each state. Given that my data is based on self-reported prices and outcomes by manufacturing firms, I use these fixed effects to control for year-specific, state-specific and industry-specific differences. This works with varying degrees of success for the outcome variables I consider.

The EA 2003 also mandated that each state set up an independent State Electricity Regulatory Commission (SERC). There is variation in the timing of the set up of these regulatory bodies across various states, and this does not often coincide with the year they executed restructuring reforms. I also run the following specification in order to further soak up any variation that may be due to the set up of the SERC. I include the indicator variable $SERC_{st}$ which takes a value of 1 for all years including and after the year y in which a state s sets up the SERC, and 0 otherwise.

$$Y_{fist} = \gamma + \sum_{z=1}^{3} \left[\sum_{y=-9}^{-1} \delta_{zy} Unbundled_{zs} \mathbb{1}(t - T_s = y) + \sum_{y=1}^{7} \delta_{zy} Unbundled_{zs} \mathbb{1}(t - T_s = y) \right] + SERC_{st} + \theta_s + \theta_t + \theta_i + \nu_{fist} \tag{3.2}$$

The final results I present in the next section are created using regressions with Equation (3.2). There is some possibility of endogeneity in the timing across states, as well as the exact nature of restructuring in various states. For instance, it may be possible that the states that restructured earlier were the states which had the greatest increase in propensity to improve efficiency of thermal power plants or reduce outages. Or, it may be true that the states that unbundled first also had increasing institutional capacity to allow for higher efficiency in generation. The availability of data limits my ability to address these completely. My results on outages, however, are consistent with those in Allcott et al. (2016); Cropper et al. (2011), who use different identification strategies.

Given that the source of variation in this data is at state level, I run a robustness check where I reduce the dataset to state-year level, by averaging the outcomes of interest to state level. I then run the same regression as above (without industry fixed-effects and linear state time trends), weighted using the frequency of firms in each state. These results are consistent with the results from the specification in Equation (3.2), albeit noisier. This is understandable given that there are far fewer observations when I reduce the dataset to state-year level. The results from these regressions are included in the appendix (Figures 3.10 and 3.11). I also use this specification to look at luminosity data as a check to see if it is consistent with my other results (figure in next section).

$$Y_{st} = \gamma + \sum_{z=1}^{3} \left[\sum_{y=-9}^{-1} \delta_{zy} Unbundled_{zs} \mathbb{1}(t - T_s = y) + \sum_{y=1}^{7} \delta_{zy} Unbundled_{zs} \mathbb{1}(t - T_s = y) \right] + SERC_{st} + \theta_s + \theta_t + \nu_{st} \tag{3.3}$$

I try to mitigate concerns of endogenous timing of the reforms as far as possible by controlling for state fixed effects, year fixed effects, industry fixed effects and state-time trends. I can reasonably argue for conditional exogeneity considering that in Figure 3.2, the pre-trends of average electricity price are similar across all states, after controlling for the above fixed effects. The event study graphs make it easy to study these pre-reform trends in order to determine the validity of this design. Since there are no differential pre-reform trends, we should expect that in the absence of the reform, the trends would be parallel across regions, conditional on the controls.

I also consider an alternative model for the event study, which forces there to be no differential pre-trends for the different types of restructuring. Imposing the condition that there are no pre-reform trends is a very strong restriction, given that I do not see clear evidence in the data that there are no pre-trends, particularly for the outcome variables other than electricity price. However, I use the results from this specification only to test whether the effect of a particular type of restructuring is jointly different from 0 for all the post-treatment years. I use the following specification for this model:

$$\begin{aligned} Y_{fist} = \gamma + & \sum_{z=1}^{3}\left[\sum_{y=1}^{7} \delta_{zy} Unbundled_{zs} \mathbb{1}(t - T_s = y)\right] \\ & + SERC_{st} + \theta_s + \theta_t + \theta_i + \nu_{fist} \end{aligned} \tag{3.4}$$

I show results using the above specification in the Appendix. I use these results only to check whether there is a significant aggregate effect of restructuring on the outcomes of interest.

3.7 Results

In this section, I present the results of the event study analysis using Equation (3.2). In Figure 3.1, I study the effect of restructuring on the intensive margin of electricity consumption. While, there there is some weak evidence of pre-reform trends, the figure still demonstrates a trend break after restructuring.

The line representing complete separation, with one distribution company (coefficients of $Unbundled_2$: δ_{2y}) shows the additional effect of completely separating the three sectors over merely separating generation and distribution from transmission. Therefore, the total effect of complete separation can be calculated by adding the coefficients of $Unbundled_1$ and $Unbundled_2$ (δ_{1y} + δ_{2y}). The total effect in the case of this kind of complete separation with one distribution appears to be positive. The net effect of complete separation with multiple distribution companies on electricity purchase appears to be zero (adding the coefficients of $Unbundled_1$, $Unbundled_2$ and $Unbundled_3$: δ_{1y} + δ_{2y} + δ_{3y}). The line representing the baseline level of restructuring, i.e. only separating generation and distribution from transmission (δ_{1y}) stops two years after the treatment year. This is because only 4 states had this limited level of structuring, and the last year available in the data for these states is a maximum of two years after restructuring.

Similarly, in the case of the unit price of electricity in Figure 3.2, the net effect of complete sep-

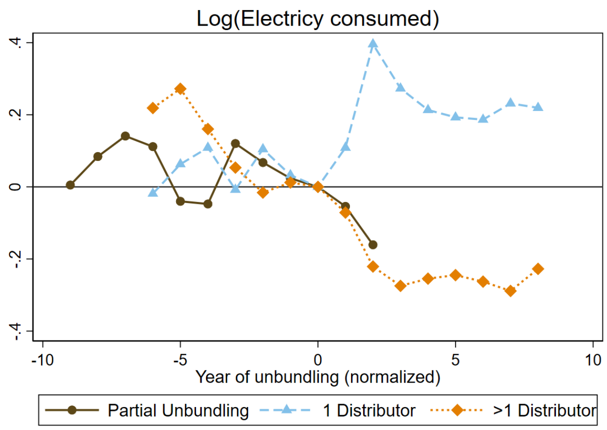


Figure 3.1: The effect of restructuring on electricity purchase (intensive margin)

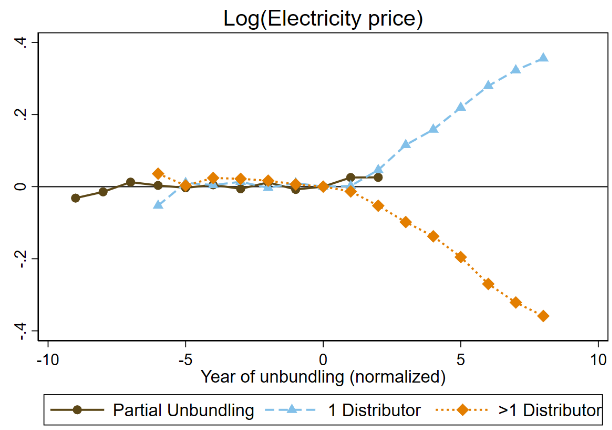


Figure 3.2: The effect of restructuring on electricity price

Note: Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The graphs plot coefficients from regressions following Equation (3.2). The left hand panel shows the results from a regression with 507,224 firm level observations, while the right hand side panel shows results from a regression run on 507,942 observations. For confidence intervals, please see Figures 3.12 and 3.13 in the Appendix. Standard errors clustered at the state level.

aration with one distribution company is positive, while there is no net effect of any other type of restructuring. These results indicate that in states where there was complete separation with one distribution company, manufacturing firms increased their consumption of electricity. This is consistent with prices increasing in these states if higher consumption levels pushed these firms to higher tiers of the price tariff, raising the average price per unit of electricity consumed due to increasing block price tariffs. I have included the confidence intervals for the additional effect of $Unbundled_2$ (where generation, transmission and distribution are separated, and one distribution company is set up) in Figures 3.12 and 3.13. These effects are significantly different from zero, in the post-periods.

However, it is unclear why such a strong effect may only be observed in states that completely unbundled all three sectors, but formed only a single distribution company rather than more. One possible reason is that there was a differential improvement in the quality of the electricity supply (reduced blackouts) in states that passed this specific legislation. This could plausibly result in firms purchasing more electricity if it is more reliable than before. This may further lead to re-optimization of other inputs within the firm in such states. Given that a high proportion of firms state that electricity blackouts are a major constraint to their production, an improvement in the quality of electricity supply may well lead to such behavior. Figure 3.3 supports this hypothesis. It shows an increase in lights density in states completely unbundled all three sectors, and formed only a single distribution company. This evidence is consistent with the hypothesis that blackouts

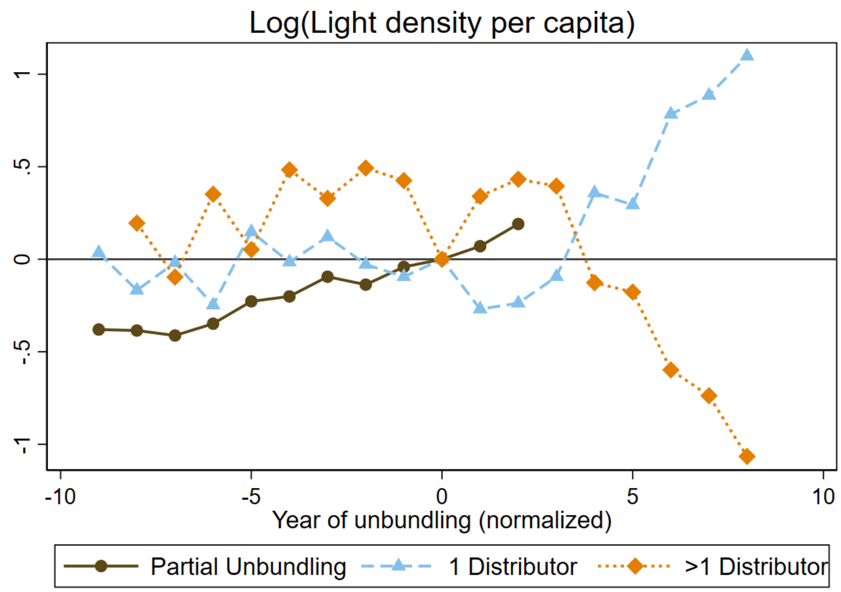


Figure 3.3: The effect of restructuring on luminosity (satellite data)

Note: This graph uses satellite nighttime light density data as a proxy for electricity consumption. I use 228 observations to run a regression following Equation (3.3), and the coefficients are plotted in this graph. For confidence intervals, please refer to Figure 3.14 in the Appendix. Standard errors clustered at the state level.

may be differentially falling in these states. However, it may merely be reflective of increased economic activity in these states, which also strengthens the findings in Figure 3.1 where there is an increase in electricity consumption among manufacturing firms, likely increasing their economic activity.

This does not rule out the fact that price tariffs may have also gone up for all firms. If so, this would be consistent with Figures 3.4 and 3.5.[2] While some firms may respond to reduced blackouts by increasing electricity purchase, even if average electricity price tariffs increase to some extent, there may be smaller firms who respond more to the price of electricity, as opposed to the quality. For these firms, a reduction in electricity purchases, as well as an increase in own-generated electricity is plausible.

Figures 3.4 and 3.5 indicate a decline in the purchase of electricity on the extensive margin, as well as an increase in own-generated electricity on the extensive margin in those states that have complete separation of the three electricity sectors, with one distribution company. This would be consistent with small firms stopping electricity purchases when prices go up, and generating their own electricity. Once again, the net effects of restructuring for states that enacted other

[2]Figures A5 and A6 include the confidence intervals for the coefficients of $Unbundled_2$

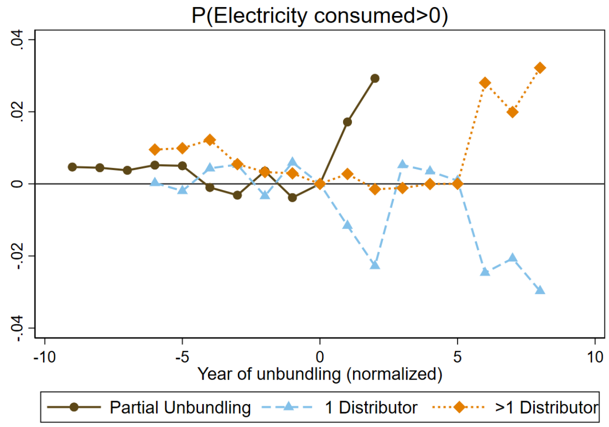


Figure 3.4: The effect of restructuring on electricity purchase (extensive margin)

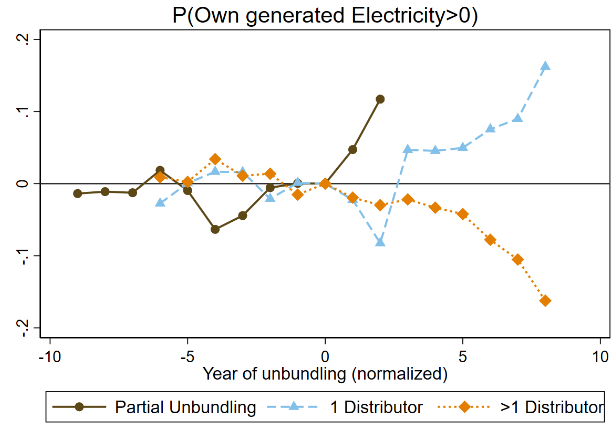


Figure 3.5: The effect of restructuring on the probability of using a generator

Note: Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The graphs plot coefficients from regressions following Equation (3.2). The left hand panel shows the results from a regression with 514,724 firm level observations, while the right hand side panel shows results from a regression run on 520,184 observations. For confidence intervals, please see Figures 3.15 and 3.16 in the Appendix. Standard errors clustered at the state level.

interpretations of the reforms appear to be zero.

The results in Figures 3.6 and 3.7 [3] appear to lend credence to the idea that larger firms for whom the quality of electricity is a major factor, increase their electricity consumption, and re-optimize production. These figures show that the days spent on non-manufacturing activity (these are manufacturing firms) falls, while the total worker days rises. Once again, this occurs only in states with complete unbundling and one distribution company. This may be interpreted as evidence that these manufacturing firms also increase their labor input in addition to electricity purchases. This is consistent with the observation that blackouts were a major constraint to their production, since they may have had to shut down production on days when there was no power. However, if the electricity supply improved in quality, there would be more days during which their workers would be able to perform their duties, particularly if they are electricity intensive firms.

This hypothesized re-optimization by larger firms is borne out in the Figures 3.8 and 3.9 as well, where an increase in electricity purchases, and subsequent increase in worker days is accompanied by an increase in the total output as well as output productivity (units of output per worker day)[4]. Again, this occurs only for firms in states where there is a single distribution company, and complete separation of generation, transmission and distribution. This increased economic activity is consistent with the findings on the lights data in Figure 3.3.

[3]Figures A7 and A8 include the confidence intervals for the coefficients of $Unbundled_2$

[4]Figures A9 and A10 include the confidence intervals for the coefficients of $Unbundled_2$

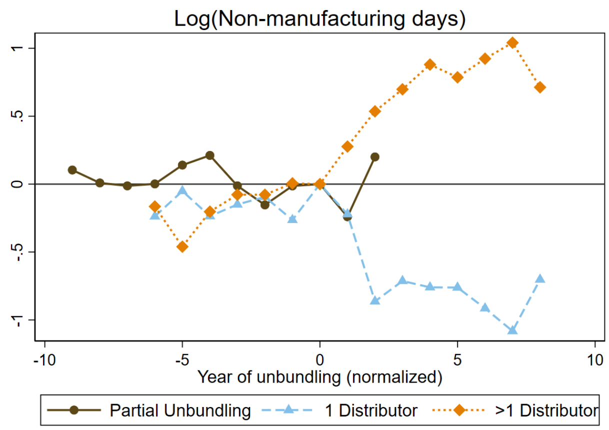


Figure 3.6: The effect of restructuring on non-manufacturing days

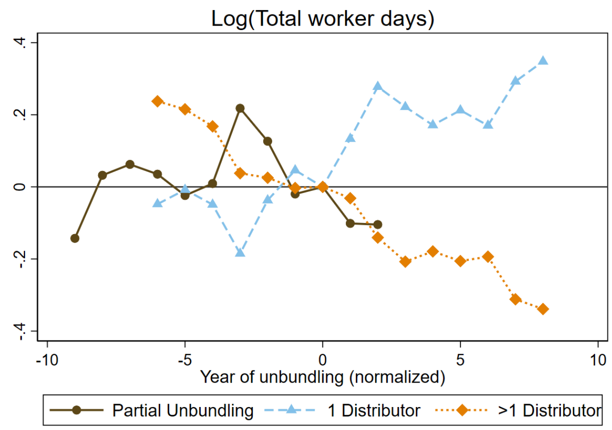


Figure 3.7: The effect of restructuring on total worker days

Note: Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The graphs plot coefficients from regressions following Equation (3.2). The left hand panel shows the results from a regression with 80,750 firm level observations, while the right hand side panel shows results from a regression run on 506,734 observations. For confidence intervals, please see Figures 3.17 and 3.18 in the Appendix. Standard errors clustered at the state level.

The above hypothesis would explain these results, but it is still not clear why these patterns would differentially occur only in states with a single distribution company. Why would they not occur, for instance, with states with multiple distribution companies, but with the same separation of the three sectors? One explanation may owe itself to the peculiarities of the Indian experience with restructuring. As mentioned before, there is a lot of political opposition to raising electricity prices in India. However, continuing not to raise tariffs even in the presence of net losses perpetuates a cycle of poor quality electricity and frequent blackouts. It may be possible that in states with only one distribution company, a hike in electricity tariffs, at least to cover all costs, is administratively easier to pass and justify. With multiple distribution companies, with varying levels of overhead and therefore costs, it may be harder to explain price tariff increases in one distribution area versus another. Therefore, such states may be constrained by one or two distribution companies having very low costs, and therefore being able operate even with lower tariffs. However, having different tariffs in different parts of the same state may be politically less palatable. Unfortunately, due to lack of data on quality of electricity, distribution company data, wholesale and retail price tariffs, this remains conjecture.

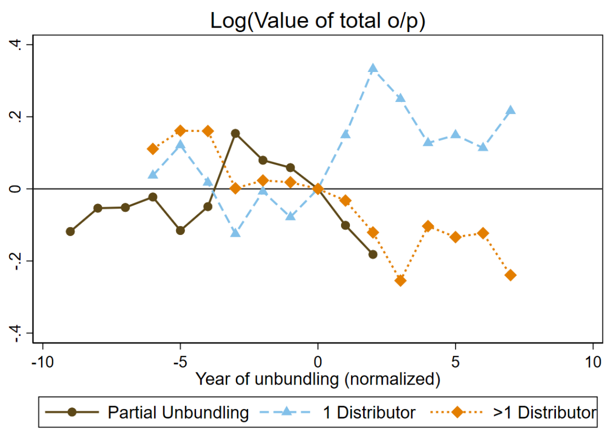


Figure 3.8: The effect of restructuring on total output

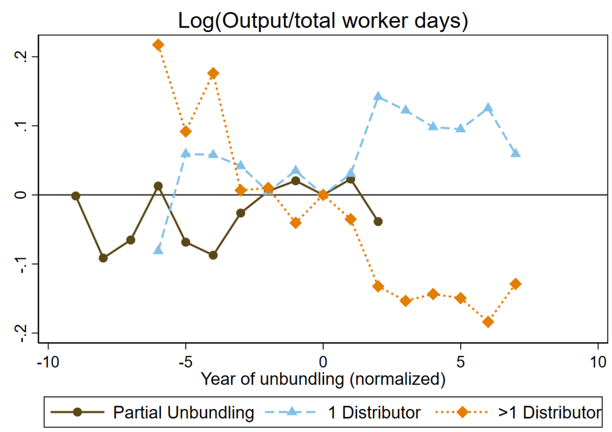


Figure 3.9: The effect of restructuring on total output productivity

Note: Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The graphs plot coefficients from regressions following Equation (3.2). The left hand panel shows the results from a regression with 382,953 firm level observations, while the right hand side panel shows results from a regression run on 186,652 observations. For confidence intervals, please see Figures 3.19 and 3.20 in the Appendix. Standard errors clustered at the state level.

3.8 Conclusion

The Indian experience with restructuring appears to have mixed results. Given the data limitations faced in this analysis, it is difficult to study the detailed effects of restructuring on wholesale and retail prices, as well as the incidence of blackouts and voltage fluctuations. However, in spite of using a calculated measure of the average price of electricity, the data reveals interesting patterns.

Nineteen out of thirty-two states and union territories experienced some form of restructuring in the electricity sector. Some of these states separated their generation and distribution sectors from the transmission sectors. A few states completely unbundled generation, transmission and distribution, ending up with a single distribution company, while other set up multiple distribution companies. The main motivation behind all these types of restructuring was to encourage more private participation and competition in generation and electricity retail, while reducing inefficiencies in electricity production and recovering costs of production.

This paper uses manufacturing sector outcomes as a means of evaluating the performance of the electricity reforms. The manufacturing sector is one of the largest paying consumers of electricity in India, and would therefore stand to be affected if the reforms affects the price and quality of electricity. In the absence of specific data on wholesale and retail prices or electricity, as well as information on blackouts, I use the derived average price of electricity in order to evaluate the reforms. However, even after using this approximated price, which masks the effect of the reforms on wholesale and retail tariffs, I find that in states that completely unbundled generation, transmission and distribution (only one distribution company) there is evidence of an increase in electricity purchases on the intensive margin. This is accompanied by higher average prices of electricity, possibly because an increase in electricity consumption pushes firms to a higher tier of an increasing block pricing tariff. Firms in these states also experience a rise in total output, as well as productivity. I find that in these states, there is also a differential increase in luminosity, which I interpret as a proxy for economic activity. In this context, however, while it does not necessarily imply a fall in blackouts, it at least does not contradict the hypothesis that there may have been a differential improvement in electricity quality in these states, via reduced blackouts.

However the results also indicate that there is a fall in electricity purchase on the extensive margin, along with an increase in firms that generate their own power. This may be because tariffs may have gone up in these states, deterring some smaller firms, that do not need high amounts of electricity. However, there may be an improvement in quality of electricity, in terms of fewer blackouts, which may explain what is observed with firms that increase electricity purchased. This may be interpreted as evidence that blackouts may be a bigger constraint for such firms, than the price.

One possible hypothesis to explain these results could be that it is administratively and politically easier to pass a tariff increase when there is one distribution company. However, with multiple distribution companies, with varying cost structures, and therefore varying break-even average prices, it may be harder for state legislators to justify different prices in different parts of the same state. However, in order to properly examine this hypothesis, comprehensive data on the quality of electricity, distribution company data, wholesale and retail price tariffs, political environments in different states, is needed. Gaining a better understanding of what caused differential effects in some states would be useful in designing future reform policy. Better reforms not only benefit the electricity sector, but also large consumers of electricity like manufacturing firms.

3.9 Additional Tables and Figures

Table 3.3: Different styles of restructuring in various states

States		
Type 1	**Type 2**	**Type 3**
Tamil Nadu	Maharashtra	Andhra Pradesh
Punjab	Meghalaya	Delhi
Himachal Pradesh	Uttaranchal	Gujarat
	Assam	Haryana
	West Bengal	Karnataka
	Chattisgarh	Madhya Pradesh
		Orissa
		Rajasthan
		Uttar Pradesh
		Bihar

Type 1 Generation and distribution separated from transmission

Type 2 Generation, transmission and distribution separated, single distribution company

Type 3 Generation, transmission and distribution separated, multiple distribution companies

Table 3.4: Electricity intensive manufacturing industries included in the data

List of included manufacturing industries
Manufacture Of Food Products And Beverages
Manufacture Of Textiles
Manufacture Of Wood And Of Products Of Wood And Cork,Except Furniture
Manufacture Of Articles Of Straw And Plating Materials
Manufacture Of Paper And Paper Products
Manufacture Of Coke, Refined Petroleum Products And Nuclear Fuel
Manufacture Of Chemicals And Chemical Products
Manufacture Of Other Non-Metallic Mineral Products
Manufacture Of Basic Metals
Manufacture Of Fabricated Metal Products, Except Machinery And Equipments
Manufacture Of Machinery And Equipment
Manufacture Of Electrical Machinery And Apparatus
Manufacture Of Other Transport Equipment
Manufacture Of Furniture
Manufacture Of Office, Accounting And Computing Machinery
Manufacture Of Radio, Television And Communication Equipment And Apparatus
Manufacture Of Medical, Precision And Optical Instruments, Watches And Clocks
Manufacture Of Motor Vehicles, Trailers And Semi-Trailers

Figure 3.10: The effect of restructuring on electricity purchase on the intensive margin, for state-year level data

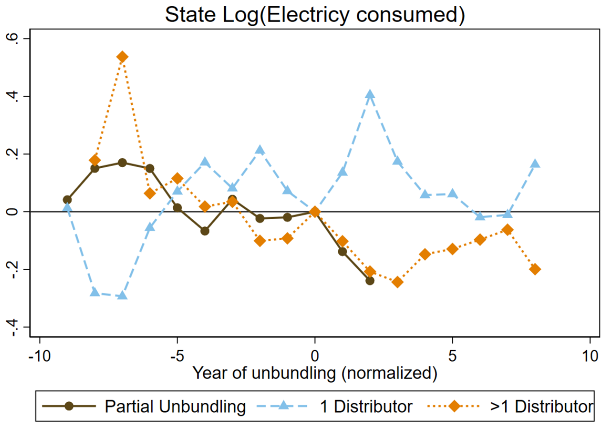
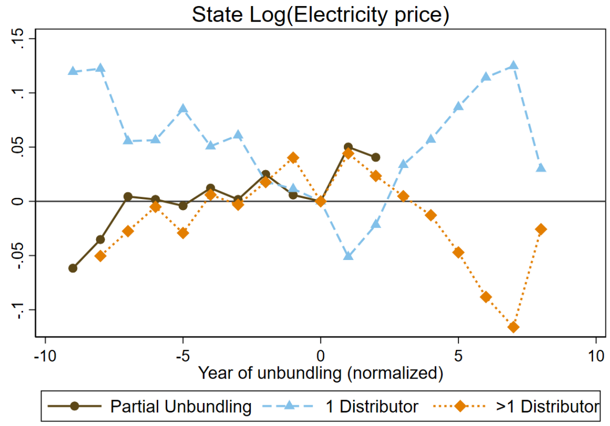


Figure 3.11: The effect of restructuring on average electricity price, for state-year level data



Note: Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The coefficients are reported from Equation (3.3)

Figure 3.12: The effect of restructuring on electricity purchase (intensive margin)

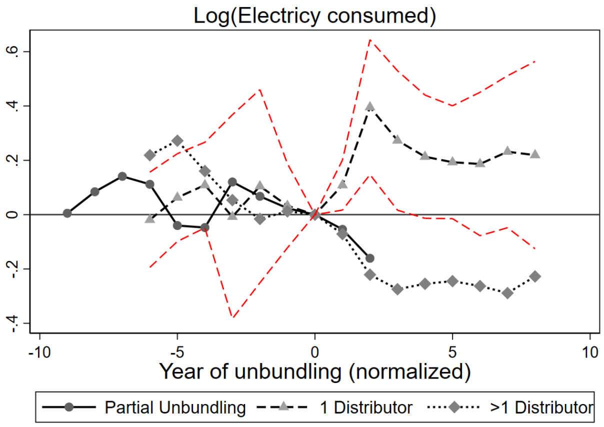
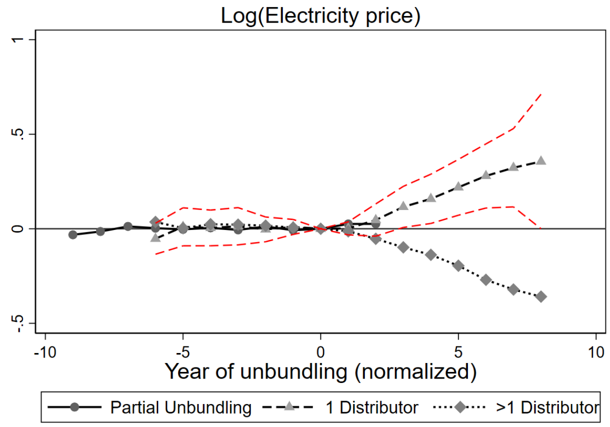
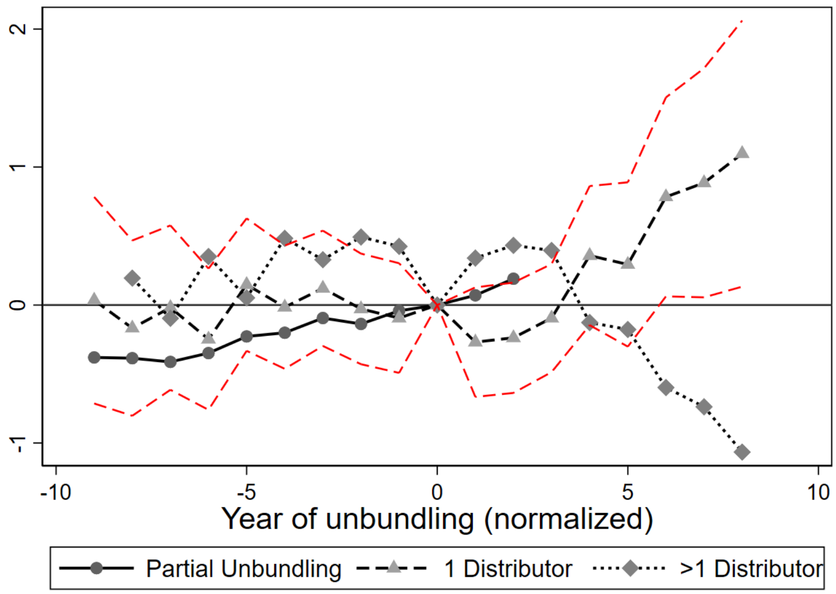


Figure 3.13: The effect of restructuring on electricity price



Note: Standard errors clustered at the state level. Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The coefficients are reported from Equation (3.2)

Figure 3.14: The effect of restructuring on luminosity (satellite data)



Note: Standard errors clustered at the state level. This graph uses satellite nighttime light density data as a proxy for electricity consumption. I use 228 observations to run a regression following Equation (3.3), and the coefficients are plotted in this graph.

Figure 3.15: The effect of restructuring on electricity purchase (extensive margin)

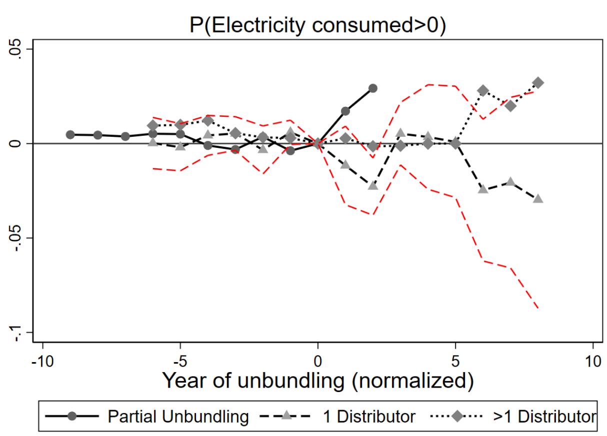
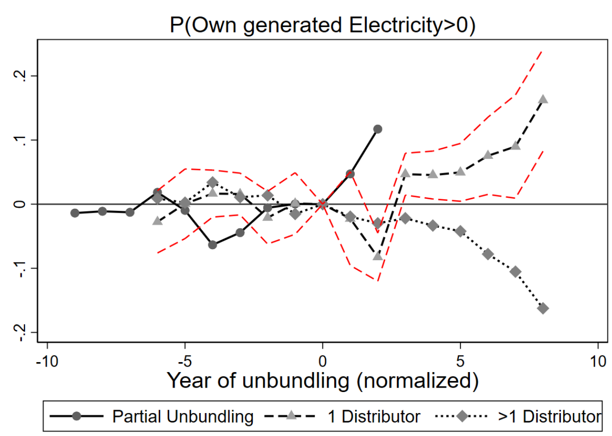


Figure 3.16: The effect of restructuring on the probability of using a generator



Note: Standard errors clustered at the state level. Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The coefficients are reported from Equation (3.2)

Figure 3.17: The effect of restructuring on non-manufacturing days

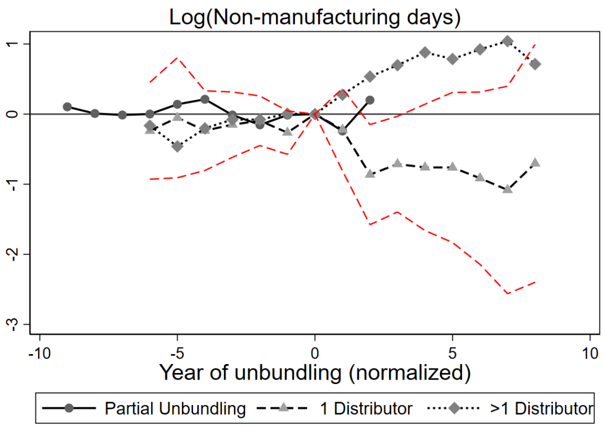
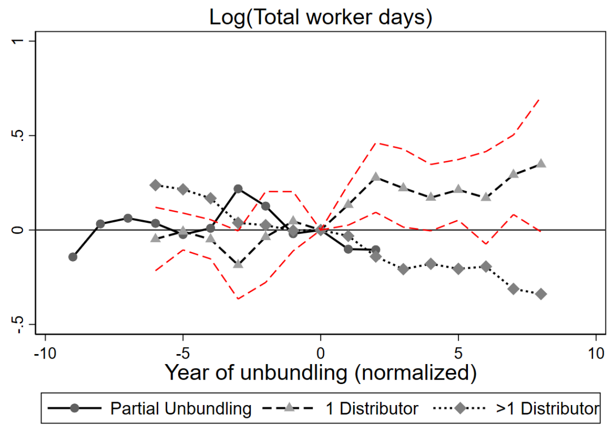


Figure 3.18: The effect of restructuring on total worker days



Note: Standard errors clustered at the state level. Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The coefficients are reported from Equation (3.2)

Figure 3.19: The effect of restructuring on total output

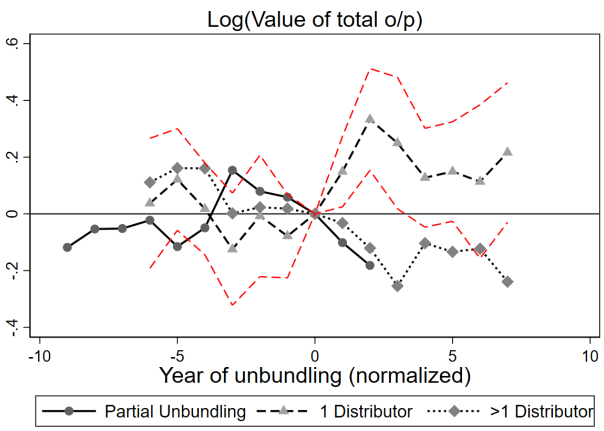
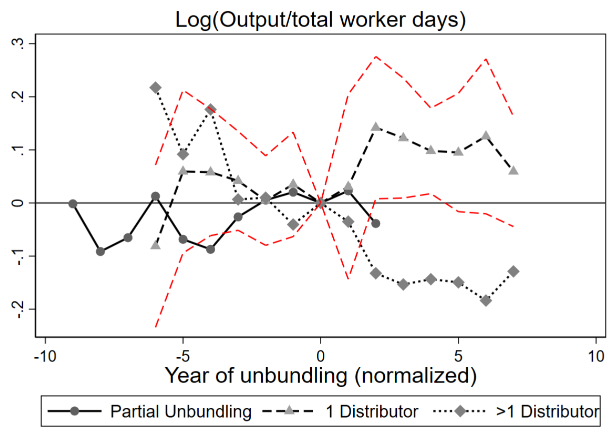


Figure 3.20: The effect of restructuring on total output productivity



Note: Standard errors clustered at the state level. Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The coefficients are reported from Equation (3.2)

Figure 3.21: The effect of restructuring on electricity purchase (intensive margin) - Model with 0 pre-trends

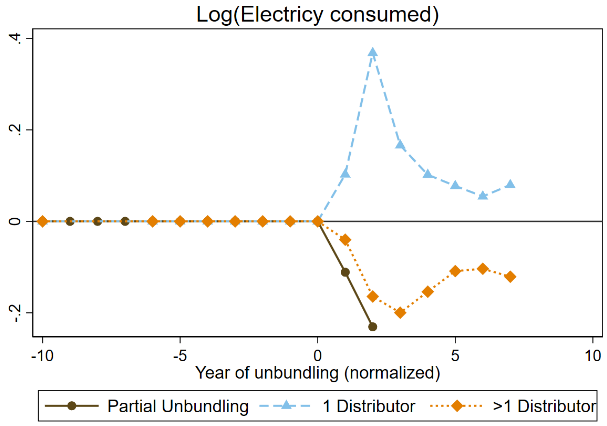
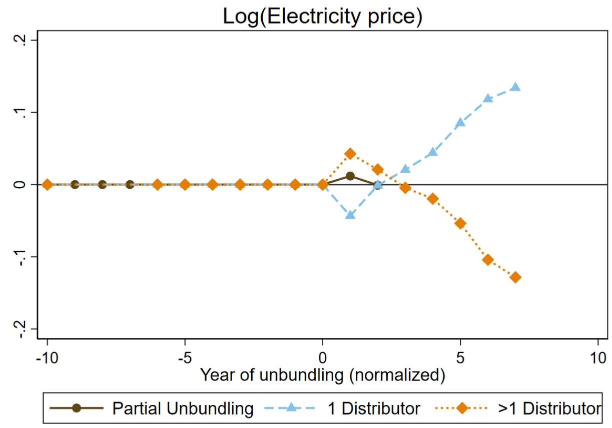


Figure 3.22: The effect of restructuring on electricity price - Model with 0 pre-trends



Note: Dataset restricted to privately-owned, electricity-intensive manufacturing firms, full-states with more than 1000 observations.The coefficients are reported from Equation (3.4)

BIBLIOGRAPHY

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.

Alam, M. (2015). Coping with Blackouts: Power Outages and Firm Choices. *Yale University (mimeo)*.

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.

Andersen, S., Harrison, G., Lau, M., and Rutstrom, E. (2008). Eliciting Risk and Time Preferences. *Econometrica*, 76(3):583–618.

Andreoni, J. and Sprenger, C. (2012). Estimating Time Preferences from Convex Budgets. *The American Economic Review*, 102(7):3333–3356.

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.

Apt, J. (2005). Competition Has Not Lowered U.S. Industrial Electricity Prices. *The Electricity Journal*, 18(2):52–61.

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.

Bacon, R. W. and John, B.-J. (2002). Global Electric Power Reform, Privatization and Liberalization of the Electric Power Industry in Developing Countries. *Energy and Mining Sector Board Discussion Paper Series*, (2).

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.

Borenstein, S., Bushnell, J. B., and Wolak, F. A. (2002). Measuring Market Inefficiencies in Californias Restructured Wholesale Electricity Market. *American Economic Review*, 92(5):1376–1405.

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. Working Paper.

Bushnell, J. B. and Wolfram, C. (2005). Ownership change, incentives and plant efficiency: The divestiture of u.s. electric generation plants. *Center for the Study of Energy Markets, Working Paper 140*.

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.

Camerer, C. and Weber, M. (1992). Recent Developments in Modeling Preferences: Uncertainty and Ambiguity. *Journal of Risk and Uncertainty*, 5(4):325–370.

Carlsson, F., Johansson-Stenman, O., and Nam, P. K. (2014). Social Preferences are Stable Over Long Periods of Time. *Journal of Public Economics*, 117(C):104–114.

Carvalho, L. S., Meier, S., and Wang, S. W. (2016). Poverty and Economic Decision-making: Evidence from Changes in Financial Resources at Payday. *The American Economic Review*, 106(2):260–284.

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

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior and Organization*, 87:43–51.

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.

Cho, I., Orazem, P. F., and Rosenblat, T. (2018). Are Risk Attitudes Fixed Factors or Fleeting Feelings? *Journal of Labor Research*, 39:1–23.

Chuang, Y. and Schechter, L. (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, 117:151–170.

Cicala, S. (2015). When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation. *American Economic Review*, 105(1):411–444.

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

Cropper, M., Malik, K., Limonov, A., and Singh, A. (2011). Estimating the Impact of Restructuring on Electricity Generation Efficiency: The Case of the Indian Thermal Power Sector. *NBER Working Papers 17383*.

Dasgupta, U., Gangadharan, L., Maitra, P., and Mani, S. (2017). Searching for Preference Stability in a State Dependent World. *Journal of Economic Psychology*, 62(C):17–32.

Davis, L. W. and Wolfram, C. (2012). Deregulation, Consolidation, and Efficiency: Evidence from US Nuclear Power. *American Economic Journal: Applied Economics*, 4(4):194–225.

Dean, M. and Sautmann, A. (2016). Credit Constraints and the Measurement of Time Preferences.

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.

Eckel, C. C. and Grossman, P. J. (2002). Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk. *Evolution and Human Behavior*, 23(4):281–295.

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.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, pages 643–669.

Erez, A. and Isen, A. M. (2003). The Influence of Positive Affect on the Components of Expectancy Motivation. *Journal of Applied Psychology*, 87(6):1055–67.

Fabrizio, K. R., Rose, N. L., and Wolfram, C. D. (2007). Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency. *American Economic Review*, 97(4):1250–1277.

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.

Forgas, J. P. (1995). Mood and Judgment: The Affect Infusion Model (AIM). *Psychological Bulletin*, 117(1):39.

Forum of Regulators (2013). Introducing Competition in Retail Electricity Supply in India. *Government of India*.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *The Journal of Economic Perspectives*, 19(4):25–42.

Fudenberg, D. and Levine, D. K. (2006). A Dual-self Model of Impulse Control. *American Economic Review*, 96(5):1449–1476.

Fudenberg, D. and Levine, D. K. (2011). Risk, Delay, and Convex Self-control Costs. *American Economic Journal: Microeconomics*, 3(3):34–68.

Fudenberg, D., Levine, D. K., and Maniadis, Z. (2014). An Approximate Dual-self Model and Paradoxes of Choice under Risk. *Journal of Economic Psychology*, 41:55–67.

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.

Harrison, G. W. (2005). *Field Experiments in Economics (research in Experimental Economics, Volume 10)*, chapter Field Experiments and Control, pages 17–50. Emerald Group Publishing Limited.

Haushofer, J., Schunk, D., and Fehr, E. (2013). Negative Income Shocks Increase Discount Rates. Technical report, University of Zurich Working Paper.

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

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies*. Oxford University Press, New York.

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

Ifcher, J. and Zarghamee, H. (2011). Happiness and Time Preference: The Effect of Positive Affect in a Random-assignment Experiment. *American Economic Review*, 101(7):3109–3129.

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.

International Energy Agency (2015). *India Energy Outlook*. International Energy Agency.

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

Joskow, P. L. (1997). Restructuring, Competition and Regulatory Reform in the U.S. Electricity Sector. *The Journal of Economic Perspectives*, 11(3):119–138.

Joskow, P. L. and Kahn, E. (2002). A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000. *The Energy Journal*, 23(4):1–36.

Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). Fairness As a Constraint on Profit Seeking: Entitlements in the Market. *The American Economic Review*, 76(4):728–741.

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.

Kirchsteiger, G., Rigotti, L., and Rustichini, A. (2006). Your Morals Might Be Your Moods. *Journal of Economic Behavior and Organization*, 59:155–172.

Knittel, C. (2002). Alternative Regulatory Methods and Firm Efficiency: Stochastic Frontier Evidence the US Electricity Industry. *The Review of Economics and Statistics*, 84(3):530–540.

Lawrance, E. C. (1991). Poverty and the Rate of Time Preference: Evidence from Panel Data. *Journal of Political Economy*, 99(1):54–77.

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.

List, J. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political Economy*, 115:482–493.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty Impedes Cognitive Function. *Science*, 341(6149):976–980.

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

Mansur, E. T. (2007). Upstream Competition and Vertical Integration in Electricity Markets. *Journal of Law and Economics*, 50:125–156.

Markiewicz, K., Rose, N., and Wolfram, C. (2004). Has Restructuring Improved Operating Efficiency at Us Electricity Generating Plants. *Center for the Study of Energy Markets, Working Paper 135*.

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.

Mullainathan, S. and Shafir, E. (2013). *Scarcity: Why Having Too Little Means so Much*. Macmillan.

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

Necker, S. and Ziegelmeyer, M. (2016). Household Risk Taking After the Financial Crisis. *The Quarterly Review of Economics and Finance*, 59:141–160.

Oswald, A. J., Proto, E., and Sgroi, D. (2015). Happiness and Productivity. *Journal of Labor Economics*, 33(4):789–822.

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

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

Roszkowski, M. and Cordell, D. (2009). A longitudinal perspective on financial risk tolerance: Rank-order and mean level stability. *International Journal of Behavioural Accounting and Finance*, 1:111–134.

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. Working Paper.

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.

Schildberg-Hörisch, H. (2018). Are Risk Preferences Stable? *Journal of Economic Perspectives*, 32(2):135–54.

Sen, A. and Jamasb, T. (2012). Diversity in Unity: An Empirical Analysis of Electricity Deregulation in Indian States. *The Energy Journal*, 33(1):83–130.

Shah, A. K., Mullainathan, S., and Shafir, E. (2012). Some Consequences of Having Too Little. *Science*, 338(6107):682–685.

Shapiro, M. D. and Slemrod, J. (1995). Consumer Response to the Timing of Income: Evidence from a Change in Tax Withholding. *The American Economic Review*, 85(1):274–283.

Sinayev, A. and Peters, E. (2015). Cognitive Reflection Vs. Calculation in Decision Making. *Frontiers in psychology*, 6:532.

Singh, A. (2006). Power Sector Reform in India: Current Issues and Prospects. *Energy Policy*, 34(16):2480–90.

Spears, D. (2011). Economic Decision-making in Poverty Depletes Behavioral Control. *The B. E. Journal of Economic Analysis and Policy*, 11(1).

Stephens, M. (2003). ”3rd of The Month”: Do Social Security Recipients Smooth Consumption between Checks? *The American Economic Review*, 93(1):406–422.

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

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam. *The American Economic Review*, 100(1):557–571.

Tanaka, Y., Fujino, J., Ideno, T., Okubo, S., Takemura, K., Miyata, J., Kawada, R., Fujimoto, S., Kubota, M., Sasamoto, A., et al. (2014). Are Ambiguity Aversion and Ambiguity Intolerance Identical? A Neuroeconomics Investigation. *Frontiers in Psychology*, 5.

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).

Tongia, R. (2003). The Political Economy of Indian Power Sector Reforms. *Program on Energy and Sustainable Development Working Paper 4*.

Wolfram, C. (1999). Measuring Duopoly Power in the British Electricity Spot Market. *American Economic Review*, 89(4):805–826.