Essays in Development and Labor Economics in Brazil

Alexander Poulsen

A dissertation

submitted to the Faculty of

the department of Economics

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College

Morrissey College of Arts and Sciences

Graduate School

April 2021

Essays in Development and Labor Economics in Brazil

Alexander Poulsen

Dissertation Committee:

Arthur Lewbel (Chair)

Anant Nyshadham

Theodore Papageorgiou

Abstract

In my dissertation I studied the political economy of economic development as well as the the urban wage premium, all using data from Brazil. In the first chapter, which is joint work with Carlos Varjão, we analyze the effect that increased political opposition in the city council has on corruption and public service provision at the local level. In the second chapter I study the sources of the high wage premium observed in cities, including firm sorting, firm and occupational matching, and compensating wage differentials. Finally, in the third chapter I study what happens to the provision of public education when a school teacher is elected to the city council (which actually occurs quite frequently). More detailed summaries of each chapter follow below.

Chapter 1: In "*Political Opposition, Legislative Oversight, and the Performance of the Executive Branch*", we study the effect that increased political opposition has on corruption and other measures of the mayor's performance in Brazil. The separation of powers between the executive and legislative branches is a cornerstone of democracy. This system of checks and balances, however, can be circumvented by partisan loyalties if legislators strategically avoid exerting oversight when their own party controls the executive branch. It is thus an empirical question whether the separation of powers prevents the abuse of

power in practice. We answer this question by measuring the extent to which members of political opposition parties in a city council effectively check the mayor's performance in Brazil. We employ a regression discontinuity design to estimate the causal effect of an additional politically opposed legislator, and we find that political opposition increases oversight action and decreases corruption, with the effect fully concentrated on mayors facing reelection pressure. We trace the impact of oversight, via a reduction in healthcare spending irregularities, all the way to impacts on healthcare service delivery and health outcomes.

Chapter 2: In "*Decomposing the Urban Wage Premium in Brazil: Firms, Matching, and Compensating Wage Differentials*" I study the sources of the high wage premium observed in cities. In this chapter I used detailed employer-employee matched data from Brazil to understand 3 important elements of the urban wage premium: (1) the role of firms sorting into cities, (2) the role of firm and occupational matching in creating agglomeration economies, and (3) the role of compensating wage differentials. I first exploit identification from multi-city firms to show that positive selection of high-wage firms into larger cities accounts for 44% of what is often considered 'agglomeration economies'. Then I show that improved firm and occupational matching together account for 87% of agglomeration effects. I then turn my attention to compensating wage differentials— a possible explanation for the high-wage firms in cities. I estimate revealed-preference valuation of jobs, and show that jobs in cities in fact have better non-wage characteristics, and so high urban wages cannot be due to compensating wage differentials. This evidence together suggests that in Brazil, cities exist because they provide thick labor markets where high-wage firms and high-wage workers can go to find productive matches.

Chapter 3: In "*Teachers in Politics: Teacher-Politicians, Gender, and the Representation of Public Education*" I study what happens to public education

in a city when a school teacher is elected to the city council, and I find that it depends on the gender of the teacher. Using a regression discontinuity design that exploits close elections, I find that when a female teacher is elected to the city council, the city hires both more teachers and more *qualified* teachers, and pays them more. Having a female teacher on the city council also increases the likelihood that the city's schools have necessary teaching resources, books, and financing, and possibly increases student test scores. No significant effect is found for male teachers elected to the city council. This difference may be due to different political career concerns for men versus women, a simple amplification of existing gender policy preference differences, or some mixture of the two.

Acknowledgements

This work would not have been possible (or at least would have been far more difficult) without the help of many, many people. First I must thank my parents, Eric and Jill, for nurturing my curiosity about the world from a young age, and supporting me in many ways throughout my formal education (which has been most of my life at this point). I thank my professors during undergrad at BYU who mentored me, gave me opportunities to start working on research, and helped me navigate the maze of PhD program admissions, specifically Joe Price, Rick Evans, and Brigham Frandsen.

I'm grateful to the Boston College economics department for giving me an academic home for the past 6 years, teaching me, and allowing me opportunities to share knowledge with others as well. I want to thank my advisors Anant Nyshadham, Arthur Lewbel, and Theo Papageorgiou, for numberless hours of discussions about the content of this dissertation. Without their guidance, there is no way I would have been able to produce this research.

I also want to thank all of my friends in Brazil, who make the numbers and figures in this dissertation come to life. It is due to my experiences in Brazil that I had the drive and curiousity to pursue this research. Without my experience in Brazil, visiting people of all stations of life every day for two years, I would not have had the motivation nor the expertise to complete the work herein. If this work is dedicated to anyone, it is dedicated to them.

I want to thank all of my friends in Boston who have made my non-schoolwork time so meaningful and enjoyable, and have helped Boston feel like home. Lastly I want to thank my wife Katie for her love and support— she knows more than anyone else the ups and downs that this research involved.

Contents

Chapter 1

Political Opposition, Legislative Oversight, and the Performance of the Executive Branch

1.1 Introduction

The separation of powers between government branches is a central tenet of liberal democracy. While the executive branch implements policies, the legislative branch has oversight responsibilities that can deter the abuse of power. But these institutional checks can be circumvented if partisan legislators strategically avoid exerting oversight when their own party controls the executive branch.[1] Hence, the effectiveness of separation of powers in practice depends critically on the degree of partisan competition, yet this has received little attention in the literature. While theory gives some guidance on this issue (Person et al, 1997, 2000; and Acemoglu et al, 2013; for example), there is little empirical work evaluating how well legislative oversight prevents rent seeking in practice.

In this paper, we provide direct evidence on the role that partisan competition has on curtailing corruption under the separation of powers. Particularly, we measure the extent to which political opposition parties in a city council provide a binding check on a mayor's performance and thus enhance the quality of public services in Brazil.

Intuitively, legislators' incentives to inform voters about a mayor's wrongdoings should depend on their party affiliation, since electoral prospects of mayors

[1] In fact, legal scholars Levinson and Pildes (2006) make the case that the true separation of powers is the separation of power between political parties, not branches of government.

and legislators from the same party are usually closely aligned. Hence, allied legislators have little incentive to investigate and report a mayor's corrupt activities since this can hurt their own reelection prospects. On the other hand, opposition legislators could perhaps improve their reelection chances by exposing a corrupt mayor. Moreover, this check on corruption should mainly be effective when the incumbent mayor is not term limited and cares about the impact of being exposed on his reelection chances. We formalize this intuition by adapting a canonical political agency model (Besley, 2006) to accommodate legislators' partisan incentives to provide information to voters. The model predicts that when mayors with reelection incentives face strong legislative opposition, there should be more oversight investigations and less corruption.

To test these predictions, we use a regression discontinuity (RD) design to estimate the causal impact of an additional political opponent in the city council on mayoral performance.[2] We follow a body of literature that exploits close elections as an RD design (Lee, 2008; Pettersson-Lidbom, 2008; Caughey and Sekhon, 2011; Eggers et al., 2015), so that the estimator compares municipalities where the opposition party just barely won an additional seat in the legislature to those where the opposition (by a small margin) did *not* win an additional seat. We trace the effects of partisan opposition through four links in the underlying causal chain. We find that legislative opposition: (*i*) increases legislative oversight activities, (*ii*) decreases corrupt activities by the mayor (particularly in the health sector), (*iii*) improves the quality of public health service provision, and (*iv*) enhances health outcomes.

More specifically, we first substantiate a mechanism through which opposition legislators constrain the executive. The creation of a Parliamentary Commission of Inquiry (CPI) is a major oversight instrument that legislative bodies in Brazil have, and the president of the city council - which is elected by the

[2]As these councils are small (typically 9 members) one additional opposition legislator represents a substantial increase in the opposition seat share and frequently flips the legislative majority towards opposition control.

city council itself - plays an essential role for the creation and effectiveness of CPIs. The city council president decides whether a CPI has a valid reason to be created, and if so, appoints legislators to the commission. We show that an additional opposition party legislator increases the probability that an opposition legislator is city council president by 14 p.p and increases the likelihood that an investigative commission is created to look into mayors' wrongdoings by 6 p.p. (55% increase).

Second, we assess if this increase in oversight inhibits rent-seeking behavior. We find that an additional opposition party legislator reduces the instances of corruption found by auditors by 0.846 standard deviation units, with the impact fully concentrated on mayors with reelection incentives. We also find that the benefits of reduced corruption are concentrated in reducing irregularities in healthcare spending. This makes sense given that municipalities have a great deal of control over the quality of local healthcare provision and the fact that healthcare is one of the most salient issues for voters and politicians in Brazil (Fujiwara, 2015).

Third, we look at whether this reduction in embezzlement of health funds improves public service delivery, which would happen if decrease in embezzlement increases the likelihood that government resources reach their intended health programs. Our estimates suggest that this is indeed the case; an additional opposition legislator increases the probability that a physician will be present at the local public health clinic, decreases waiting lines, and reduces irregular hiring practices of health workers.

Fourth, we find evidence that this improvement in public health service provision translates into better health outcomes. Results suggest that for uneducated mothers (those most likely to depend on public health services), one additional opposition legislator decreases infant mortality by 3.4 per 1000 births and the rate of preterm births by 0.8 per 1000 births. This impact of legislative

oversight on health outcomes is large and robust to various specifications and falsification tests.

Regarding mechanisms, we test for heterogeneous effects based on which seat is captured by the extra opposition legislator. We find evidence that the impact of an additional opposition legislator is driven by situations when it flips the majority in the local city council, although results of a seat by seat analysis are imprecise due to small sample sizes. This result suggests that to promote effective oversight, the opposition needs to control key institutional features— like launching an investigative commission— that are only available once they have a majority in the legislative branch.

We also test two of the main mechanisms in the model. First, we provide evidence of partisan differentiated incentives for oversight by showing that the electoral prospects of mayors and legislators from the same party— but not the opposition— are closely aligned. Using the fact that the timing of the public dissemination of the federal audit results is random (Ferraz and Finan, 2008), we show that when a mayor is revealed to be corrupt and there is an AM radio station present to disseminate findings, electoral success of mayor-allied legislators falls by 10 p.p., but there is no effect for opposition legislators. Additionally, we evaluate whether the presence of opposition legislators reduces the reelection chances of incumbent mayors. Opposing legislators reduce reelection chances of mayors when reelection incentives are not strong enough to force mayors to pretend to be non-corrupt, even when they know that they will be reported by legislators and this will end up costing their reelection. Results suggest that an additional opposition legislator decreases mayors' reelection chances by 6.2 p.p.

The main contribution of this paper is to show that the effectiveness of the separation of powers in preventing abuses of power depends on the degree of partisan competition. We do so by exploiting exogenous variation in the number

of politically opposed legislators, and showing its effect on a chain of outcomes ranging from politician behavior to the welfare of constituents. This relates to the literature on political institutions and politician performance. While this literature has principally considered the direct effect of electoral incentives on performance (Cascio and Washington, 2013; Dal Bó and Rossi, 2011; Ferraz and Finan, 2011; Lim, 2013; Martinez-Bravo et al. 2014; Fujiwara, 2015; Gulzar and Pasquale, 2017), a much smaller strand of this literature has examined the effect of checks from other branches of government (Alt and Lassen, 2008; Litschig and Zamboni, 2015).[3] So while previous literature shows that voters keep politicians in line, we show that the system of checks and balances allows politicians to keep other politicians in line. [4]

Our results further relate to the literature on divided government. Alesina and Rosenthal (1996) build a model where the election of a divided government arises from moderate voters preferring moderate policies. This paper proposes another rationale for the formation of divided government: to incentivize good performance from the executive. This could be part of the explanation for the high degree of split-ticket voting in Latin America (Ames et al. 2009), a region with high corruption and poor public services. While much attention has also been given to the potential of divided government to generate legislative gridlock (Fiorina, 1992; Krehbiel, 1998; Mayhew, 1991; Nzelibe and Stephenson, 2009), there is little investigation on the potential of divided government to reduce rent extraction by the executive.

There is also a large recent literature that studies the impact of centralized audits on corruption and public service provision (Olken, 2007; Ferraz and Finan, 2008; Litschig and Zamboni, 2016; Nishijima, Ellis, and Cati, 2016, Lichand et al., 2016; Avis, Ferraz, and Finan, 2018). Our work can be seen

[3]Related to these checks from other branches of government, Britto and Fiorin (2016) show that larger city council sizes in Brazil increase corruption.

[4]We also contribute to a related literature on democracy and welfare (Besley and Kudamatsu, 2006; Kudamatsu, 2012; Acemoglu et al. 2014; Fujiwara 2015; Madsen et al. 2015) by documenting how a specific democratic feature can improve citizens' well-being.

as complementary to this literature, since we study the effect of the auditing unit that is already built in to many democracies— the legislative branch. Our analysis suggests that strengthening legislative oversight, by ensuring that legislative branches have both the authority and capacity to investigate the executive branch, could be an effective way to improve politician performance. This article is also a contribution to the literature on political partisanship and corruption (Anduiza, Gallego, and Muñoz, 2013; Eggers, 2014), which finds that voters are more tolerant of corruption in their own political party. This paper shows that this result also holds for politicians, and that this tolerance has negative impacts on outcomes such as health.

1.2 Institutional Background

Brazil is a federal republic much like the United States. There are 3 spheres of government: the federal government, the states, and the municipalities. Executive and legislative branches exist in all three spheres and are directly elected.

Municipal governments in Brazil are made up of of the mayor, his or her appointed secretaries, and the city council (*Câmara de Vereadores*). As the executive branch, the mayor and secretaries are responsible for implementing laws and policies through the Ministries of Health, Education, Agriculture, and so on. Mayors face a two-term limit, and thus do not face reelection incentives during their second term. As the legislative branch, the city council is given the responsibility to (1) make laws, and (2) audit and review municipal spending, which includes reviewing the accounts of the mayor and his or her secretaries.

Internally, the city council elects a board of directors— a president, vice-president, and secretaries— which serves a purpose similar to the speaker of the house in the U.S. House of Representatives. Among other things, this board is responsible for proposing projects and authorizing procurement of public goods. The city council's twofold responsibility will require a corrupt mayor

to either buy off the city council, or find some other way to get around them. If monitoring responsibilities are divided up among city councillors, more city councillors from opposing parties may make it more difficult to steal money or exert low effort. Corruption frequently manifests itself in the form of fraudelent projects, which must be approved by the city council. More opposing politicians in the legislature could mean more power to block these projects from being approved. Thus, the legislature could improve executive performance and/or restrain the executive's rent-seeking either through (1) legislation it does or does not choose to pass or (2) through its auditing responsibilities.

If city councilors notice some irregularities in the mayor's accounts, they are charged with creating a Parliamentary Commission of Inquiry (CPI) in order to investigate possible malfeasance. Legislators may threaten the executive with denunciation, but if they are politically opposed, these threats will be much more credible both because political leverage stands to be gained and because politicians may not want their political enemies to have access to additional resources gained from rents.

The creation of CPIs is a right of legislative bodies at municipal, state and federal level guaranteed by the Brazilian constitution[5]. CPIs need the approval of 1/3 of the legislative body to be initiated. A CPI must be created with a very specific purpose and a time limit (usually 6 months). The CPI has judicial investigative powers. It can have access to bank, telephone and tax information and can call witnesses to be interrogated. The CPI ends with a document describing all the findings of the investigation and is passed forward to prosecutors to take the appropriate actions. The city council can also use the wrongdoings uncovered by the investigation to impeach the mayor.

The leader of the city council plays an important role in the effectiveness of the CPI. Local city council leaders are elected by the city council itself, are responsible for deciding if the purpose of the investigation is narrow enough

[5]Art. 58 Paragraph 3

for the commission to be created, and are also responsible for deciding which legislators that will work on the CPI[6]. Once a CPI is created, it elects a president who is responsible to guide the investigations. Hence, it is common practice for council leaders— if allied with the mayor— to claim that the CPI does not have a well defined purpose to block it or create a CPI with legislators that have no interest in investigating the mayor.

Some examples are useful to illustrate how allies of the mayor can block the creation or effective investigation by CPIs. In 2015, a councilman in the municipality of Nova Friburgo in the state of Rio de Janeiro tried 2 times to create a CPI to investigate irregularities in the procurement process for medicine acquisition by the municipal government. Although he had enough votes to create the CPI, the leader of the city council (who is from the same party as the mayor) did not aprove the CPI because he claimed that the purpose of the investigation was not narrow enough[7]. Two months later, the leader of the city council approved the creation of the CPI due to popular pressure.

Another example comes from the city of Rio de Janeiro. Opposition legislators wanted to create a CPI to investigate several reports of schemes to defraud procurement contracts on municipal public works for the 2016 Olympics. Again, although the opposition had enough votes to create the CPI, the city council leader (who is from the same party as the mayor) claimed that the purpose of the investigation was too broad and did not approve the creation of the CPI. The opposition went to the courts and a judge decided that the purpose of the investigation was very well defined and ordered the city council president to accept the creation of the CPI[8]. After the CPI was created, the city council leader decided that 4 of the 5 members of the commission would be from the mayor's party. The opposition did not accept this, claiming that with this

[6]In principle, he should respect the party composition of the house, but this is commonly ignored as the examples below show.

[7]http://avozdaserra.com.br/noticias/cpi-da-saude-aprovada-em-nova-friburgo

[8]http://esportes.estadao.com.br/noticias/jogos-olimpicos,justica-do-rio-obriga-camara-municipal-a-instalar-cpi-da-olimpiada,1855211

composition the investigation would be compromised. The investigations have been suspended since then[9].

As the examples above illustrate, it can be extremely hard for the opposition to exert effective oversight of the executive actions if it does not have a substantial presence in the city council.

On the other hand, there is anecdotal evidence that suggests that corruption in Brazil may be institutionalized and that party doesn't matter; in other words, corruption could be largely due to 'cultural norms'. As one Brazilian anticorruption organization has said, 'It seems there is some unwritten pact, a type of code of honor among the corrupt... and they meet their terms, even when they are political enemies' (Chizzotti et al. 2012). Corruption schemes may be inhereted from previous administrations, despite the transfer of power from one party to another, with city councillors recieving monthly payments from the mayor to keep quiet.

Politics is highly fragmented in Brazil due to a proportional represention system— in our data 25 different parties had mayors elected in the municipalities of Brazil. This fragmentation makes pure majorities in the legislature almost impossible to get, and so coalitions are key to getting representatives elected and advancing a policy agenda. Four main parties dominate the political landscape in Brazil, with smaller parties generally allying themselves with larger parties according to current political issues. Parties form coalitions for the election of both mayors and legislators, with candidates running under a specific party and coalition. These political coalitions are seen by some as exacerbating the problem of corruption. Coalitions between parties are often formed based on promises to be fulfilled after the election, which could lead to fraudelent schemes in order to transfer money to party leaders as a reward. In 2015 legislation was introduced in the Brazilian National Congress to prohibit

[9]http://g1.globo.com/rio-de-janeiro/noticia/2016/09/justica-do-rio-determina-reabertura-da-cpi-da-olimpiada.html

coalitions in proportional elections, but the legislation did not pass.

Brazil has an open party-list proportional system where seats are allocated according to the D'Hondt Method[10], which will be discussed later as an important part of our identification strategy. The D'Hondt Method is widely used—44 countries[11] in the world use some form of it, principally in Europe and Latin America. The findings in this paper will have important implications for the merits of the D'Hondt Method versus alternative methods like the Webster Method[12] or the Huntington-Hill Method[13].

Brazilian municipalitites are an ideal setting to study the effect of political opposition on politician performance because we can observe a cross-section of thousands similar local-level governments, and because governance is highly decentralized in Brazil, meaning that local-level politics can have a large impact on important outcomes.

1.3 Data

1.3.1 Audit Data

One of the most difficult parts of studying corruption or politician effort is getting accurate and informative data. In this paper we use rich data provided by an anticorruption program in Brazil. Beginning in the year 2003, the Brazilian Federal Government began a lottery program in which every few months, municipalities from across the country would be randomly chosen to

[10]First introduced by Thomas Jefferson to allocate seats to states in the US House of Representatives, but it is most often associated with the Belgian mathematician Victor D'Hondt who introduced it a century later

[11]According to Wikipedia, these are Albania, Argentina, Armenia, Austria, Belgium, Brazil, Bulgaria, Cambodia, Cape Verde, Chile, Colombia, Croatia, the Czech Republic, Denmark, the Dominican Republic, East Timor, Ecuador, Estonia, Fiji, Finland, Guatemala, Hungary, Iceland, Israel, Japan, Kosovo, Luxembourg, Macedonia, Moldova, Montenegro, Netherlands, Northern Ireland, Paraguay, Peru, Poland, Portugal, Romania, Scotland, Serbia, Slovenia, Spain, Turkey, Uruguay, and Wales.

[12]Introduced by American statesman Daniel Webster. This method was formerly used to allocate seats to the states in the US House of Representatives.

[13]This is the method currently in use to allocate seats to states in the US House of Representatives; however, it is not currently used by any legislature to allocate seats to parties.

be audited with respect to all federal funds they had received in recent years. When a city is chosen by the lottery, a team of federal employees spends a few weeks in the city, inspecting receipts, budgets, bank statements, as well as the physical premises of projects targeted by federal monies, to ensure that projects have taken place. Programs audited include primary schooling, health clinics, poverty relief, and road construction. Audits are administered by the Controller-General of the Union (CGU), an agency within the System of Internal Control of the federal government's executive branch. The federal auditors are highly paid and trained professionals, and Ferraz and Finan (2008) find no evidence that they are successfully bribed by municipal governments to manipulate audit reports.

After being in the city for a few weeks reviewing documents, federal auditors write up a report (usually between 50 and 200 pages) listing all 'irregularities' pertaining to each program and service item[14] within the program, including how much money was involved in each project and service item. About 80% of funds audited are related to either health, education or social programs. We use data from audits performed during the 3 mayoral terms from 2005-2016.[15] Our data comes requested from the CGU, and lists all irregularities found in audits from the 20th to the 40th lottery.

Following classifications in Ferraz and Finan (2011), we examine 3 main types of corruption found in the reports: (1) embezzlement (diversion of funds), (2) fraud in procurement, and (3) overinvoicing. In the reports, what these respectively look like is (1) the local government spends public money and doesn't prove where it went (presumably transferred to private bank accounts, etc.); (2) the bidding process for public projects is simulated or manipulated, either using fake firms or 'friendly' firms; or (3) government officials pay higher-than-

[14] A 'service item' indicates a given set of transfers from the federal government to the municipal government for a given program. This term has been used interchangebly with the term 'service order' in the literature that uses these audit reports.

[15] Terms are 4 years long

market prices for goods or projects (and then presumably receive a kickback from the private providers). As outcome variables we use each of these 3 types of corruption, an aggregate corruption variable, as well an aggregate of all other irregularities not classified as embezzlement, fraud, or overinvoicing. This includes serious procedural errors and other irregularities.

In our data, we observe all irregularities found by auditors over the 1,200 audits performed from the years 2006 to 2015.[16] In order to create measures for corruption, we use regular expressions (i.e. keywords search) to search for words and phrases that isolate an irregularity as corruption. Details on these regular expressions can be found in the appendix in section 1.11. Given the regular expressions we use, it is clear that our fraud and overinvoicing variables are measuring corruption at some level of the government (whether the mayor directly or his or her subordinates). Our embezzlement variable measures situations in which money was spent by the municipal government and there are no receipts or documents showing that the money was actually spent as specified, thus opening the way for diverting public resources for private purposes. While some of these could be mere procedural errors, we present evidence in the results section that they are generally not. Our measure of corruption is very closed related to Ferraz and Finan (2011). The main difference is that instead of reading the reports and manually classifying corruption episodes like Ferraz and Finan (2011), we use this regular expression method to classify corruption episodes.[17]

Audit reports are subdivided by ministry (of health, education, etc.) and then by service item, and the CGU classifies each irregularity as either 'major' or 'minor' based on potential monetary losses to the government. Our main

[16]We omit the data from lotteries 29 and 30 because these audits happened in the second half of the first year of a term, making it unclear if corruption findings are due to the current or previous administration. Corruption found in audits performed in the first 6 months of a term are attributed to the previous administration.

[17]So while we did not read all the reports, we did read a large sample of the reports in order to know which expressions and phrases regularly indicated that corruption had taken place.

corruption outcome variables will be the amount of service items found to have been involved in each type of corruption that were defined as 'major' irregularities by the CGU. The variable 'Total Corruption' is the sum of irregularities associated with embezzlement, fraud, or overinvoicing, for each municipality. Summary statistics are provided in table 1.1.

Table 1.1: Audit Data Summary Statistics

	count	mean	sd	min	max
Total Audit Service Items	788	24.01269	9.552229	7	95
Total Corruption	788	.8109137	1.598453	0	13
Embezzlement	788	.2525381	.7634811	0	8
Fraud	788	.2918782	.8910597	0	8
Overinvoicing	788	.3274112	.9977917	0	12
Other Irregularities	788	3.332487	3.435536	0	30

Each variable is the count of audit service items where the given type of irregularity was found and also classified as a 'major' irregularity by the CGU.

The unit of observation is an audited municipality and the sample is audited municipalities between 2006 and 2015 that fit the requirements for the RD design (detailed in section 3.4.1).

1.3.2 Public Service Provision Data

When the CGU sent auditors to inspect documents for the program outlined above, they also surveyed local residents to assess the quality of public service provision, particularly to assess the quality of Brazil's Family Health Plan (*Programa Saúde da Família*). A high fraction of healthcare in Brazil is provided by the government (Family Health Plan covers over 90% of Brazilians) and is implemented at the local level. For the poor of Brazil, Family Health Plan is generally their only way to receive care. If public funds are being stolen, municipal health clinics will be underfunded and will not be able to provide people with proper healthcare. Similarly, if the municipal government is exerting low

effort, healthcare providers may not be hired or incentivized to come to work, and people will not recieve care.

In each municipality, CGU auditors picked a random sample of residents to interview (22 families on average) and asked a series of questions relating to the quality of care received at local health clinics. In addition, this dataset also has nonsurvey data in which the auditors assessed the quality of public service provision based on documentation provided by the local government, including if Community Health Agents (CHA)[18] had been hired by irregular means. The questions asked are shown in table 3.2 in the appendix and summary statistics are provided in panel A of table 1.2.

[18]Community Health Agents in Brazil are government employees (within the Brazilian Unified Health System) with only basic healthcare training and report to a physician or nurse. They are generally selected from members of the community and make regular visits to families and promote good health.

Table 1.2: Healthcare Data Summary Statistics

Panel A: Survey Data			
	count	mean	sd
Nurse Present	5334	.9478815	.2222868
Dentist Present	4386	.7763338	.4167484
Physician Present	5359	.8303788	.3753347
Irregular Hiring	487	.3326489	.4716465
Lines at Health Unit	5278	.5617658	.4962173
CHA Visits	6730	.9257058	.2622686
Panel B: DATASUS Data			
Preterm Rate, Uned	45191	8.960052	7.841882
Infant Mort. Rate, Uned	45192	14.76405	30.48358
Preterm Rate, Educ.	45258	8.147874	5.623468
Infant Mort. Rate, Educ.	45260	8.873812	17.66823

The unit of observation of all variables in panel A is a respondent in an audited municipality and all variables are binary indicating the respondent's answer to the questions in table 3.2. Sample sizes differ slightly by variable because some respondents may not have needed to see each type of medical professional during their last visit. The 'Irregular Hiring' variable has much lower sample size because it is based on responses of auditors rather than survey respondents. Panel B contains infant health indicators for uneducated and educated mothers. Preterm rate is the fraction of births born before 37 weeks, infant mortality rate is the number of infant deaths per 1000 live births. The unit of observation is municipality-year between 2005 and 2016. In both panels we only include municipalities that fit the requirements for the RD design (detailed in section 3.4.1)

1.3.3 Health Outcome Data

If poor politician performance has an adverse affect on public service provision, we may expect to see some negative effect on outcomes that these public programs are targeted at. Thus, we investigate the effect that the marginal opposing city counselor has on two infant health outcomes: fraction of infants born preterm[19] (before 37 weeks) and the infant mortality rate per 1000 live births. We obtain this data from DATASUS, the data arm of Brazil's Unified Health System (*Sistema Único de Saúde*), the system that implements the Family Health Program. One advantage of this dataset is that unlike the CGU data, this data is available for every year in our sample and almost every municipality in Brazil (around 5,500). Thus we have a very high sample size and thousands of clusters.

Underfunding of Brazil's Family Health Program is most likely to affect low-income families who depend on it for their healthcare (wealthier families generally seek private care; Alves and Timmins, 2003). While health outcomes are not available by income level, they *are* available by education level. Thus, we classify mothers who received 1-7 years of schooling as 'uneducated' and mothers who received 8+ years of schooling as 'educated'.[20] In our data 34% of births are to uneducated mothers, 64% are to educated mothers, and 2% of mother's education levels are unreported. Summary statistics are in panel B of table 1.2.[21]

[19]Medical research shows that there a link between receiving adequate prenatal care and having a preterm birth. See Vintzileos et al. (2002)

[20]This is similar to the classification in Fujiwara (2015), though Fujiwara classifies 'uneducated' mothers as recieving 1-9 years of schooling. Though it may be ideal to reuse Fujiwara's classification, education level data for mothers is only provided in bins, and the Brazilian government reported the data in different bins during the period we study.

[21]While only 2% of mother's education levels are unreported for births, 35% of mother's education levels are unreported for infant deaths, leading to some measurement error for the infant mortality rate for educated or uneducated mothers. We assume this is because healthcare workers are in general less preoccupied with gathering this information from a mother after she has lost her infant, regardless of her educational attainment. However, even *if* lower education mothers are more likely to have their education levels unrecorded, our RD design still ensures that this measurement error is uncorrelated with regressors, and thus it will not bias our estimates.

1.3.4 CPI

We constructed the dataset on the creation of CPIs by searching on Google for any news reporting the creation of a CPI[22] in a municipality between the years of 2013 and 2016[23]. To limit the scope of this data collection, we follow Cattaneo et al. (2015) and use a data-driven method of determining the bandwidth around which treatment can be considered as if random. The method basically amounts to performing balance checks for a vector of covariates with successively smaller bandwidths until you can fail to reject that all the covariates are balanced at a conservative level. We performed these checks using the 12 covariates that we use for our placebo test in Table 4 and arrived at a bandwidth that includes 674 municipalities.

1.3.5 Election Data

We observe data on all candidates, parties, coalitions, votes received, and seats won, for the 2004, 2008, and 2012 Brazilian municipal elections. This data is available on the website of the Superior Electoral Court of Brazil (TSE). The identity of council leaders in 2005 comes from the database "Censo Legislativo 2005".

The summary statistics in table 1.3 paint a picture of what the average city council looks like. We can see that on average the city council has 9.3 seats and that there are about 4.6 different coalitions of parties competing for votes. While various separate coalitions may compete with one another in the elections for city council, multiple coalitions often unite into one super-coalition for the election of a single candidate for mayor. Thus the average mayor has 1.8 coalitions that supported him or her in the election. On average the mayor's coalition wins 4.8 of the seats in the city council, and as we can see from table

[22]We did not include CPIs that explicitly investigated only acts of legislators and not the executive.

[23]We focus on the most recent mayor term because it is the easiest one to find news report online.

1.20, the typical mayor's coalition is the highest ranked coalition in terms of vote share. As seen in figure 1.9 in the appendix, in the bulk of municipalities, the opposing coalition has between 33% and 55% of the seats.

Table 1.3: Municipality Characteristic Summary Statistics

Variable	Mean (S.D.)	Variable	Mean (S.D.) .
Tot. City Council Seats	9.311	AM Radio	0.213
	(1.963)		(0.409)
Second Term Mayor	0.279	TV Station	0.102
	(0.448)		(0.302)
Mayor's Coalitions's seats	4.767	Urbanization Rate	63.263
	(1.548)		(21.978)
Num. Coalitions in Municipality	4.654	Illiteracy Rate	16.36
	(2.5)		(9.862)
Num. Coalitions Supporting Mayor	1.852	Avg. Monthly Income	433.133
	(1.105)		(196.475)
Mayor Coalition Rank	1.462	Population 2010	34135.8
	(0.818)		(224619.0)
Anti-Mayor Coalition Vote Share	50.688	Tot. Votes Cast	17692.5
	(14.251)		(116821.2)
Opposition Seat Share	0.484		
	(0.155)		
Observations	4192		4192

Municipalities fitting the requirements for the RD design (detailed in section 3.4.1) in the 2004 election (the year corresponding to most of the audit data). Means are presented in the main row, standard deviations are below in parenthesis.

1.4 Theoretical Framework

We present here a simple two-period[24] political agency model modified to include a legislative body to guide the interpretation of our empirical results. The main objective of this model is to illustrate how equilibrium legislative investigations, corruption and mayor's reelection rates can differ when there is a strong legislative opposition versus when there is not.

There is one voter (v), one incumbent mayor (m_a), one challenger (m_b) and two legislature candidates (l_a and l_b) each coming from a party a or b. If the mayor and the legislator are from the same party[25], we will call this an unified government and if the mayor and the legislator are from different parties, we will call this a divided government. Our empirical results suggest that an extra opposition legislator has the most impact on politicians behavior when this extra legislator exogenously flips the control of the local city council. Hence, our analysis here will compare equilibrium level of corruption and oversight when the game begins with an unified government versus when it begins with a divided government.

All agents want to maximize the discounted sum of their utilities:

$$U_i = u_{i1} + \delta u_{i2} \tag{1.4.1}$$

The voter has to elect a mayor and a legislator in the second period. The utility of the voter is $u_v = \theta - C$ if he elects a mayor and a legislator from the same party (unified government) or $u_v = \theta - C - D$ if he elects a mayor and a legislator from different parties (divided government). θ is a random shock with a uniform [0,1] distribution, C is how much the elected mayor decided to steal and D is a loss of utility from having a divided government. D captures the idea that if the voter did not worry about corruption he would always choose

[24] This maps directly to the fact that mayors can only serve two consecutive terms in Brazil.
[25] m_a and l_a or m_b and l_b

one of the parties to control both branches of government to avoid gridlock or any other negative consequence from divided government. Voters don't observe θ.

There are two types of mayoral candidates: corrupt (c) and non-corrupt (nc). The probability of a mayor from party a being corrupt is p and from party b being corrupt is $p\prime$, with $p < p\prime$ and $\frac{2p}{p+1} < p\prime$. Corrupt mayors have a utility $u_{cm} = E + C - B$ if elected and zero otherwise. Non-corrupt mayors have utility $u_{ncm} = E$ if elected and zero otherwise. E is a ego rent politicians enjoy if elected; $C \in \{0, \bar{c}\}$, with $\bar{c} < 1$,is how much corruption rent the elected mayor extracts and $B \in [0, C]$ is how much the elected mayor gives to the legislature as bribe. Hence, by the very definition of his utility, a non-corrupt mayor does not extract rents.

Finally, the utility of legislators is $u_l = V + B$ if elected and zero otherwise. V is an ego rent legislators enjoy if elected and B is how much bribe they accepted from the mayor.

The timing of the game is as follows. Nature draws the type of the incumbent (M_a), the type of the challenger (M_b) and θ. The incumbent mayor observes θ and chooses the pair (B_o, C), where B_o is the bribe offer to the legislator. The elected legislator observes (B_o, C), if $C > 0$ he can report C to the voter[26] or take no action. If he reports the mayor, he cannot accept the bribe and $B = 0$, if he does not report the mayor, he accepts the bribe and $B = B_o$. If $C = 0$ the legislator has no action to take. Voters observe their utility and whether the mayor's actions were reported and decide whether to reelect the mayor and/or the legislator. Finally, in the second period, nature draws θ again, the elected mayor chooses a new pair (B_o, C) and the elected legislator decides to report corruption activities or not.

We start solving the game with the last period. Because it is the last period, a legislator will accept any bribe $B_o > 0$ and will be indifferent between

[26]This reporting is based on hard evidence, he cannot lie about C.

reporting the executive or not if $B_o = 0$. If a mayor is corrupt, he will make $B_o = 0$ and $C = \bar{c}\,\forall\theta$ because there are no reelection incentives. Since a corrupt mayor will always steal in a divided or unified government in the second period, the voter always elect a unified government in the second period to avoid paying the cost D. But voters still have to decide if they will elect a unified government under party a or b in the second period. This decision boils down to the voter's belief about the probability of the first period mayor been a corrupt type given the utility or reports he gets. Hence, if $P(c/u, r) < p\prime$, he reelects the mayor, otherwise, he does not. Therefore mayors of a corrupt type might want to behave as non-corrupt mayors in the first period to get reelected and enjoy ego and corruption rents in the second period.

The incentives of legislators to accept a bribe or report a mayor in the first period depends if it is allied with the mayor or not. Since,under a unified government, the reelection of the legislator is completely tied to the reelection of the mayor, he will never reveal any wrongdoings of the mayor because this would get the mayor kicked out of office along with him. On the other hand, under a divided government, the reelection of the legislator is tied to the incumbent mayor been kicked out of office. Hence, he will reveal any wrongdoing of the mayor unless the bribe offered by the mayor is high enough so that he is willing to forgo his reelection.

To solve the first period of the game, we will need to break it down in two cases depending on how much a corrupt mayor values getting reelected and staying in office versus and how much they value extracting rents today but getting kicked out of office. We will focus on cases where $\bar{c} < \delta V$ in all sub cases below.[27].

[27]Otherwise, legislators are cheap to bribe and the equilibrium will involve corrupt mayors' bribing legislators to get away with corruption. For this parameter values, there is no difference in equilibrium corruption, oversight and reelection probability between a divided or unified goverment.

Case 1: $\bar{c} > \dfrac{E}{1-\delta}$

Under a Unified Government in period 1

As in all cases under a unified government, the legislator will never be better off reporting the mayor because by doing so the voter will know that the mayor is a corrupt type and will not reelect any of the branches, hence, $B_o = B = 0$ and the legislative never reports.

In this specific case, the rents a corrupt mayor can extract today are too large compared with what he can get tomorrow. Hence, he makes $C = \bar{c}\ \forall\theta$. Voters reelect the mayor if $U_{v1} > 0$ and do not reelect the mayor if $U_{v1} \leqslant 0$. Voter beliefs are $P(c/u_{v1}) = 1$ if $u_{v1} < 0$, $P(c/u_{v1}) = p$ if $1 - \bar{c} > u_{v1} > 0$ and $P(c/u_{v1}) = 0$ if $1 - \bar{c} < u_{v1}$.

Under a Divided Government in period 1

As in all cases under a divided government, the legislator will accept the bribe and won't report the mayor if $B_o \geq \delta V$ and will report the mayor otherwise.

In this specific case, the rents a corrupt mayor can extract today are too large compared with what he can get tomorrow and the legislator is too expensive to buy off. Hence, he makes $C = \bar{c}$ and $B_o = 0\ \forall\theta$. Legislators will always report the mayor and he will not be reelected.

Case 2: $\bar{c} < \dfrac{E}{1-\delta}$

Under a Unified Government in period 1

As we argued before under a unified government, the legislator will never be better off reporting the mayor. Hence, $B_o = B = 0$ and the legislative never reports.

In this specific case, the rents a corrupt mayor can extract tomorrow are enough to prevent him from stealing if he knows he will get caught. Hence, he makes $C = \bar{c}$ for $\theta > \bar{c}$ and $C = 0$ for $\theta < \bar{c}$. Voters always reelect the mayor. Voter beliefs are $P(c/u_{v1}) = \frac{2p}{p+1}$[28] if $\bar{c} > u_{v1} > 0$, $P(c/u_{v1}) = p$ if $1 - \bar{c} > u_{v1} > \bar{c}$ and $P(c/u_{v1}) = 0$ if $1 - \bar{c} < u_{v1}$.

Under a Divided Government in period 1

As we argued before under a divided government, the legislator will accept the bribe and won't report the mayor if $B_o \geq \delta V$ and will report the mayor otherwise.

In this specific case, the rents a corrupt mayor can extract tomorrow are enough to prevent him from stealing if he knows he will get caught. Also, the legislator is too expensive to be bought off. Hence, he makes $C = 0$ and $B_o = 0$ $\forall\theta$. Legislators won't have anything to report and the mayor will always get reelected.

Result 1 : Expected corruption C under a unified government is greater than or equal to expected corruption under a divided government

If $\bar{c} < \dfrac{E}{1-\delta}$:

$E(C) = p(1 - \bar{c}))\bar{c}$ in unified government and $E(C) = 0$ under a divided

[28]Remember that by assumption $\frac{2p}{p+1} < p\prime$.

government

If $\bar{c} > \frac{E}{1-\delta}$:

$E(C) = p\bar{c}$ in unified government and in a divided government

Result 2 : The probability of the legislator reporting the mayor $P(Report)$ under a unified government is smaller than or equal to the probability of reporting under a divided government

If $\bar{c} < \frac{E}{1-\delta}$:

$P(Report) = 0$ in unified government and in a divided government

If $\bar{c} > \frac{E}{1-\delta}$:

$P(Report) = 0$ in unified government and $P(Report) = p$ under a divided government

Result 3 : The probability of the mayor getting reelected $P(Reelected)$ under a divided government is greater than or equal to the probablity of reelection under a divided government

If $\bar{c} < \frac{E}{1-\delta}$:

$P(Reelected) = 1$ in unified government and in a divided government

If $\bar{c} > \frac{E}{1-\delta}$:

$P(Reelected) = p(1 - \bar{c}) + 1 - p$ in unified government and $P(Reelected) = 1 - p$ under a divided government

This simple model generates these 3 predictions that we can take to the data. According to our model, municipalities with a first-term mayor and more opposition legislators in the city council are (1) more likely to have investigations launched into the mayor's actions, (2) less likely to have corruption episodes, and (3) less likely to reelect the incumbent mayor.

1.5 Econometric Model

1.5.1 Constructing the Running Variable

In order to identify causal effects we use a regression discontinuity (RD) model which exploits close elections.[29] As a key part of our RD model we construct a running variable that serves as a measure for how close a given election was. We use detailed knowledge of Brazil's electoral system in order to construct this running variable. Brazil has an open party-list proportional system, with seats allocated according to the D'Hondt Method and with coalitions treated as single parties. In order to illustrate how the D'Hondt Method works, consider the following example.

Imagine three different coalitions are competing for 6 seats in a fictional city council. The coalition of parties A & B receives 100,000 votes, the coalition of parties C & D receives 80,000 votes, and party E, which is running as an isolated party, receives 20,000 votes. First, the 'electoral quotient' is calculated, which is the total amount of votes cast divided by the number of seats available. In our case, the electoral quotient is $(100,000 + 80,000 + 20,000)/6 = 33,333$.

[29]For identification of causal effects using the RD design see Hahn, Todd, and Van der Klaauw (2001). For a primer on RD see Imbens and Lemieux (2008).

Only coalitions whose raw vote count exceeds the electoral quotient are eligible to be awarded seats. Thus, party E is already disqualified from winning seats, since it only received 20,000 votes. After this, a series of quotients is calculated for each coalitition, according to the formula

$$Q_s = \frac{V}{s+1}$$

where V is the total of votes the coalition received and s is the round of calculation (or number of seats already awarded to the coalition). In an election where n seats are available, coalitions are awarded 1 seat for each quotient they have among the highest n quotients.

This is illustrated in the table below, where the coalition of A & B and the coalition of C & D have both been awarded 3 seats, since both have 3 quotients among the top 6 quotients.

D'Hondt Method Example (6 seats available)

	Q_0	Q_1	Q_2	Q_3	Q_4	Seats Won
Parties A+B	$100,000^*$	$50,000^*$	$33,333^*$	25,000	20,000	3
Parties C+D	$80,000^*$	$40,000^*$	$26,666^*$	20,000	16,000	3
~~Party E~~	~~20,000~~	~~10,000~~	~~6,666~~	~~5,000~~	~~4,000~~	0
Note: Asterisks denote quotients in the top 6. Each quotient a coalition has in the top 6 of quotients earns them a seat.						

Because of the proportional representation system, only the last seat filled in each municipality (city) can be considered as being 'on the margin' of being flipped from one coalition to another. In the example above, that is the seat awarded to the coalition of parties C & D for their quotient of 26,666 (which was the lowest quotient of the top 6). In this case, the coalition of parties A & B was also on the margin of winning this seat, because of their quotient of

25,000 (which was the highest quotient *not* in the top 6).[30] We consider these two quotients (26,666 and 25,000) as the 'marginal quotients'.

We construct the running variable for our sharp RD design as the opposition coalition's margin of victory (or loss) between quotients for the last seat filled in the legislature, scaled by the total amount of votes cast. Thus, for each municipality m, the running variable is formally defined as

$$R_m = \frac{Q_{m,opposition} - Q_{m,mayor-coalition}}{TotalVotes_m},$$

where $Q_{m,opposition}$ is the marginal quotient of the opposition coalition, $Q_{m,mayor-coalition}$ is the marginal quotient of the mayor's coalition, and $TotalVotes_m$ is the total amount of votes cast for city councillors in the election. In the example above, if the coalition of parties A & B was allied with the mayor, and the coalition of parties C & D was the opposition, the running variable would then be calculated as $R_m = \frac{26{,}666-25{,}000}{200{,}000} = 0.0083$.

Notice that in the example above the opposition is capturing its 3rd seat in the city council, but in our dataset the marginal opposition quotient can correspond to the opposition capturing any seat (the 1st, 2nd, 3rd and so on). Because our dataset is not large enough to precisely estimate the treatment effect at each individual threshold (the effect of the 1st, 2nd, or 3rd seat occupied by an opposing city councillor), we stack all of the thresholds and our estimated treatment effect is a weighted average of the treatment effect at various seats (as discussed in Cattaneo et al., 2016) rather than the treatment effect at a single seat.

In practice, there are often 3 or more coalitions in the legislative election, rather than just a pro-mayor and an opposition coalition. We make the simplifying assumption that parties that are not formally allied with the mayor in

[30]Note that parties A & B didn't miss winning that seat by 1,666 votes, they actually missed it by 4*1,666=6664 votes, since the total votes was scaled down for purposes of the quotient.

the election coalitions are against him or her, and are considered the political opposition.[31]

We drop observations from the analysis if they meet any of the following criteria[32]. :

- The two marginal quotients are from the same coalition, and thus the concept of 'being on the margin' is no longer valid. This happens when the marginal seat is passing from one opposition coalition to another, or from one pro-mayor coalition to another. (25% of obs)

- The quotients we calculate from vote totals do not acurately predict the amount of seats awarded to each party. This is may happen due to a gap in coalition data. (1.5% of obs)

- Coalitions are such that two parties are allied in the election for city councilors but not in the election for mayor. (1.5% of obs)

Figure 1.1 shows the variation that this identification strategy in generating. On average, on the left side of the discontinuity the opposition represents 45% of the local city council, whereas on the on the right side of the discontinuity it represents 55%. Notice that beyond just increasing the opposition share in the city council, on average our variation also flips the opposition from minority to majority on the city council.

Note that it is hard to interpret the correlation between our running variable and outcomes of interest outside the cutoff. As Figure 1.1 makes clear,

[31]This identification strategy bears some resemblance to that of Folke (2014), though our strategy is simpler as we estimate only the effects of opposition in general, not the effects of specific parties.

[32]In table 3.3 as well as table 1.13 (in the appendix, section 1.9.1), we compare election and municipal characteristics of the RD sample and the non-RD sample. As can be seen, municipalities in the RD sample have slightly fewer coalitions overall. This is expected given that when there are fewer coalitions it is more likely that the marginal seat is passing from a pro-mayor party to an anti-mayor party, or vice-versa (if there are only 2 coalitions, then a seat can only pass from a pro-mayor coalition to an anti-mayor coalition, but if there are 4 coalitions, 2 for the mayor and 2 against, then the marginal seat may pass from one pro-mayor coalition to another pro-mayor coalition, invalidating the RD), but in general dropped municipalities are qualitatively similar to municipalities in the RD sample.

higher values of the running variable are not necessarily associated with a larger presence of opposition in the local city council. This happens because our running variable is stacked across all seat thresholds rather than at a single seat threshold as discussed earlier.

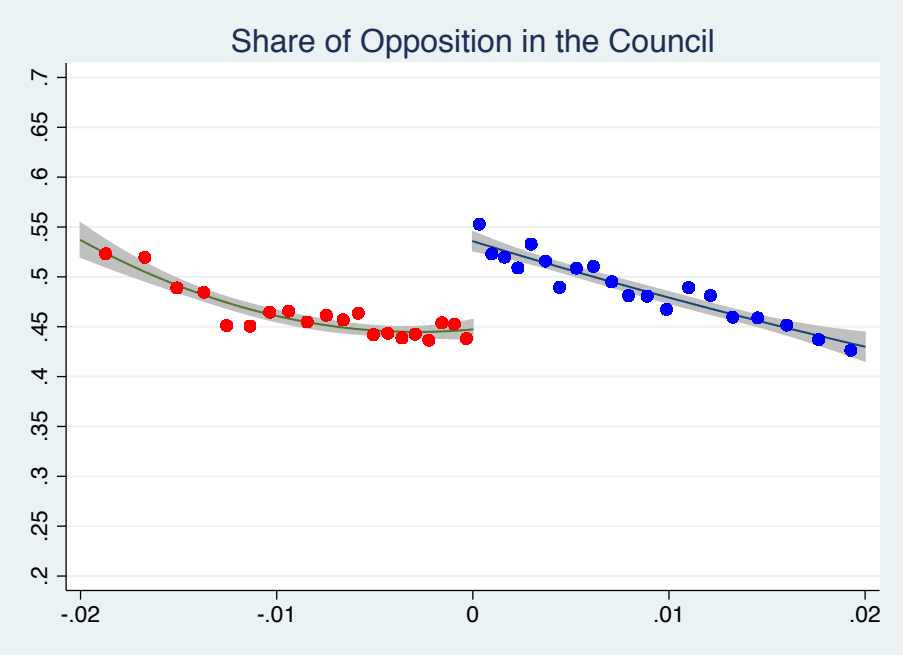


Figure 1.1: Share of Opposition in the City Council

Notes: Here we present the RD plot for the opposition seat share in the local city council. This plot presents equally sized bins with means of the dependent variable inside each one. It also presents the prediction for the dependent variable from a regression of the dependent variable on the running variable and on the running variable squared and plots the resulting line along with a 95% confidence interval.

1.5.2 Effect on Corruption

With this running variable we estimate a series of sharp RD models. The first model, which estimates the effect of political opposition on various measures of corruption, can be written

$$c_{m,t} = \alpha_1 + \alpha_2 T_{m,t} + \alpha_3 f(R_{m,t}) + \alpha_4 f(R_{m,t}) * T_{m,t} + \varepsilon_{m,t} \tag{1.5.1}$$

for municipality m during term t.[33] Where $c_{m,t}$ is the amount of corrupt violations found in the municipality, the function $f(*)$ is a polynomial specification

[33] A small number of municipalities were audited twice in the same term. In this case we only consider the first audit report.

for the running variable, and $T_{m,t} = I[R_{m,t} > 0]$ denotes treatment status. Thus, α_2 is the main parameter of interest. For each outcome, we report in section 3.5 our preferred specification in tables (using the CCT optimal bandwidth from Calonico, Cattaneo, and Titiunik (2014) and a quadratic polynomial specification), and in section 1.10.1 in the appendix as a robustness check, we report a number of alternate RD specifications, varying the bandwidth and polynomial order. In the appendix we also report CCT's bias-corrected and robust variance estimates from Calonico, Cattaneo, and Titiunik (2014).

1.5.3 Effect on Public Service Provision

The second model estimates the effect of political opposition on the quality of public service provision. This model differs from the first model in that we now observe data at the individual level rather than the municipality level, also it is done as a linear probability model[34], since survey answers are yes/no answers. The model can be written

$$ps_{i,m,t} = \beta_1 + \beta_2 T_{i,m,t} + \beta_3 f(R_{i,m,t}) + \beta_4 f(R_{i,m,t}) * T_{i,m,t} + \epsilon_{i,m,t} \qquad (1.5.2)$$

for individual i in municipality m during term t. Where $ps_{i,m,t}$ is an indicator for the respondent answering 'yes' to a given question relating to public service provision, and now β_2 is the main parameter of interest. In this model standard errors are clustered at the municipality level. Once again, we report in tables our preferred specification and the appendix contains a variety of robustness checks.

It should be noted that this model does not limit the channel through which political opposition may affect public service provision. Given the nature of our identification strategy, we cannot distinguish between an improvement in public services and health outcomes that is due to the reduction in corruption that

[34]Results are the same for probit/logit models.

we see, versus an improvement that is a direct result of increased opposition. For example, increased opposition might directly induce higher mayor effort in the provision of public goods, independent of whether or not it reduces corruption. However, it is clear from the results presented in this paper that political opposition improves politician performance.

1.5.4 Effect on Health Outcomes

The third model estimates the effect of the marginal politicaly opposed legislator on various health outcomes. In this model we again observe municipalities, though this model differs from the first model in that now we observe data for each year within an electoral term. The model can be written

$$h_{y,m,t} = \gamma_1 + \gamma_2 T_{y,m,t} + \gamma_3 f(R_{y,m,t}) + \gamma_4 f(R_{y,m,t}) * T_{y,m,t} + u_{y,m,t} \qquad (1.5.3)$$

for municipality m in year y during term t. Where $h_{y,m,t}$ is a health outcome measure and now γ_2 is the main parameter of interest. In this model standard errors are also clustered at the municipality level. For each of these models, tables 1.5 through 1.8 contain our preferred specification and the appendix section 1.10.1 contains alternate specifications to test robustness.

1.5.5 RD Design: Smoothness and Balance Tests

The key assumption for RD to be valid is the smoothness or 'no precise manipulation' assumption. We test for manipulation of the running variable using the test outlined in McCrary (2008) and using the local polynomial methods put forth in Cattaneo, Jansson, and Ma (2017). We find no evidence of manipulation, failing to reject the null hypothesis of no manipulation with a p-value of 0.97. Figure 3.2 presents visual evidence of this.

As additional evidence for the validity of the RD design, in tables 1.28 to 1.30 in the appendix, we present results from placebo tests for each of the 3

models we present above, estimating the effect of treatment on a variety of municipal characteristics where no treatment should be found. Our results show a broad covariate balance across the 32 variables we test.

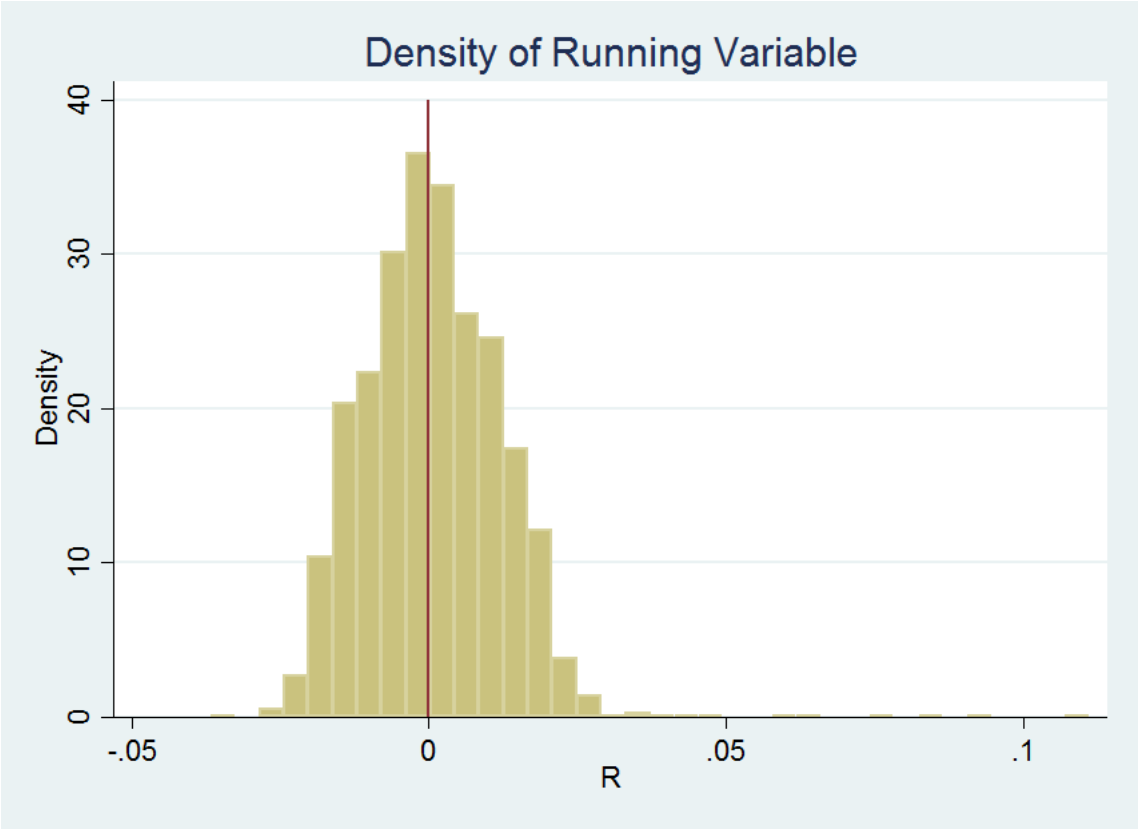


Figure 1.2: Density of the running variable (margin of victory/loss for the marginal legislator politically opposed to the mayor). Using McCrary's (2006) density test, we find no evidence of manipulation of the running variable.

1.6 Results

1.6.1 Effects on Legislative Oversight

We first show results for the most immediate outcomes: whether the president of the city council belongs to the opposition and whether a CPI investigation was opened by the city council. For the control of the city council (whether or not the city council president is a member of the mayor's coalition), we use the RD model outlined above and find that flipping one city council member from pro-mayor to anti-mayor increases the likelihood that the city council will be

controlled by an opposition legislator by 14 percentage points, amounting to a 43% change in probability. This is unsurprising given that most city councils have 9 seats, and thus, treatment entails an 11.11 percentage point change in composition. Figure 1.3 presents the corresponding RD plot[35].

Next we examine the effect that the marginal opposition legislator has on the probability that the mayor will be investigated for wrongdoing using a CPI. As outlined in section 1.3.4, in order to limit the scope of our data collection, we use the algorithm outlined by Cattaneo, Frandsen, and Titiunik (2015) in order to find the bandwidth around which treatment can be considered random, and then we regress treatment on CPI opening for only this bandwidth. Here we find that the marginal opposition legislator increases the likelihood of a CPI investigation by about 6 percentage points, amounting to a 55% increase in the likelihood of an investigation.

Table 1.4: Effect of Opposition on Council Functioning

	Opposition Council President	CPI Opened
Mean	.322	.1086
T	0.140***	0.0596**
	(0.0459)	(0.0264)
Observations	3216	674

Effect of the marginal opposing legislator. S.E. in parenthesis.
For 'Opposition Countil President', we use the RD estimator and the mean reported is the overall mean. For 'CPI Opened', we use the randomization algorithm outlined in Cattaneo, Frandsen, and Titiunik (2015), and the mean reported is the control mean.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[35]As a matter of consistency, we use a 0.02 bandwidth in all RD plots presented. But it is important to note that the bandwidth used in the estimation tables vary for each dependent variable according to the CCT optimal bandwidth procedure.

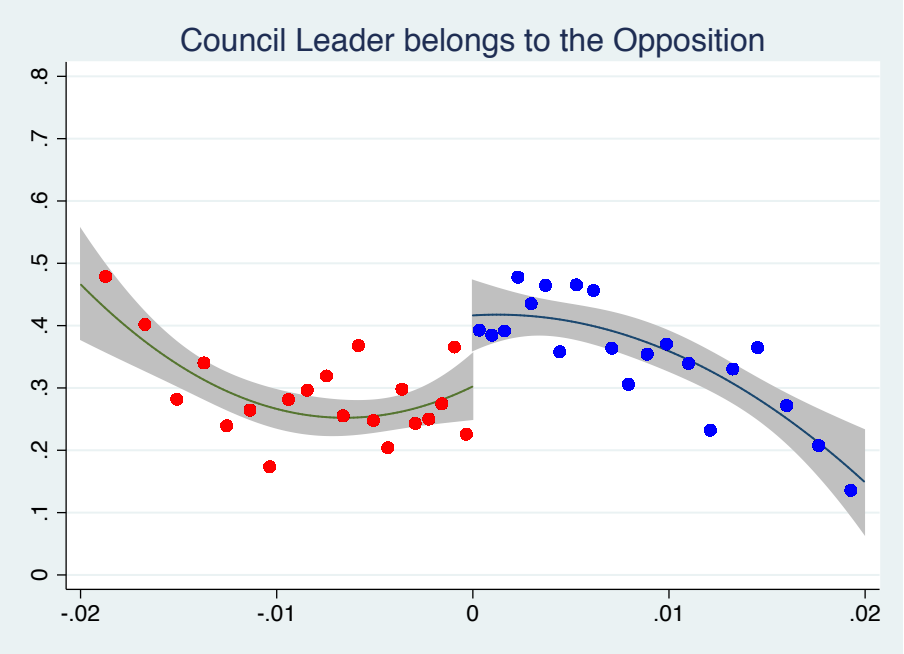


Figure 1.3: Council Leader Belongs to Opposition

Notes: Here we present the RD plot of the probability that the city council president is a member of the opposition's coalition. This plot presents equally sized bins with means of the dependent variable inside each one. It also presents the prediction for the dependent variable from a regression of the dependent variable on the running variable and on the running variable squared and plots the resulting line along with a 95% confidence interval. For consistency, we use a 0.02 bandwidth for all RD plots, which is generally wider than the CCT optimal bandwidth used for coefficient estimates.

1.6.2 Effect on Corruption

Our estimates for the treatment effect in equation 2.3.1 are reported in table 1.5. Here we see that when the mayor faces reelection incentives (first term mayors), the marginal politically opposed legislator can decrease the amount of items audited found to have corruption by approximately 1.353 items, which amounts to a 0.846 decrease in standard deviation units[36]. Figure 1.4 presents the corresponding RD plot. This effect is driven largely by embezzlement, which is decreased by 0.889 items. This is a sizable effect given that the average municipality only has 1.6 items audited found to be involved in some

[36]We find no effects for mayors without reelection incentives (second term mayors). While our theoretical model considers only electoral costs to corruption, there are also potentially criminal costs of corruption. The fact that we find no effect for second term mayors suggests that information discovered and revealed by the city council is politically damaging rather than criminally damaging, or perhaps that malfeasant actions are observable by the legislature (may be used for political purposes), but not verifiable (prosecutable).

type of corruption.[37] Note that there is no significant effect on 'Other Irregularities', evidence that our embezzlement variable is indeed measuring cases of embezzlement and not mere procedural errors.

Compared to the effect on embezzlement, the effects on fraud and overinvoicing are small, in addition to being statistically insignificant. Going back to the discussion in section 3.2, this is evidence that the legislative branch can restrain the executive branch's rent-seeking through its auditing responsibilities, but not through its lawmaking responsibilities. Restraining the executive from transferring money from government bank accounts to his or her own private bank account is fairly straightforward as long as the city council requires that the mayor and his secretaries provide receipts and other documentation for all expenditures they make. However, knowing which public projects proposed by the mayor are likely to be fraudulent may be more difficult. These findings also run contrary to the idea that corruption in Brazil is completely institutionalized, independent of party affiliation.

These results are robust to a variety of alternate specifications as shown in the appendix in section 1.10.1.

[37] For municipalities where corruption was found, the average is 2.8 items found to be involved with some type of corruption.

Table 1.5: Corruption Outcomes, First Term Mayors

	Total Corruption	Embezzlement	Fraud	Overinvoicing	Other Irreg.
T	-1.353**	-0.889***	-0.329	-0.270	0.443
	(0.578)	(0.296)	(0.309)	(0.279)	(1.065)
Optimal BW	0.006	0.007	0.008	0.009	0.009
Obsevations	244	260	294	329	352

Treatment effect of the marginal opposing legislator on corruption, using the CCT optimal bandwidth and a uniform kernel. Standard errors in parenthesis. Outcomes are the count of items audited with a violation. Other irregularities captures all other irregularities found by auditors that were not classified as embezzlement, fraud, or overinvoicing, and includes a host of procedural and other errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, we briefly examine the effect of political opposition on corruption by government sector. In our audit data the largest 3 sectors in terms of irregularities are the ministries of Health, Education, and Social Development, together making up 85% of the irregularities discovered by auditors, with various other ministries making up the remainder. For this analysis we again estimate the model from equation 2.3.1, but using the total corruption variable differentiated for these 3 different sectors of the government. In table 1.6 we find that effect of political opposition on corruption is largely concentrated in the health sector, with some evidence of effects in the education sector. This motivates the next section of our analysis, in which we estimate the effect of political opposition on healthcare provision and health outcomes.[38]

[38]We also examined possible effects on education outcomes, which sector is statistically significant at the 10% level in table 1.6, but we found no clear effects.

Table 1.6: Total Corruption by Government Sector, First Term Mayors

	Ministry of Health	Ministry of Education	Ministry of Social Dev.
T	-0.907***	-0.470*	0.0259
	(0.314)	(0.273)	(0.103)
Optimal BW	.007	.008	.009
Observations	260	297	307

Treatment effect of the marginal opposing legislator on total corruption, estimated by government sector. We use the CCT optimal bandwidth and a uniform kernel. Standard errors in parenthesis. Outcomes are the count of items audited with a violation for each sector.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

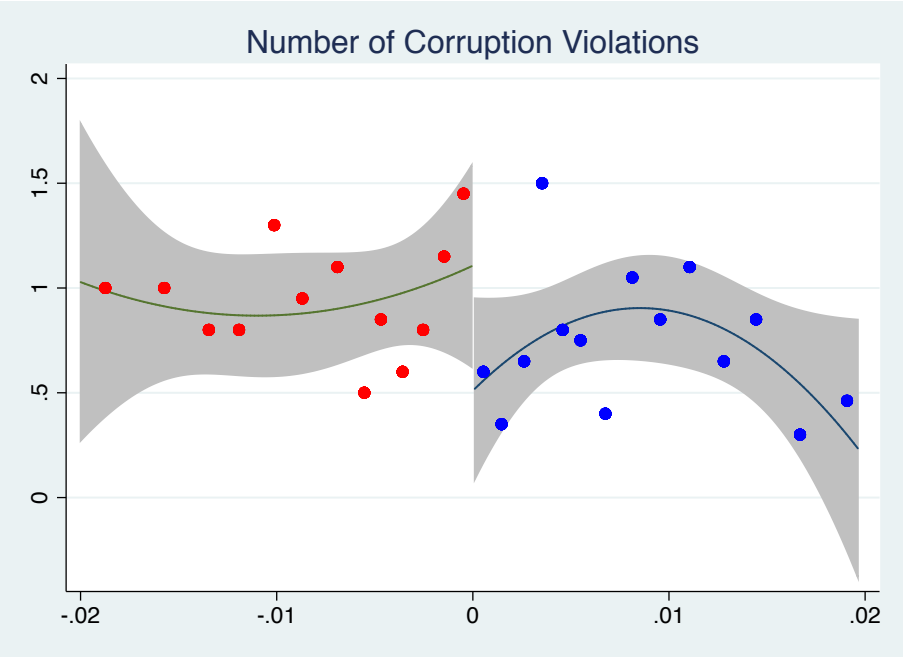


Figure 1.4: Total Corruption

Notes: Here we present the RD plot of the amount of items audited found to have corruption. This plot presents equally sized bins with means of the dependent variable inside each one. It also presents the prediction for the dependent variable from a regression of the dependent variable on the running variable and on the running variable squared and plots the resulting line along with a 95% confidence interval. For consistency, we use a 0.02 bandwidth for all RD plots, which is generally wider than the CCT optimal bandwidth used for coefficient estimates.

1.6.3 Effect on Healthcare Provision

If the marginal politically opposed city councillor can decrease the amount of healthcare money embezzled by the executive branch, we may hope that the money not being stolen is finding its way to the programs it is meant for. Complementary to this, there may be some direct effect of higher politician effort on public service provision. Thus, we now examine the pass-through

effects on public service provision, particularly healthcare. In Brazil a large amount of healthcare is provided by the government, with local government health clinics staffed by physicians, dentists, and nurses. If money is being stolen from public coffers, local governments will be cash-constrained and may be unable to hire the adequate amount of healthcare professionals, or unable to pay the professionals they have already hired. Approximately 51% of physicians in Brazil work in both the public and private sectors.[39] If these physicians are unpaid for their public work because of missing funds, they will likely substitute towards their private practice, leaving the public clinics understaffed and unable to provide care to those who depend on Brazil's public healthcare. Additionally, it may be that mayors with more opposition in the city council exert higher effort in assuring that health clinics are adequately staffed[40]. We examine this effect on public service provision in table 1.7, which contains our estimates for the treatment effect from equation 2.3.2. Because of our small amount of clusters in the healthcare provision data, in this table we pool first and second term mayors together.

Estimates in panel A of table 1.7 suggest that flipping a city counselor from being allied with the mayor to being opposed to the mayor results in a 34 p.p. increase in the likelihood a dentist will be present to see patients, and a 28 p.p increase in the likelihood that a physician will be present to see patients. Figure 1.5 presents the corresponding RD plots.

Results in panel B show that political opposition also improves public services by decreasing the likelihood that patients will have to wait in lines to

[39]Meanwhile 27% work exclusively in private medicine and 22% work only in public medicine (Scheffer M. et al. 2015).

[40]In fact, in collecting the CPI data, we found multiple investigations opened specifically to investigate missing doctors at local health clinics:
https://www.radioaltouruguai.com.br/tres-passos-instalada-cpi-para-investigar-supostas-irregularidades-no-cumprimento-de-horario-dos-profissionais-de-postos-de-saude/
http://www.unicanews.com.br/politica/vereador-quer-cpi-para-investigar-falta-de-medicos-nas-unidades-de-saude/10165
http://www.al.ms.gov.br/Noticias/64542/denuncias-de-cobrancas-irregulares-falta-de-medicos-e-ma-gestao-de-recursos-marcam-reuniao-da-cpi-da-saude

recieve care at local health units (resulting in a 40 p.p decrease in probability), and decreasing the likelihood that the municipal government uses irregular hiring practices for healthcare professionals (while the coefficient is not interpretable due to the steep slope of the regression function, it is clear that there is a large effect).

In recent years, Brazil has experienced a significant shortage of physicians at public health clinics. There are various news articles documenting this (Falcão and Amorim; 'Postos de saúde') and the Brazilian Federal government has responded with an ongoing federal program which started in 2013, *Mais Médicos* (More Doctors), which is aimed at recruiting more physicians for public health clinics, including importing them from other countries. Experts cited in these news articles suggest that this public sector shortage of physicians could be due to under financing of the public health system. Our results suggest that a significant share of the physician shortage in the public sector could explained by missing funds due to corruption, and that much of this corruption could be stopped by legislators who will fulfill their oversight responsibilities.

Table 1.7: Public Service Provision, First and Second Term Mayors

Panel A: Presence of Healthcare Professionals			
	Nurse Present	Dentist Present	Physician Present
T	0.150	0.339**	0.277**
	(0.128)	(0.139)	(0.128)
Optimal BW	0.008	0.008	0.009
Observations	2711	2233	2751
Clusters	127	121	130
Panel B: Community Health Agents and Lines			
	Irregular Hiring	Lines at Health Unit	Healthcare Visits
T	-1.055***	-0.402**	-0.0369
	(0.256)	(0.170)	(0.0487)
Optimal BW	0.007	0.007	0.006
Observations	243	2439	2482
Clusters	74	119	117

Treatment effect of the marginal opposing legislator on public service provision outcomes, using the CCT optimal bandwidth and a uniform kernel. Clustered standard errors in parenthesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

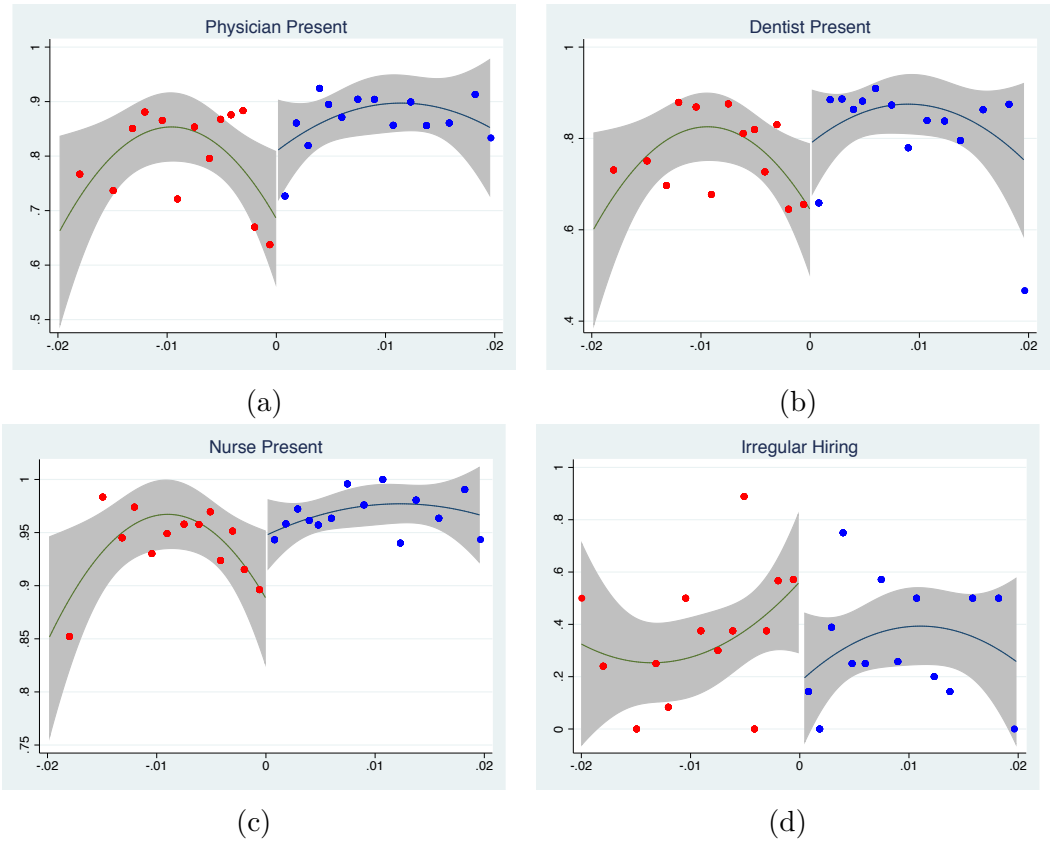


Figure 1.5: **Effect on Healthcare Provision**

Notes: Here we present the RD plots for particular outcomes of interest. These plots present equally sized bins with means of the dependent variable inside each one. They also present the prediction for the dependent variable from a regression of the dependent variable on the running variable and on the running variable squared and plots the resulting line along with a 95% confidence interval. For consistency, we use a 0.02 bandwidth for all RD plots, which is generally wider than the CCT optimal bandwidth used for coefficient estimates.

1.6.4 Effect on Health Outcomes

Going further down the causal chain, we now estimate the effect that the marginal opposing legislator has on actual health outcomes, particularly infant health. If physicians are missing from local health clinics due to missing funds or low politician effort, it is natural to expect to see negative health outcomes for the population these clinics serve. Alves and Timmins (2003) show that poorer (uneducated) Brazilians are those who are most likely to use the public health system, whereas wealthier Brazilians generally seek care from private providers. Thus, we expect to see some effect on health outcomes for the uneducated, especially for infants, who are among the most vulnerable.[41] Table

[41] It is also relevant to note that many researchers use infant mortality rate as a proxy measure for the overall health of a population.

1.8 contains our estimates for the treatment effect from equation 2.3.3. Figure 1.6 presents the corresponding RD plots.

The most interesting finding from this section is that in panel A of table 1.8 we see that in municipalities where the mayor faces reelection incentives, the marginal opposing legislator can improve executive performance to the extent that on average it lowers the infant mortality rate by 3.4 deaths per 1000 live births for uneducated mothers. This estimate is robust to a whole range of specifications as shown in figure 1.22 in section 1.10.1 in the appendix. Given that the overall infant mortality rate in Brazil is approximately 15 per 1000 live births, this is a large effect for only a single city councillor to have.

The coefficient on the rate of preterm births is also significant at the 10% level, preterm birth being a leading cause of child deaths (*Preterm Birth* 2016)[42]. This estimate is also robust to different specifications as shown in figure 1.21 in section 1.10.1 in the appendix. As predicted, no effect is found on infant health outcomes for educated mothers.

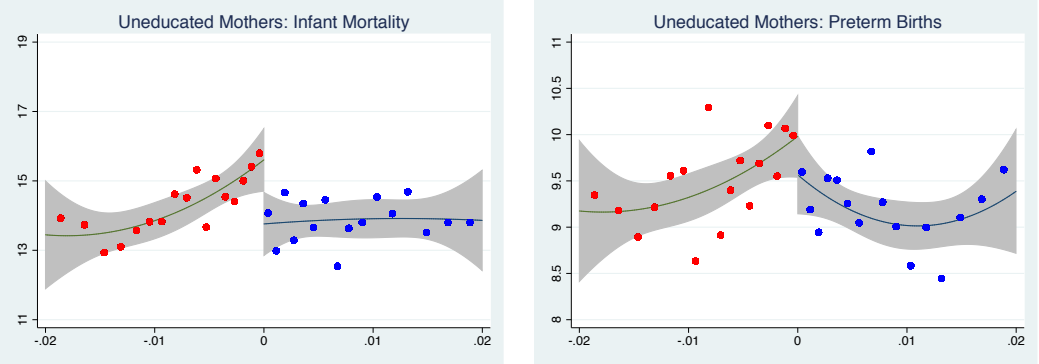
[42]We find no effects on infant mortality or preterm births for mayors without reelection incentives (second term mayors).

Table 1.8: Public Health Outcomes, First Term Mayors

Panel A: Uneducated Mothers		
	Preterm	Infant Mort. Rate
T	-0.824*	-3.402***
	(0.431)	(1.254)
Optimal BW	0.009	0.010
Observations	17297	19819
Clusters	3580	3945
Panel B: Educated Mothers		
	Preterm	Infant Mort. Rate
T	-0.377	-0.228
	(0.304)	(0.765)
Optimal BW	0.009	0.007
Observations	18645	15136
Clusters	3772	3238

Treatment effect of the marginal opposing legislator on public health outcomes, using the CCT optimal bandwidth and a uniform kernel. Clustered standard errors in parenthesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

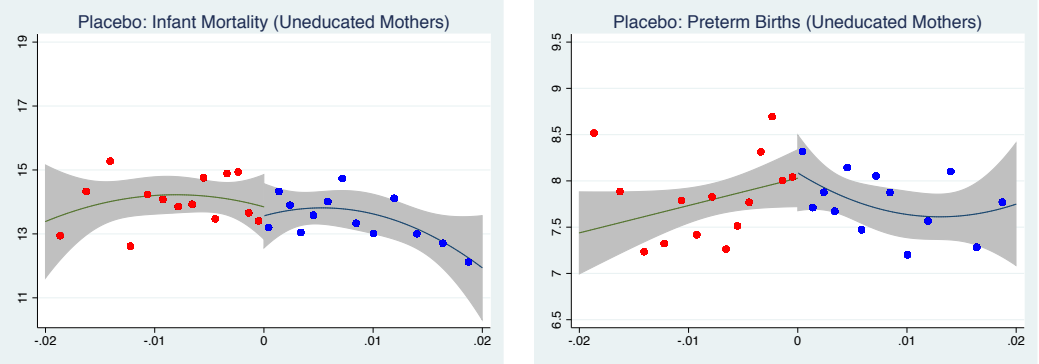


(a) (b)

Figure 1.6: **Effect on Health Outcomes**

Notes: Here we present the RD plots for particular outcomes of interest. These plots present equally sized bins with means of the dependent variable inside each one. They also present the prediction for the dependent variable from a regression of the dependent variable on the running variable and on the running variable squared and plots the resulting line along with a 95% confidence interval. For consistency, we use a 0.02 bandwidth for all RD plots, which is generally wider than the CCT optimal bandwidth used for coefficient estimates.

Leveraging the fact that the health outcomes above are in panel format (we observe the same municipality for different year), we can perform a placebo test for our estimated effects on health outcomes. We reestimate the previous RD results using health outcomes during the previous mayoral term as the dependent variable. If our identification strategy is valid, then having an extra opposition legislator in term t should have no effect on health outcomes in that municipality in term $t-1$. Figure 1.7 shows the graphical results of this exercise. Unlike Figure 1.6, Figure 1.7 shows no differences in health outcomes around the discontinuity whatsoever, strongly suggesting that our results are not been driven by any possible imbalances among municipalities around the threshold.



(a) (b)

Figure 1.7: **Placebo: Effect on Lagged Health Outcomes**

Notes: The plots show the estimated treatment effect with 95% confidence intervals for each marginal seat the can be captured by the opposition for different dependent variables.

1.6.5 Heterogeneous Effects by Seat

With the main effects established, we now examine heterogeneity of the treatment effect depending on what marginal seat the the opposition is capturing. As we discussed in Section 5, our previous models stack all of the seat thresholds and our estimated treatment effect is a weighted average of the treatment effect at all seats.

However, it is possible that the effects of an extra opposition legislator are strongest when this legislator changes which coalition has a majority in the city council (going from 4 to 5 seats in a 9-seat council, for example), giving the opposition real power in the legislative branch to open an investigation of the mayor, choose who investigates and how, and block corrupt legislation. It is also possible that these effects have little to do with specific institutional features that only a majority can use. For example, maybe more opposition legislators simply means more oversight capacity because there are more people to share the work. In this case, it should not matter which seat the opposition is capturing[43], i.e., no heterogeneous effect by seat.

For this section we focus our analysis on municipalities whose city council has only 9 seats, which makes up 90% of councils in our sample. In table 1.9

[43]In this framework, if there are decreasing marginal returns for opposition legislators, you could also imagine the effect been strongest for the first seat captured by the opposition.

we compare the treatment effect of the opposition winning its 5th seat with the effect for all other seats pooled together.[44] We find broad evidence that the marginal opposing legislator affects the mayor's performance mainly when it flips the majority in the city council. All estimated coefficients point to a much larger effect for the 5th seat. This result suggests that to promote effective oversight, the opposition needs to have access to institutional features that are only available once they have a majority in the legislative branch.

Table 1.9: Opposition Effects: Flipping the Majority vs Not Flipping the Majority

Panel A: Marginal Opposition Legislator Flips Majority						
	CPI	Council Leader	Corruption	Infant Mort.	Preterm	Reel.
T	0.0949	0.526***	-1.218	-5.018**	-1.999*	-0.151**
	(0.0909)	(0.115)	(1.179)	(2.379)	(1.021)	(0.0742)
Observations	115	542	45	2868	3712	721
Panel B: Marginal Opposition Legislator *does not* Flip Majority						
	CPI	Council Leader	Corruption	Infant Mort.	Preterm	Reel.
T	0.0148	0.00368	-0.656	-1.956	-0.711	-0.00544
	(0.0797)	(0.0590)	(0.633)	(1.471)	(0.672)	(0.0395)
Observations	211	1958	147	8782	8911	2401

Standard errors in parentheses. Treatment indicates the effect of switching one seat in the legislature from 'pro-mayor' to 'anti-mayor'. Panel A contains estimates for such occasions when this switch flips the mayor's coalition from having a majority to having a minorty, and panel B contains estimates for all other occasions. 'Reel.' refers to reelection of the mayor.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[44] We also do the analysis seat-by-seat in the appendix in Figure 1.12, but we do not report the results here because estimates are highly imprecise. Nevertheless, the general pattern of the estimates support our findings here, that the seat that flips the majority is the important one.

1.6.6 Mechanisms

Finally, we test two of the main mechanisms of the model. First, revealing a mayor is corrupt reduces the reelection chances of allied legislators. To test this we use the same identification strategy of Ferraz and Finan (2008). They explore the fact that the timing of the public dissemination of the federal audit results is random. Some municipalities had the results of the audit reported before the municipal elections and some municipalities had the results published after the election took place. They use this variation to estimate the effect of exposing a corrupt mayor on their reelection chances. We use the same strategy, but we estimate the effect of exposing the mayor on the 2008 reelection chances of legislators that belonged to the mayor's coalition.

Also motivated by Ferraz and Finan (2008), we examine the interaction of audits with the presence of AM radio stations which disseminate knowledge about the results of the audits, and our indicator for corrupt indicates when 2 or more items were found to have corruption, since Ferraz and Finan's findings suggest that voters tolerate low levels of corruption, but not high levels.

The two specifications we use are these:

$$y_i = \beta_0 + \beta_1 PreAudit_i + \beta_2 Corrupt_i + \beta_3 PreAudit_i X Corrupt_i + X_i + \varepsilon_i$$

$$\begin{aligned} y_i = &\beta_0 + \beta_1 PreAudit_i + \beta_2 Corrupt_i + \beta_3 AMRadio_i + \beta_4 PreAudit_i X Corrupt_i \\ &+ \beta_5 AMRadio_i X Corrupt_i + \beta_6 PreAudit_i X AMRadio_i + \beta_7 PreAudit_i X AMRadio X Corru \\ &+ X_i + \varepsilon_i \end{aligned}$$

where X_i is a vector of municipal charactaristics.

In table 1.10 we find that when the mayor is revealed to be corrupt and there is a local AM radio station present to disseminate the audit report results, this

lowers reelection probability for mayor-allied legislators by about 10 percentage points.[45] This result shows that beyond punishing corrupt mayors (Ferraz and Finan, 2008), voters also punish legislators that are *allied* with corrupt mayors. Thus, city counselors that are allied with the mayor lack incentives to report and investigate corruption, because revealing this corruption would likely damage their own political careers.

Table 1.10: The effect of revealing that the mayor is corrupt

	Incumbent Mayor Allies Reelected	
Pre-Audit	-0.0286	-0.0587**
	(0.0258)	(0.0289)
Pre-AuditXCorrupt	0.0465	0.0844
	(0.0576)	(0.0599)
Pre-AuditX AMRadio		0.128**
		(0.0616)
Pre-AuditX AMRadio X Corrupt		-0.179**
		(0.0826)
Observations	1629	1629
Clusters	448	448

The outcome for both columns is an indicator for if a mayor-allied member of the city council is reelected.

Standard errors in parentheses, municipal charactaristics used as controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Finally, we also test if the presence of opposition legislators reduces the re-election chances of incumbent mayors using reelection results for the 2008, 2012 and 2016 mayoral elections. According to our political agency model, opposing

[45] This is $\beta_4 + \beta_7$

legislators will reduce reelection chances of mayors when reelection incentives are not strong enough to force corrupt mayors to pretend to be non-corrupt, even when they know that they will be reported by legislators and this will end up costing their reelection. Table 1.11 shows that an additional opposition legislator decreases mayors' reelection chances by 6.2 p.p. If we estimate the effect of an additional opposition legislator in municipalities where the incumbent mayor actually ran for office, there is a reduction of 9.2 p.p. on reelection chances. These results are robust to a variety of alternate specifications as shown in the appendix in section 1.10.1.

Table 1.11: Incumbent Mayor Reelection

	Reelection	Reel. Conditional on Running	Incumbent Vote Share
T	-0.0618*	-0.0926**	-0.0215
	(0.0360)	(0.0433)	(0.0163)
Optimal Bandwidth	0.011	0.010	0.010
Observations	5539	3985	3800

Effect of the marginal opposing legislator. CCT optimal bandwidth, uniform kernel, s.e. in parenthesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

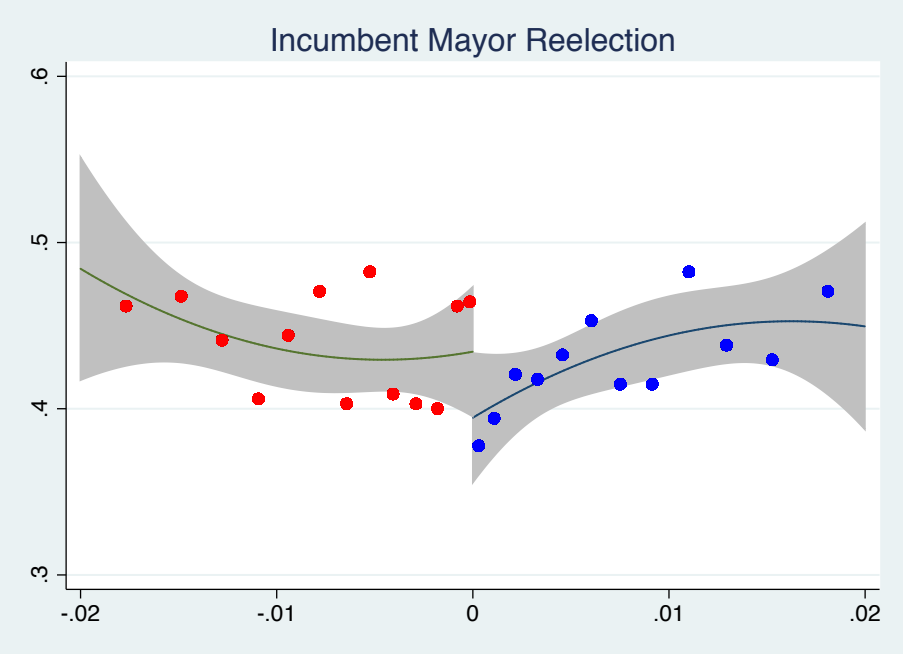


Figure 1.8: Probability of Mayor Reelection

Notes: Here we present the RD plot of the probability of the incumbent mayor getting reelected. This plot presents equally sized bins with means of the dependent variable inside each one. It also presents the prediction for the dependent variable from a regression of the dependent variable on the running variable and on the running variable squared and plots the resulting line along with a 95% confidence interval. For consistency, we use a 0.02 bandwidth for all RD plots, which is generally wider than the CCT optimal bandwidth used for coefficient estimates.

1.7 Discussion and Conclusion

Our estimates of the sizable effect of political opposition on corruption, public service delivery, and health outcomes, coupled with the fact that the typical mayor has 5 city councillors in his or her coalition (table 1.18) suggests that mayors depend significantly on having city councillors on their side in order to extract rents or exert low effort.

It is useful to benchmark the magnitudes of our estimated impacts on corruption with previous results in the literature. For example, Ferraz and Finan (2011) find that reelection incentives are associated with a 0.471 reduction in acts of corruption in Brazil and Bobonis et al (2016) find that announcing a municipality will be audited before an election reduces the number of reported corrupt violations in 1.43 in Puerto Rico. Hence, our results suggest that the increase in political opposition have an impact similar to the threat of audits.

It is also important to put these large impacts on infant mortality rate in a broader context. During our period of study infant mortality in Brazil was in fast decline (see Figure 1.11). Hence, it might be useful to benchmark our results with this broader trend and think of the impact of opposition legislators as an acceleration of this improvement process. Back of the envelope calculations show that our impact is equivalent to a 3 years acceleration in the improvement of infant mortality outcomes.[46] Finally, the magnitude of our estimated impact is also similar to the findings of Galiani et al (2005) that the privatization of water provision in Argentina reduced in a short period of time child mortality by up to 26 percent in poor areas.

This suggests that it may be favorable to have more legislators politically opposed to the mayor, but legislators are chosen by voters, not economists, so what can be done? It so happens that the D'Hondt Method used in Brazil (and many other countries) to apportion seats is known to disproportionately favor parties that are politically stronger. As can be seen in table 1.20, the typical mayor's coalition in Brazil is the largest coalition in terms of vote share. Thus, using an alternative method that does not disproportionately favor stronger parties may be a low-cost way to increase political opposition, decrease corruption, and improve public service provision.

We conduct counterfactual simulations using the Webster Method, the Danish Method, and the Huntington-Hill Method on Brazil's municipal electoral data from 2004 to 2016, and report results in tables 1.32 through 1.34 in appendix 1.10.4. The three methods only differ in the way that quotients are calculated.[47]

In the Huntington-Hill case for example, 83% of elections have no change in

[46] Between 1996 and 2016 infant mortality fell on average 0.62 a year and our estimated impact on overall mortality rate (educated plus non-educated mothers) is 1.9. Hence: $1.9/0.62 \approx 3$.

[47] The formulas for the three alternate methods are respectively $Q_s = \frac{V}{2s+1}$, $Q_s = \frac{V}{3s+1}$, and $Q_s = \frac{V}{\sqrt{s*(s+1)}}$.

the size of the mayor's coalition and in 4.4% of elections the mayor's coalition gains a member, *but* in 12.4% of cases the mayor loses a coalition member. So there is an 8% net increase in municipalities that would have an additional legislator politically opposed to the mayor[48].

Additionally, this paper emphasizes the importance of strengthening legislative oversight to combate corruption and promote the good use of public resources. Stapenhurst, Pelizzo, Olson, and von Trapp (2008) document that various tools of legislative oversight are common throughout the world, but they are not universal.[49] For example, in a sample of 39 countries— 25 of which are OECD countries— only 28% have a specialized budget research organization attached to the legislature.[50] The number is likely much lower in a sample of only developing countries. Our analysis suggests that strengthening legislative oversight—by ensuring that legislative branches have the power and capacity to audit executive spending and investigate wrongdoing— could be an effective alternative or complement to centralized audit programs.

Finally, our research gives important direction to anti-corruption auditing agencies throughout the world. Anti-corruption agencies are often resource-constrained and may not know how to best focus their efforts. Our research suggests that anti-corruption audits should focus on areas where the political opposition is weak. In places where the opposition is strong, there is lower corruption, and the legislature is capable of doing investigation. In places where the opposition is weak, however, there is more corruption, and less capacity to investigate it.

In conclusion, we use a regression discontinuity design to show that the marginal politically opposed legislator can have a large impact in reducing corruption as well as improving healthcare provision and actual health outcomes.

[48]With the caveat that these simulations don't take into account possible changes in strategic responses that voters and politicians could make in response to a change in voting methods.

[49]See table 1.1 specifically.

[50]Table 6.6

Finally, we address this work's implications for the methods of proportional representation that are most likely to encourage good governance, as well as the importance of legislative oversight to combat corruption.

1.8 References

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson."Democracy does cause growth." No. w20004. National Bureau of Economic Research, 2014.

Acemoglu, Daron, James A. Robinson, and Ragnar Torvik. "Why do voters dismantle checks and balances?" Review of Economic Studies 80, no. 3 (2013): 845-875.

Alesina, Alberto, and Howard Rosenthal. "A theory of divided government." Econometrica: Journal of the Econometric Society (1996): 1311-1341.

Alves, D., and C. Timmins (2003): "Social Exclusion and the Two-Tiered Healthcare System of Brazil," in Who's in and Who's out: Social Exclusion in Latin America, ed. by J. R. Behrman, A. G. Trujillo, and M. Székely. Washington, DC: Inter-American Development Bank. [425,441]

Alt, James E., and David D. Lassen. "Political and judicial checks on corruption: Evidence from American state governments." Economics & Politics 20.1 (2008): 33-61.

Anduiza, Eva, Aina Gallego, and Jordi Muñoz. "Turning a blind eye: Experimental evidence of partisan bias in attitudes toward corruption." Comparative Political Studies 46.12 (2013): 1664-1692.

Ames, Barry, Andy Baker, and Lucio R. Renno. "Split-ticket voting as the rule: Voters and permanent divided government in Brazil." Electoral Studies 28.1 (2009): 8-20.

Avis, Eric, Claudio Ferraz, and Frederico Finan. "Do Government Audits Re-

duce Corruption? Estimating the Impacts of Exposing Corrupt Politicians." Journal of Political Economy 126, no. 5 (2018): 1912-1964.

Besley, Timothy. "Principled agents?: The political economy of good government." Oxford University Press on Demand, 2006.

Besley, Timothy, and Masayuki Kudamatsu. "Health and democracy." American Economic Review 96, no. 2 (2006): 313-318.

Bobonis, Gustavo J., Luis R. Cámara Fuertes, and Rainer Schwabe. "Monitoring corruptible politicians." American Economic Review 106.8 (2016): 2371-2405.

Britto, Diogo, and Stefano Fiorin. "Corruption and legislature size: Evidence from brazil." Available at SSRN 2859940 (2016).

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. "The political resource curse." American Economic Review 103, no. 5 (2013): 1759-96.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. "Robust nonparametric confidence intervals for regression-discontinuity designs." Econometrica 82.6 (2014): 2295-2326.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. "Optimal data-driven regression discontinuity plots." Journal of the American Statistical Association 110.512 (2015): 1753-1769.

Cascio, Elizabeth U., and Ebonya Washington. "Valuing the vote: The redistribution of voting rights and state funds following the voting rights act of 1965." The Quarterly Journal of Economics 129.1 (2013): 379-433.

Cattaneo, Matias D., Brigham R. Frandsen, and Rocio Titiunik. "Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate." Journal of Causal Inference 3.1 (2015): 1-24.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. "Simple local polyno-

mial density estimators." Working Paper. Retrieved July 22, 2017

Cattaneo, Matias D., Rocio Titiunik, Gonzalo Vazquez-Bare, and Luke Keele. "Interpreting regression discontinuity designs with multiple cutoffs." The Journal of Politics 78, no. 4 (2016): 1229-1248.

Caughey, Devin, and Jasjeet S. Sekhon. "Elections and the regression discontinuity design: Lessons from close US House races, 1942-2008." Political Analysis (2011): 385-408.

Cerulli, Giovanni, Yingying Dong, Arthur Lewbel, and Alexander Poulsen. "Testing stability of regression discontinuity models." In Regression Discontinuity Designs: Theory and Applications, pp. 317-339. Emerald Publishing Limited, 2017.

Chizzotti, Antonio, et al. *O Combate à Corrupção nas Prefeituras do Brasil.* São Paulo: 24x7 Cultural, 2012. Online.

Dal Bó, Ernesto, and Martín A. Rossi. "Term length and the effort of politicians."The Review of Economic Studies 78.4 (2011): 1237-1263.

Dong, Yingying, and Arthur Lewbel. "Identifying the effect of changing the policy threshold in regression discontinuity models." Review of Economics and Statistics 97.5 (2015): 1081-1092.

Eggers, Andrew C. "Partisanship and electoral accountability: Evidence from the UK expenses scandal." Quarterly Journal of Political Science 9.4 (2014): 441-472.

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder Jr. "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, no. 1 (2015): 259-274.

Falcão, Jaqueline and Silvia Amorim. 'Médicos recusam empregos oferecidos por prefeituras e estados.' *O Globo* 10 Jan. 2016. Web. 1 Nov. 2017.

Ferraz, Claudio, and Frederico Finan. "Exposing corrupt politicians: the effects

of Brazil's publicly released audits on electoral outcomes." (2008).

Ferraz, Claudio, and Frederico Finan. "Motivating politicians: The impacts of monetary incentives on quality and performance." No. w14906. National Bureau of Economic Research, 2009.

Ferraz, Claudio, and Frederico Finan. "Electoral accountability and corruption: Evidence from the audits of local governments." The American Economic Review 101.4 (2011): 1274-1311.

Fiorina, Morris P. "An era of divided government." Developments in American politics. Palgrave, London, 1992. 324-354.

Folke, Olle. "Shades of brown and green: party effects in proportional election systems." Journal of the European Economic Association 12.5 (2014): 1361-1395.

Fujiwara, Thomas. "Voting technology, political responsiveness, and infant health: Evidence from Brazil." Econometrica 83.2 (2015): 423-464.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. "Water for life: The impact of the privatization of water services on child mortality." Journal of political economy 113.1 (2005): 83-120.

Gulzar, Saad, and Benjamin J. Pasquale. "Politicians, bureaucrats, and development: Evidence from India." American Political Science Review 111.1 (2017): 162-183.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. "Identification and estimation of treatment effects with a regression-discontinuity design." Econometrica 69.1 (2001): 201-209.

Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142.2 (2008): 615-635.

Krehbiel, Keith. "Pivotal Politics: A Theory of Pivotal Politics." (1998).

Kudamatsu, Masayuki. "Has democratization reduced infant mortality in sub-Saharan Africa? Evidence from micro data." Journal of the European

Economic Association 10.6 (2012): 1294-1317.

Lee, David S. "Randomized experiments from non-random selection in US House elections." Journal of Econometrics 142.2 (2008): 675-697.

Levinson, Daryl J., and Richard H. Pildes. "Separation of parties, not powers." Harvard Law Review (2006): 2311-2386.

Lichand, Guilherme, Marcos FM Lopes, and Marcelo C. Medeiros. "Is Corruption Good For Your Health?" Working Paper, 2016.

Lim, Claire SH. "Preferences and incentives of appointed and elected public officials: Evidence from state trial court judges." The American Economic Review 103.4 (2013): 1360-1397.

Litschig, Stephan, and Yves Zamboni. "Judicial presence and rent extraction." No. 1143. 2015.

Litschig, Stephan, and Yves Zamboni. "Audit risk and rent extraction: Evidence from a randomized evaluation in Brazil." Universitat Pompeu Fabra (2017).

McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test." Journal of econometrics 142.2 (2008): 698-714.

Madsen, Jakob B., Paul A. Raschky, and Ahmed Skali. "Does democracy drive income in the world, 1500-2000?." European Economic Review 78 (2015): 175-195.

Martinez-Bravo, Monica, Gerard Padró i Miquel, Nancy Qian, and Yang Yao. "Political reform in China: the effect of local elections." NBER working paper 18101 (2014).

Mayhew, David R. "Divided we govern." New Haven: Yale University, 1991.

Morais, Rogério. 'Prefeito Puranci é condenado a perda de mandato em processo aberto por dispensa de licitação.' *Grupo Fronteira Missões* 23 Oct. 2017. Web. 22 Nov. 2017.

Nishijima, Marislei, Randall Ellis, and Regina Cati. "Evaluating the impact of Brazil's central audit program on municipal provision of health services." Working Paper. 2016.

Nzelibe, Jide O., and Matthew C. Stephenson. "Complementary constraints: separation of powers, rational voting, and constitutional design." Harv. L. Rev. 123 (2009): 617.

Olken, Benjamin A. "Monitoring corruption: evidence from a field experiment in Indonesia." Journal of Political Economy 115.2 (2007): 200-249.

Olken, Benjamin A., and Rohini Pande. "Corruption in developing countries." Annu. Rev. Econ. 4.1 (2012): 479-509.

Persson, Torsten, Gerard Roland, and Guido Tabellini. "Separation of powers and political accountability." The Quarterly Journal of Economics 112.4 (1997): 1163-1202.

Persson, Torsten, Gerard Roland, and Guido Tabellini. "Comparative politics and public finance." Journal of political Economy 108, no. 6 (2000): 1121-1161.

Pettersson-Lidbom, Per. "Do Parties Matter for Economic Outcomes? A Regression Discontinuity Approach." Journal of the European Economic Association 6.5 (2008): 1037-1056.

'Postos de saúde novos e equipados não funcionam por falta de médicos.' *O Globo* 30 June 2013. Web. 1 Nov. 2017.

'Preterm Birth' World Health Organization, 9 Nov. 2016, `http://www.who.int/mediacentre/factsheets/fs363/en/`.

Rodrigues, Bárbara. 'Vereadores denunciam novamente o prefeito Gilson Filho ao TCE.' *GP1* 12 Nov. 2017. Web. 22 Nov. 2017.

Scheffer, M., A. Biancarelli, and A. Cassenote. 'Demografia médica no Brasil 2015.' Departamento de Medicina Preventiva, Faculdade de Medicina da USP. Conselho Regional de Medicina do Estado de São Paulo. São Paulo:

Conselho Federal de Medicina (2015).

Stapenhurst, Rick, Riccardo Pelizzo, David Olson, and Lisa von Trapp, eds. "Legislative oversight and budgeting: A world perspective." The World Bank, 2008.

'Vereador denuncia: prefeito contrata empresa investigada pelo próprio governo' *Câmara Municipal de Pelotas* Web. 22 Nov. 2017.

'Vereador denuncia prefeito de Ivinhema por suposto esquema de desvio de dinheiro.' YouTube, uploaded by Jornal da Nova, 19 Sept 2017, `https://www.youtube.com/watch?v=i6i_ScTR790`.

Vintzileos, Anthony M., Cande V. Ananth, John C. Smulian, William E. Scorza, and Robert A. Knuppel. "The impact of prenatal care in the United States on preterm births in the presence and absence of antenatal high-risk conditions." American journal of obstetrics and gynecology 187, no. 5 (2002): 1254-1257.

1.9 Appendix

1.9.1 Additional Summary Statistics

Table 1.12: Election Data Summary Statistics

	RD Sample	Non-RD Sample	Diff.
Tot. City Council Seats	9.311	9.344	-0.033
	(1.963)	(1.562)	(0.053)
Second Term	0.279	0.315	-0.037^{**}
	(0.448)	(0.465)	(0.015)
Mayor's Coalitions's seats	4.767	5.35	-0.583^{***}
	(1.548)	(2.319)	(0.069)
Num. Coalitions in Municipality	4.654	5.12	-0.466^{***}
	(2.5)	(2.576)	(0.082)
Num. Coalitions Supporting Mayor	1.852	2.165	-0.313^{***}
	(1.105)	(1.316)	(0.041)
Mayor Coalition Rank	1.462	1.449	0.013
	(0.818)	(0.805)	(0.026)
Anti-Mayor Coalition Vote Share	50.688	45.62	5.068^{***}
	(14.251)	(21.451)	(0.64)
Opposition Seat Share	0.484	0.425	0.059^{***}
	(0.155)	(0.236)	(0.007)
T	0.505	.	.
	(0.5)	.	.
R	0.001	.	.
	(0.011)	.	.
	4192	1275	.

Full sample for 2004 election year (the year corresponding to most of the audit data). Means are presented in the main row, standard deviations/errors are below in parenthesis.

Table 1.13: Municipal Characteristics

	RD Sample	Non-RD Sample	Diff.
AM Radio	0.213	0.236	-0.023^{*}
	(0.409)	(0.425)	(0.013)
TV Station	0.102	0.127	-0.026^{***}
	(0.302)	(0.334)	(0.01)
Judiciary District	0.472	0.492	-0.019
	(0.499)	(0.5)	(0.016)
Avg. Monthly Income	433.133	446.383	-13.25^{**}
	(196.475)	(202.788)	(6.447)
Illiteracy Rate	16.36	15.69	0.67^{**}
	(9.862)	(9.812)	(0.315)
Urbanization Rate	63.263	65.589	-2.325^{***}
	(21.978)	(22.171)	(0.708)
Population 2010	34135.8	34273.96	-138.153
	(224619.0)	(97499.66)	(4420.171)
Tot. Votes Cast	17692.5	17334.42	358.079
	(116821.2)	(49890.38)	(2282.044)
Observations	4192	1275	.

Full sample for 2004 election year (the year corresponding to most of the audit data). Means are presented in the main row, standard deviations/errors are below in parenthesis.

Table 1.14: Conditional Summary Statistics

	count	mean	sd	min	max
Total Audit Service Items	432	25.38657	10.13098	9	95
Total Corruption	432	1.479167	1.916961	0	13
Embezzlement	432	.4606481	.984015	0	8
Fraud	432	.5324074	1.149545	0	8
Overinvoicing	432	.5972222	1.287019	0	12
Other Irreg.	432	4.118056	3.786954	0	30

RD Sample, conditional on corruption being found

Table 1.15: Public Service Assessment Questions

Variable	Survey Questions
Nurse Present	*When you needed to be seen at the Family Health Unit, was there a nurse there to serve you?*
Dentist Present	*When you needed to be seen at the Family Health Unit that has a dentist, were you served?*
Physician Present	*When you needed to be seen at the Family Health Unit, was there a physician there to serve you?*
Healthcare Visits	*Does the family recieve visits from Community Health Agents?*
Lines at Health Unit	*Have you or someone in your family had to wait in lines to recieve care?*
	Auditor's Assessment
Irregular Hiring	*Are there Community Health Agents that were contracted irregularly?*

	Total City Council Seats	
	freq	pct
9	778	88.41
10	42	4.77
11	28	3.18
12	5	0.57
13	12	1.36
14	4	0.45
15	5	0.57
16	1	0.11
18	2	0.23
19	3	0.34
Total	880	100.00

Table 1.16

	Num. Coalitions in Municipality	
	freq	pct
2	190	21.59
3	185	21.02
4	142	16.14
5	137	15.57
6	85	9.66
7	54	6.14
8	38	4.32
9	19	2.16
10	8	0.91
11	11	1.25
12	4	0.45
13	4	0.45
15	3	0.34
Total	880	100.00

Table 1.17

	Mayor's Coalition Rank (In Vote Share)	
	freq	pct
1	591	67.16
2	231	26.25
3	41	4.66
4	11	1.25
5	3	0.34
6	2	0.23
8	1	0.11
Total	880	100.00

Table 1.20

	Mayor's Coalition's Seats	
	freq	pct
0	4	0.45
1	7	0.80
2	39	4.43
3	95	10.80
4	177	20.11
5	273	31.02
6	174	19.77
7	73	8.30
8	25	2.84
9	8	0.91
10	3	0.34
11	1	0.11
12	1	0.11
Total	880	100.00

Table 1.18

	Num. Coalitions Supporting Mayor	
	freq	pct
1	444	50.45
2	253	28.75
3	124	14.09
4	44	5.00
5	11	1.25
6	3	0.34
7	1	0.11
Total	880	100.00

Table 1.19

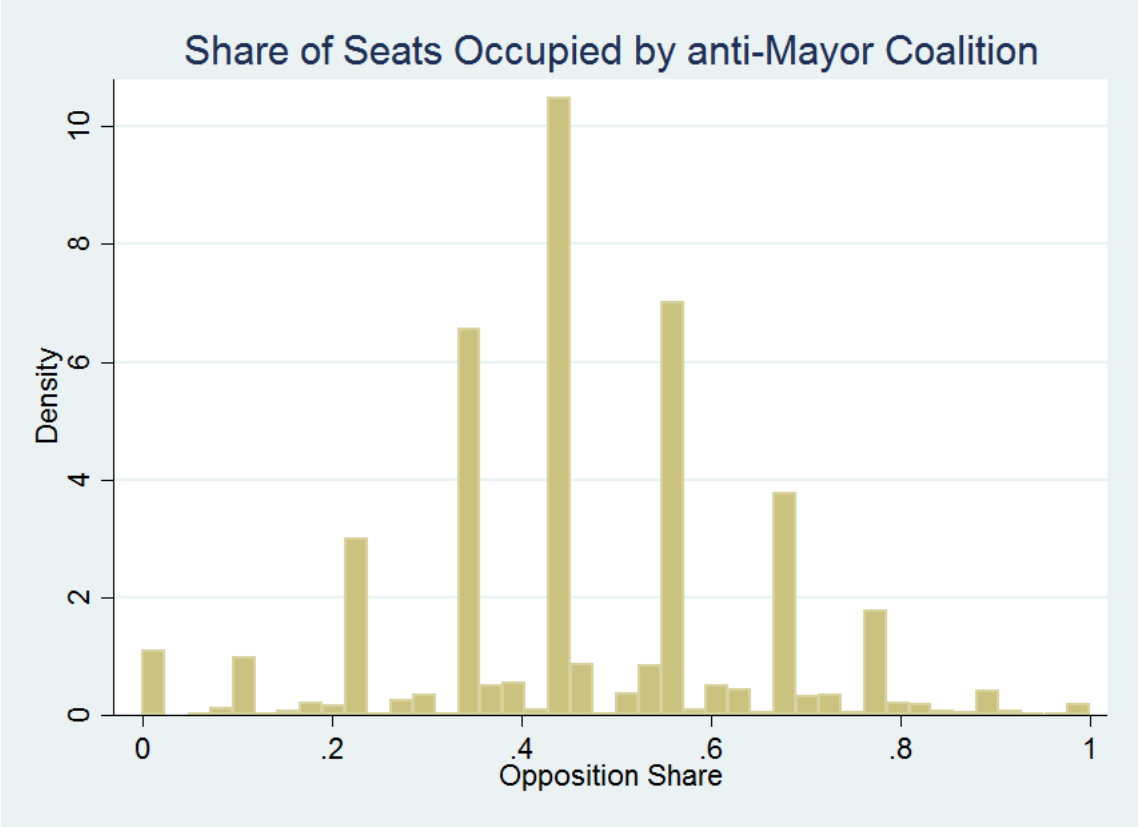


Figure 1.9: The density of opposition seat shares. The density clusters at levels that are fractions of 9 because the majority of municipalities have 9 seats.

Figure 1.10 illustrates somewhat the variation from which we are identifying the treatment effect. Particularly, it shows how this estimator can be seen as a weighted average at various levels of voter preference, because we have data close to the cutoff for a variety of total anti-mayor coalition vote shares.

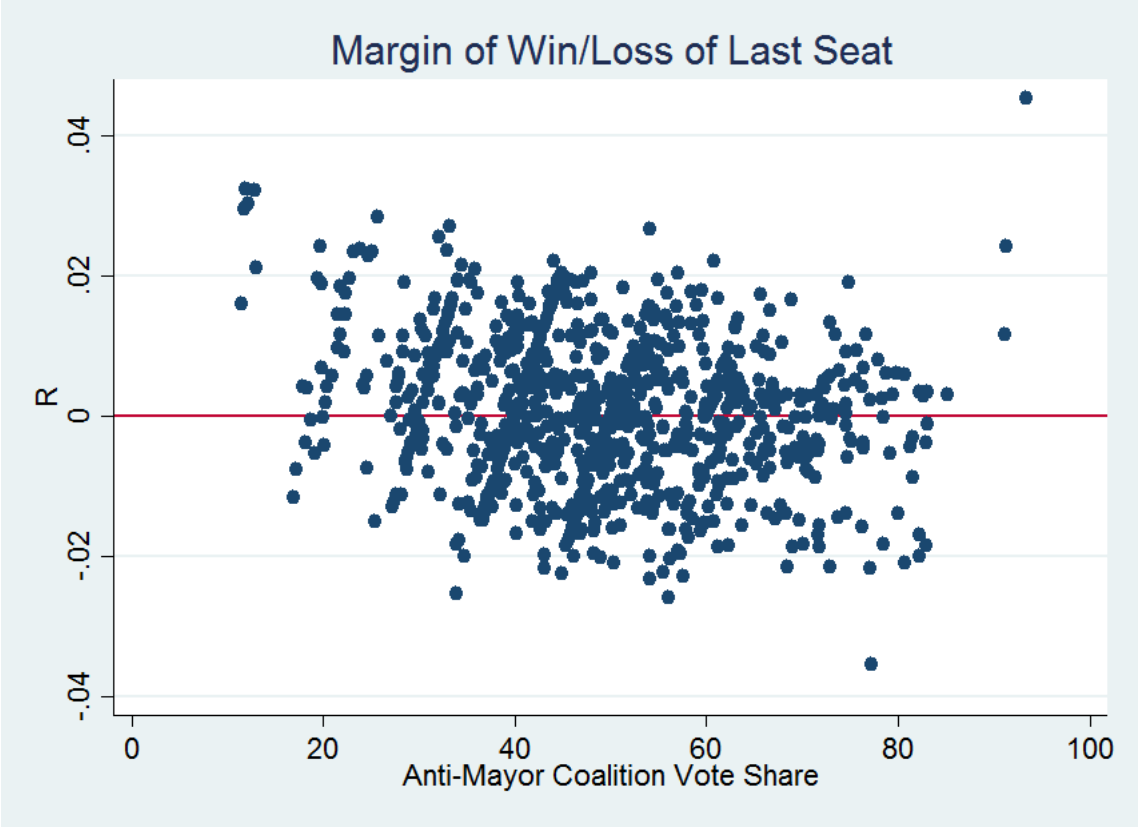


Figure 1.10: This figure illustrates that the estimator is a weighted average of the treatment effect at a variety of thresholds (voter preference levels). The faint pattern of upward-sloping lines comes from municipalities where there were only 2 coalitions, in which case vote share maps 1-to-1 into the running variable for a given disputed seat in the legislature.

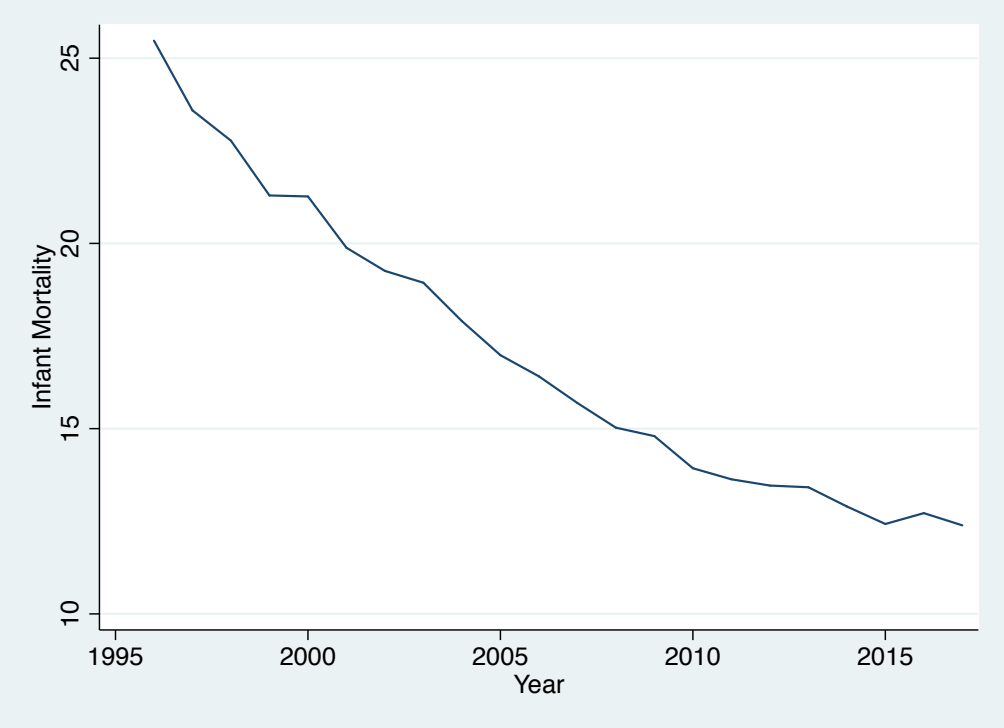


Figure 1.11: Infant Mortality in Brazil

1.9.2 Additional Results

In the model shown in 2.3.1 we use conventional standard errors. Heteroskedasticity-robust standard errors are biased in small samples, so we conduct the Breusch-Pagan test to test for possible heteroskedasticity. We fail to reject the null hypothesis of homoskedasticiy for all of our main outcome variables, with all F-statistics arbitrarily close to zero. This is unsurprising, since due to the RD design, we do not expect variance of the error term to be different between municipalities where the anti-mayor coalition barely won an additional seat versus municipalities where they barely didn't win an additional seat. Results available upon request.

Below we report results from Calonico, Cattaneo and Titiunik's RD robust estimator. Their 'conventional' standard errors are heteroskedasticity robust standard errors.

Tables 1.25 through 1.26 show results for all mayors, rather than just first term mayors

Table 1.21: CCT Robust Estimation

	Total Corruption	Embezzlement
Conventional	-1.353**	-0.889*
	(0.566)	(0.471)
Bias-corrected	-1.464***	-0.960**
	(0.566)	(0.471)
Robust	-1.464**	-0.960*
	(0.628)	(0.561)
Observations	552	552

Treatment effect using CCT robust confidence intervals.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.22: CCT Robust Estimation

	Physician Present	Dentist Present
Conventional	0.133	0.291*
	(0.141)	(0.155)
Bias-corrected	0.167	0.320**
	(0.141)	(0.155)
Robust	0.167	0.320*
	(0.159)	(0.174)
Observations	3854	3150

Treatment effect using CCT robust confidence intervals.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.23: CCT Robust Estimation

	Preterm	Infant Mort. Rate
Conventional	-0.824*	-3.402***
	(0.429)	(1.249)
Bias-corrected	-0.851**	-3.559***
	(0.429)	(1.249)
Robust	-0.851*	-3.559***
	(0.480)	(1.342)
Observations	30725	30726

Treatment effect using CCT robust confidence intervals.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.24: Corruption, Extensive Margin (1/0)

	Total Corruption	Embezzlement	Fraud	Overinvoicing	Other Irreg.
T	-0.330*	-0.101	-0.139	-0.0652	-0.0471
	(0.177)	(0.123)	(0.117)	(0.122)	(0.133)
Obs	253	282	336	324	274

Treatment effect of the marginal opposing legislator on corruption, using the CCT optimal bandwidth and a uniform kernel. Standard errors in parenthesis. Outcomes are the count of items audited with a major violation. Other irregularities captures all other irregularities found by auditors that were not classified as embezzlement, fraud, or overinvoicing, and includes a host of procedural and other errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.25: Corruption Outcomes, First and Second Term Mayors

	Total Corruption	Embezzlement	Fraud	Overinvoicing	Other Irreg.
T	-0.876*	-0.642**	-0.271	-0.106	0.235
	(0.474)	(0.256)	(0.259)	(0.293)	(0.911)
Obs	381	342	402	399	505

Treatment effect of the marginal opposing legislator on corruption, using the CCT optimal bandwidth and a uniform kernel. Standard errors in parenthesis. Outcomes are the count of items audited with a major violation. Other irregularities captures all other irregularities found by auditors that were not classified as embezzlement, fraud, or overinvoicing, and includes a host of procedural and other errors.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.26: Public Health Outcomes, First and Second Term Mayors

Panel A: Uneducated Mothers		
	Preterm	Infant Mort. Rate
T	-0.547	-2.324**
	(0.340)	(0.983)
Observations	25297	27702
Clusters	4259	4278
Panel B: Educated Mothers		
	Preterm	Infant Mort. Rate
T	-0.275	-0.346
	(0.250)	(0.538)
Observations	25403	25673
Clusters	4259	4278

Treatment effect of the marginal opposing legislator on public health outcomes, using the CCT optimal bandwidth and a uniform kernel. Clustered standard errors in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Our seat-by-seat analysis is contained in figure 1.12, which plots the estimated treatment effects with 95% confidence intervals for our main outcomes of interest for each seat which we have enough data to get reasonable estimates. The coefficients estimated are imprecise due to small sample sizes for some seats, but a faint pattern emerges from the plots. It is usually the case the case that the strongest effects appear when the opposition captures the 5th seat and gains a majority.

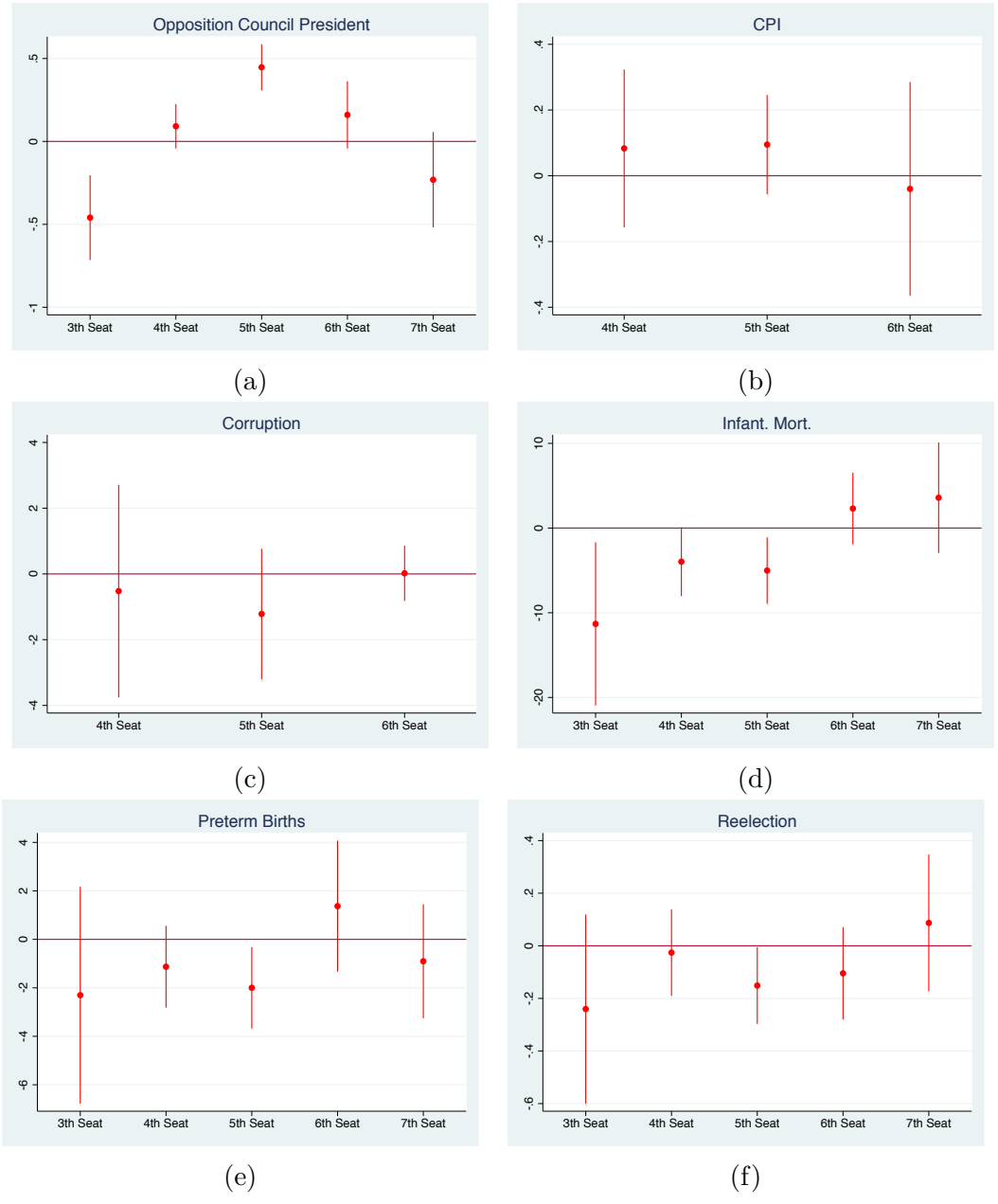


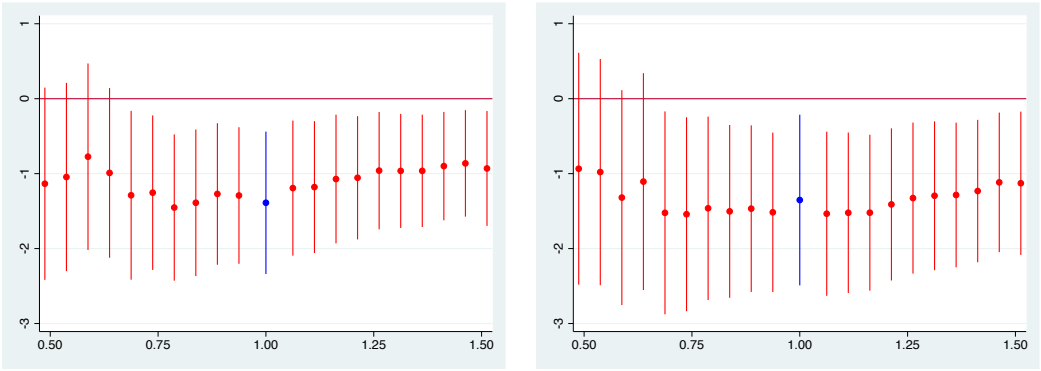
Figure 1.12: **Seat by Seat Effects**

1.10 Robustness Checks

1.10.1 BW Sensitivity Exercise

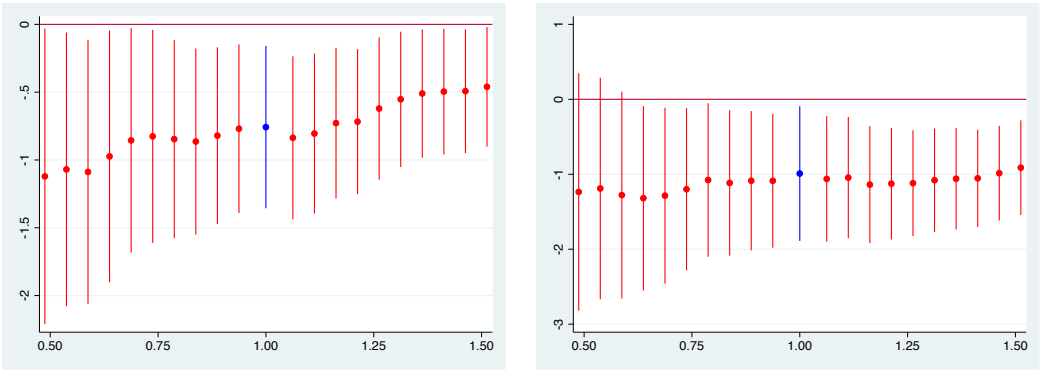
The following plots show the estimated treatment effects with 95% confidence intervals from a variety of alternate model specifications for outcome variables of interest, including various bandwidth selections and polynomial specifications. In the plots, 1.0 is the optimal BW selected by the Calonico et al. (2015) procedure and 0.5, for example, is half the optimal BW.

Figure 1.13: **BW and Polynomial Sensitivity: Corruption Violations**



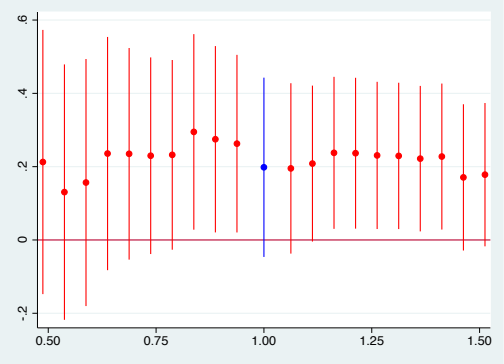
(a) First Degree Polynomial (b) Second Degree Polynomial

Figure 1.14: **BW and Polynomial Sensitivity: Embezzlement Violations**

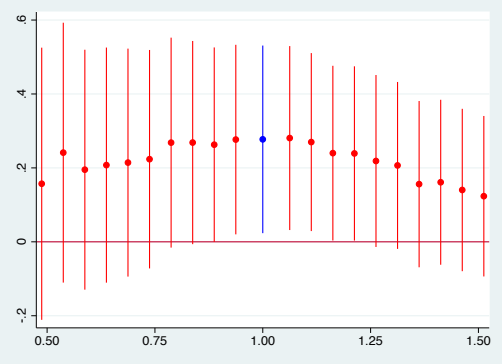


(a) First Degree Polynomial (b) Second Degree Polynomial

Figure 1.15: **BW and Polynomial Sensitivity: Physician Present**

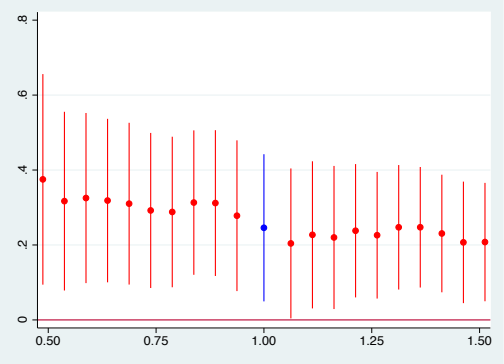


(a) First Degree Polynomial

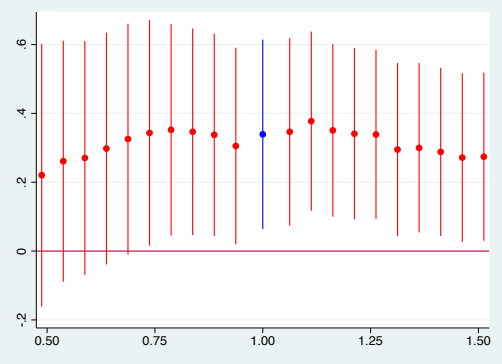


(b) Second Degree Polynomial

Figure 1.16: **BW and Polynomial Sensitivity: Dentist Present**

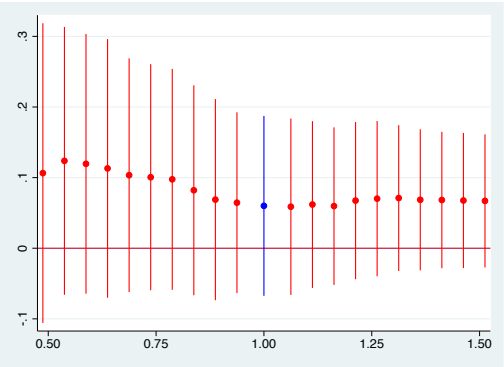


(a) First Degree Polynomial

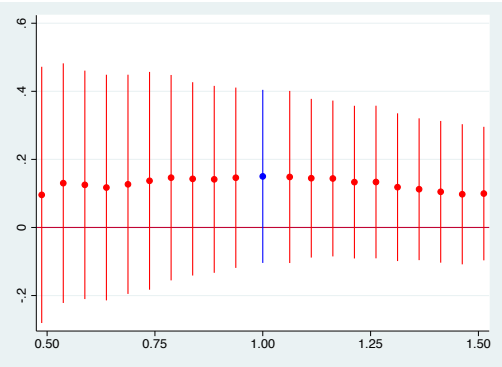


(b) Second Degree Polynomial

Figure 1.17: **BW and Polynomial Sensitivity: Nurse Present**

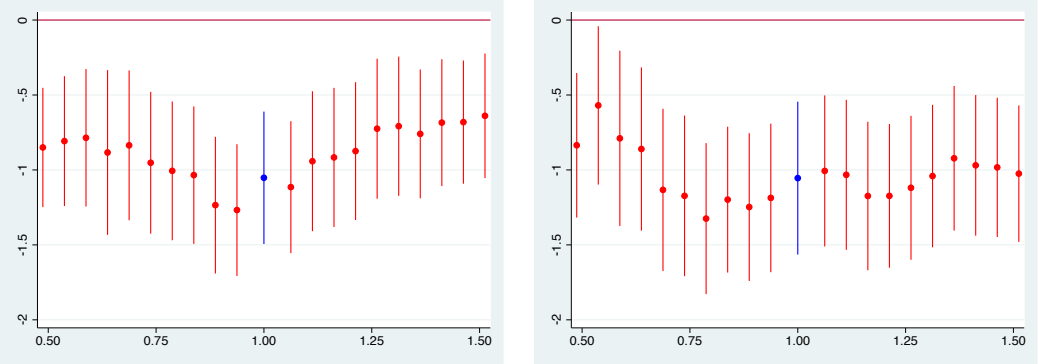


(a) First Degree Polynomial



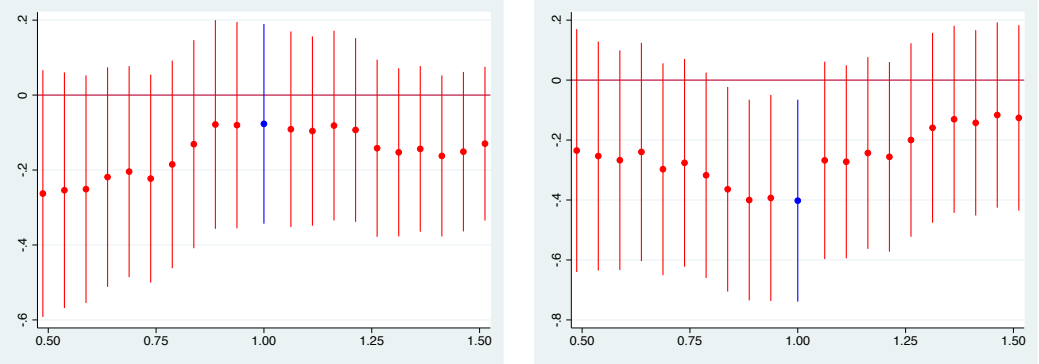
(b) Second Degree Polynomial

Figure 1.18: **BW and Polynomial Sensitivity: Irregular Hirings**



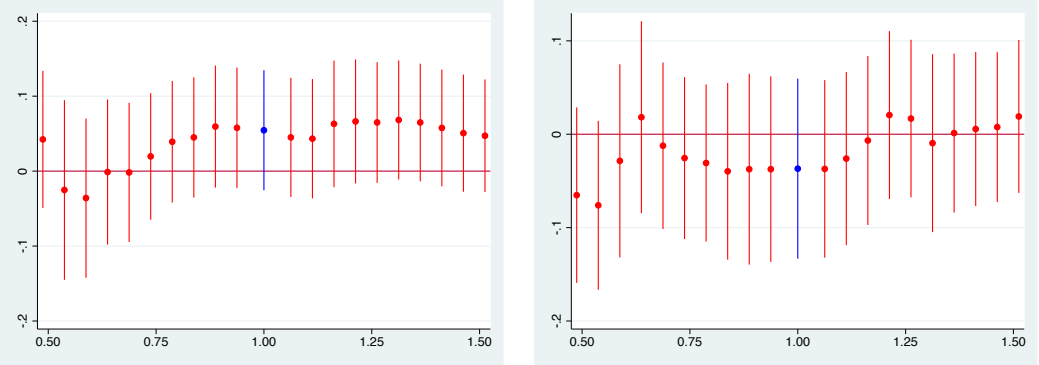
(a) First Degree Polynomial (b) Second Degree Polynomial

Figure 1.19: **BW and Polynomial Sensitivity: Lines at Health Unit**



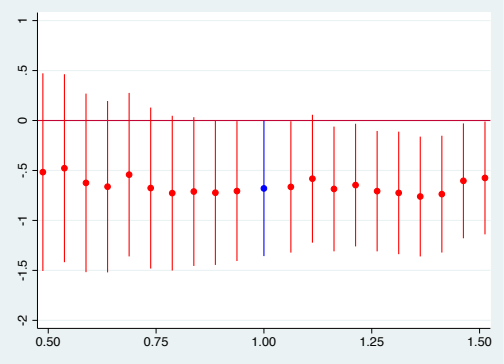
(a) First Degree Polynomial (b) Second Degree Polynomial

Figure 1.20: **BW and Polynomial Sensitivity: Healthcare Visits**

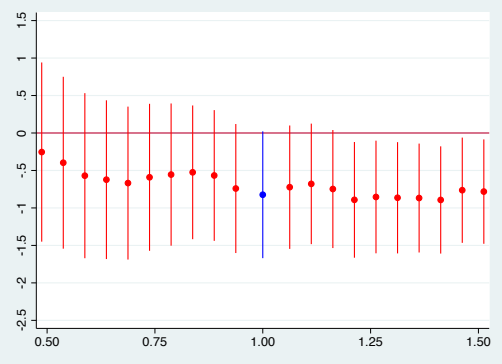


(a) First Degree Polynomial (b) Second Degree Polynomial

Figure 1.21: **BW and Polynomial Sensitivity: Preterm Births Uneducated Mothers**

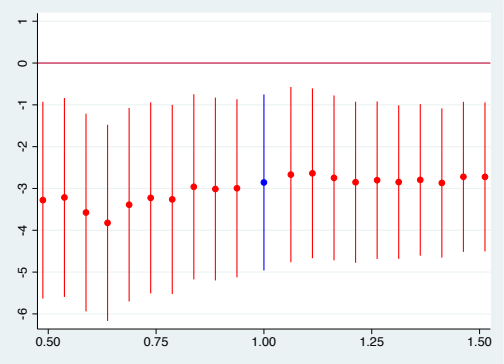


(a) First Degree Polynomial

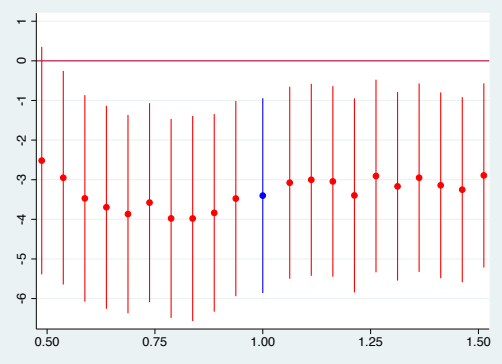


(b) Second Degree Polynomial

Figure 1.22: **BW and Polynomial Sensitivity: Infant Mortality Uneducated Mothers**

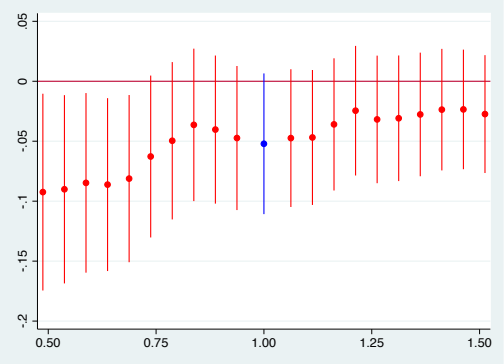


(a) First Degree Polynomial

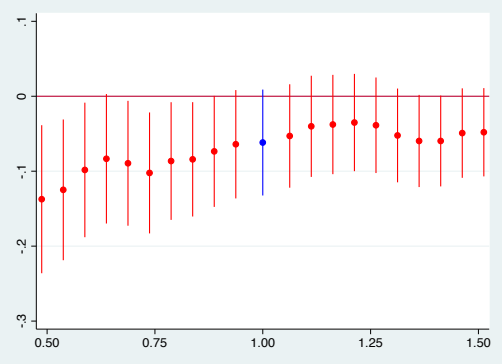


(b) Second Degree Polynomial

Figure 1.23: **BW and Polynomial Sensitivity: Mayor Reelection**



(a) First Degree Polynomial



(b) Second Degree Polynomial

1.10.2 Stability (External Validity) Test

Because regression discontinuity estimators estimate the treatment effect only for a given value of the running variable, we test the stability of our RD estimates using the method put forward by Dong and Lewbel (2015) and Cerulli et al. (2017), which uses the treatment effect derivative (TED) as a measure of the stability of our treatment effect estimates. Intuitively, this method tests what would happen to treatment effect estimates if the RD cuttoff were changed. Cerulli et al. suggest a rule of thumb for treatment effect stability, which is that if the TED is statistically significant and the relative TED is below approximately 1, estimates may be unstable, or likely to change if the RD cutoff were changed. We perform this test and find that our estimates of the effect on corruption and the infant mortality rate are stable, while our estimates for the effect of treatment on the presence of dentists and physicians shows signs of instability. Cerulli et al. point out that 'instability does not mean that the RD estimates are invalid, but rather that they need to be interpreted cautiously. In contrast, a finding of stability (i.e., a small TED) suggests some external validity, since it implies some other people, those away from but near the cutoff, likely have treatment effects of similar magnitudes to those right at the cutoff.'

Table 1.27: Stability (External Validity) Test

	Tot. Corruption	Dentist	Physician	Inf. Mort.
Treatment	-1.353**	0.339**	0.277**	-3.402***
	(0.578)	(0.139)	(0.128)	(1.254)
TED	-223.9	156.9**	133.8**	-434.1
	(437.6)	(74.06)	(63.29)	(479.8)
Bandwidth	0.006	0.008	0.009	0.010
Relative TED	0.995	0.258	0.236	0.757
Observations	244	2233	2751	19819
Clusters		121	130	3945

Using the test set forth in Dong and Lewbel (2015) and Cerulli et al. (2017)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.10.3 Balance Tests

Table 1.28: Balance Tests, Eq. 2.3.1 (Corruption Outcomes)

Marginal Legislator Characteristics				
	Male	College	High School	Coalition Size
T	-0.00119	-0.0654	-0.0255	0.00726
	(0.0846)	(0.107)	(0.137)	(0.395)
Obs	484	406	428	382
	Native Mun.	Native State	Mayor Age	Mayor Win Margin
T	0.0102	-0.0676	-0.946	-0.0347
	(0.124)	(0.0825)	(2.508)	(0.0370)
Obs	494	469	457	464
Municipality Characteristics				
	AM Radio	TV Station	Internet Prov.	Metro
T	0.0122	0.0239	-0.207	0.00116
	(0.126)	(0.0761)	(0.138)	(0.0681)
Obs	389	457	417	421
	Judiciary	Illiteracy Rate	Med. Month Inc.	Bookstore
T	0.0154	-0.839	-7.715	0.150
	(0.123)	(2.585)	(35.90)	(0.127)
Obs	497	479	466	417
	Urban. Rate	Pct. < 25 yrs	Council Wage	Hrs Worked
T	-7.514	1.222	-123.3	-0.259
	(5.629)	(1.787)	(277.5)	(1.879)
Obs	488	428	497	481
	GDP per Cap	GDP per Cap Growth	Population	Pop. Growth
T	-2.361	-0.0120	-14809.1	-0.00616
	(3.153)	(0.0362)	(9838.8)	(0.00902)
Obs	416	427	524	394
Party Seat Shares				
	PMDB Share	PSDB Share	PFL Share	PP Share
T	0.0106	0.0218	0.0210	0.0258
	(0.0424)	(0.0370)	(0.0291)	(0.0315)
Obs	427	434	366	491
	PT Share	PTB Share	PDT Share	Other Share
T	-0.0434	0.00877	-0.0782***	0.0460
	(0.0279)	(0.0314)	(0.0261)	(0.0553)
Obs	427	410	417	510

Effect of the marginal opposing legislator. CCT optimal bandwidth, uniform kernel, s.e. in parenthesis. The characteristics in the first panel correspond to the last legislator who won a seat in the city council, except for the last two, which correspond to the mayor.

Table 1.29: Balance Tests, Eq. 2.3.2 (Survey Outcomes)

Marginal Legislator Characteristics				
	Male	College	High School	Coalition Size
T	0.149	-0.156	0.182	0.0509
	(0.161)	(0.200)	(0.246)	(0.727)
Obs	182	152	160	141
	Native Mun.	Native State	Mayor Age	Mayor Win Margin
T	-0.224	-0.394**	-4.167	-0.0462
	(0.220)	(0.158)	(4.329)	(0.0797)
Obs	186	176	173	171
Municipality Characteristics				
	AM Radio	TV Station	Internet Prov.	Metro
T	0.0898	0.137	0.0928	0.00910
	(0.215)	(0.123)	(0.250)	(0.112)
Obs	143	173	156	165
	Judiciary	Illiteracy Rate	Med. Month Inc.	Bookstore
T	0.164	-2.276	-29.21	0.304
	(0.218)	(4.692)	(62.38)	(0.235)
Obs	189	178	176	156
	Urban. Rate	Pct. < 25 yrs	Council Wage	Hrs Worked
T	-9.817	3.097	274.0	0.923
	(9.670)	(3.288)	(505.6)	(3.026)
Obs	183	160	189	180
	GDP per Cap	GDP per Cap Growth	Population	Pop. Growth
T	-1.079	0.00674	1771.0	-0.0161
	(3.854)	(0.0694)	(7517.3)	(0.0116)
Obs	155	159	199	145
Party Seat Shares				
	PMDB Share	PSDB Share	PFL Share	PP Share
T	0.00353	0.0964	0.0307	-0.0745
	(0.0784)	(0.0750)	(0.0541)	(0.0534)
Obs	158	160	131	183
	PT Share	PTB Share	PDT Share	Other Share
T	-0.00613	-0.0101	-0.0325	0.00637
	(0.0507)	(0.0584)	(0.0440)	(0.0972)
Obs	158	152	154	192

Effect of the marginal opposing legislator. CCT optimal bandwidth, uniform kernel, s.e. in parenthesis. The characteristics in the first panel correspond to the last legislator who won a seat in the city council, except for the last two, which correspond to the mayor.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.30: Balance Tests, Eq. 2.3.3 (Health Outcomes)

Marginal Legislator Characteristics				
	Male	College	High School	Coalition Size
T	0.0155	-0.0547*	-0.0169	-0.344***
	(0.0215)	(0.0289)	(0.0340)	(0.0993)
Obs	7368	6292	6540	5751
Clust	4735	4325	4421	4078
	Native Mun.	Native State	Mayor Age	Mayor Win Margin
T	0.0136	0.0227	0.227	0.0136
	(0.0318)	(0.0218)	(0.650)	(0.00840)
Obs	7305	7070	6879	7159
Clust	4712	4629	4549	4668
Municipality Characteristics				
	AM Radio	TV Station	Internet Prov.	Metro
T	0.0363	0.0136	-0.0332	-0.0170
	(0.0320)	(0.0235)	(0.0341)	(0.0193)
Obs	6057	6854	6364	6494
Clust	4232	4534	4356	4171
	Judiciary	Illiteracy Rate	Med. Month Inc.	Bookstore
T	0.0291	0.265	-5.067	0.0567*
	(0.0308)	(0.617)	(9.718)	(0.0325)
Obs	7430	7259	6996	6516
Clust	4756	4692	4599	4411
	Urban. Rate	Pct. < 25 yrs	Council Wage	Hrs Worked
T	-1.118	-0.500	41.23	-0.112
	(1.412)	(0.433)	(85.35)	(0.508)
Obs	7303	6518	7533	7160
Clust	4709	4410	4731	4565
	GDP per Cap	GDP per Cap Growth	Population	Pop. Growth
T	-0.861	0.0000621	17675.4	0.0000647
	(0.833)	(0.00924)	(20618.5)	(0.00200)
Obs	6427	6445	7766	6037
Clust	4368	4375	4831	4204
Party Seat Shares				
	PMDB Share	PSDB Share	PFL Share	PP Share
T	-0.0152	0.00399	0.0144**	0.00193
	(0.00969)	(0.00835)	(0.00701)	(0.00746)
Obs	6745	6803	5889	7402
Clust	4495	4512	4146	4742
	PT Share	PTB Share	PDT Share	Other Share
T	0.00221	-0.00523	-0.0137*	0.0272*
	(0.00677)	(0.00680)	(0.00717)	(0.0144)
Obs	6548	6343	6362	7649
Clust	4424	4349	4355	4820

Effect of the marginal opposing legislator. CCT optimal bandwidth, uniform kernel, s.e. in parenthesis. The characteristics in the first panel correspond to the last legislator who won a seat in the city council, except for the last two, which correspond to the mayor.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.31: Public Health Outcomes, First Term Mayors, With Controls

Panel A: Uneducated Mothers				
	Low Birth Weight	7+ P.N. Visits	Preterm	Infant Mort. Rate
T	-0.613*	-1.001	-0.841**	-2.609*
	(0.320)	(1.399)	(0.347)	(1.414)
Observations	14420	14741	17604	17135
Clusters	2988	3039	3443	3376
Panel B: Educated Mothers				
	Low Birth Weight	7+ P.N. Visits	Preterm	Infant Mort. Rate
T	-0.153	-0.0769	-0.344	-0.363
	(0.196)	(1.201)	(0.279)	(0.760)
Observations	14828	14371	14136	13020
Clusters	3047	2978	2942	2778

Treatment effect of the marginal opposing legislator on public health outcomes, using the CCT optimal bandwidth and a uniform kernel. Clustered standard errors in parenthesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.10.4 Counterfactuals

Webster Method		
	Freq.	Pct.
-2	8	0.05
-1	1627	10.15
0	13794	86.02
1	607	3.79
Total	16036	100.00

Difference in mayor's coalition size from using Webster's Method rather than the D'Hondt Method.

Table 1.32

Danish Method		
	Freq.	Pct.
-2	19	0.12
-1	2227	13.89
0	13019	81.19
1	769	4.80
2	2	0.01
Total	16036	100.00

Difference in mayor's coalition size from using the Danish Method rather than the D'Hondt Method.

Table 1.33

Huntington-Hill Method		
	Freq.	Pct.
-2	16	0.10
-1	1989	12.40
0	13317	83.04
1	712	4.44
2	2	0.01
Total	16036	100.00

Difference in mayor's coalition size from using the Huntington-Hill Method rather than the D'Hondt Method.

Table 1.34

1.11 Regular Expressions Used to Measure Corruption

Each violation found by the CGU auditors includes a description of the irregularity. We use regular expressions to search for words and phrases that isolate an irregularity as a certain type of corruption. We arrived at these words and phrases after personally reading through audit reports and taking note of the language used by the auditors. We have inspected a sample of irregularities flagged by these regular expressions and found that they were indeed corruption.

Fraud: Any irregularites containing

- *simulação* OR *simulado(a)* OR *simulações* ['simulation' OR 'simulated' OR 'simulations' (of the bidding process)]
- *montagem* ['assemblage/rigging' (of the bidding process)]
- *fraude* OR *fraudulento* OR *fraudar* ['fraud' OR 'fraudulent']
- *fachada* OR *fantasma* ['façade' OR 'phantom' (referring to fake firms)]

Overinvoicing: Any irregularites containing

- *superfatura* ['overinvoice']
- *sobrepreço* ['overprice']
- *preço superior* OR *preços superiores* ['higher price' OR 'higher prices']

Embezzlement: Any irregularites containing

- *falta de comprovação/comprovante* AND (*pagamento* OR *despesa* OR *aplicação*)

  ['lack of proof/receipt' AND ('payment' OR 'expenditure' OR 'aplication')]

- *não comprovação* AND (*despesa* OR *aplicação*)

  ['no proof/receipt' AND ('expenditure' OR 'aplication')]

- (*pagamento* OR *despesa* OR *aplicação*) AND (*sem comprovação/comprovante* OR *sem documentação comprobatório*)

  [('payment' OR 'expenditure' OR 'application' (of resources)) AND ('without proof/receipt' or 'without supporting documentation')]

- *não apresentação* AND (*documentos comprobatórios* OR *comprovação/comprovante*) AND (*pagamento* OR *despesa* OR *aplicação*)

  ['no presentation' AND ('proving documents' OR 'proof/receipt') AND ('payment' OR 'expenditure' OR 'application')]

- (*pagamento* OR *despesa* OR *aplicação*) AND (*sem documento fiscal* OR *sem suporte documental*)

  [('payment' OR 'expenditure' OR 'application') AND ('without fiscal documents' OR 'without documental support')]

- (*fiscal* OR *fiscais*) AND (*falsa* OR *inidônea* OR *fria*)

  [('fiscal' OR 'fiscals') AND ('false' OR 'disreputable' OR 'cold')] ('cold notes' in portuguese are false fiscal notes)

- *utilização* AND *recursos* AND *sem* AND *comprovação/comprovante* AND *despesa*

  ['utilization' AND 'resources' AND 'without' AND 'proof/receipt' AND 'expenditure']

- *ausência* AND *comprovação/comprovante* AND (*despesa* OR *pagamento*) AND length of description less than 75 characters

['absence' AND 'proof/receipt' AND ('expenditure' OR 'payment') AND length of description less than 75 characters]

Note: The character restriction on the last bullet point is because there were some irregularities with very long descriptions that included all of the chosen words, but on inspection, were clearly not corruption.

Chapter 2

Decomposing the Urban Wage Premium in Brazil: Firms, Matching, and Compensating Wage Differentials

2.1 Introduction

It is well documented that there is considerable heterogeneity in wages for workers with similar skills and experience. One literature has documented significant pay differences between rural and urban places (Glaeser and Maré 2001), and sought to understand why such pay differences exist. Meanwhile, another literature has documented significant pay heterogeneity across firms (Abowd, Kramarz, and Margolis 1999). Firms, meanwhile do not sort uniformly across cities, and thus, a significant share of spacial wage heterogeneity may be due to the way in which firms sort into cities.

A key challenge with studying how these two forms of wage heterogeneity relate to each other is the difficulty of obtaining data that follows workers over time and separately identifies both their city and employer. I study the urban wage premium in the context of Brazil using a detailed employer-employee matched administrative dataset, and using this data I answer 3 important questions about the urban wage premium: (1) What role does firm sorting play in the urban wage premium? (2) How important is firm and occupational matching for agglomeration economies? (3) Are wages high in cities because of compensating wage differentials?

I first show that a significant share of agglomeration economies in Brazil

that are estimated using traditional methods— 44%— is actually due to positive selection of high-wage firms into larger cities, so they are in fact not really agglomeration economies at all, suggesting that estimates of the agglomerations economies that do not account for firm selection may significantly overestimate the benefits of agglomeration. While worker selection has long been considered an issue to be dealt with in the estimation of agglomeration economies (starting with Glaeser and Mare, 2001), firm selection into larger cities has not received much attention, despite this large literature documenting significant firm-level pay premia and discounts due to compensating wage differentials and rent sharing (see Abowd et al., 1999; Card et al., 2013; and Sorkin, 2018; among many others). I exploit the existence of multi-city firms (firms with multiple establishments spanning at least two separate cities) to disentangle firm premia from city premia for the first time.

The intuition for this estimation is similar to that of AKM models (Abowd, Kramarz, and Margolis, 1999), where identification of firm and worker premia comes from 'movers', workers who move firms, and identification only exists for an interconnected set of firms and workers. In the present case, identification comes from 'overlappers'— firms with multiple establishments across different cities, and similarly, effects are identified only for an interconnected set of cities, firms, and individuals.

Agglomeration effects have been theorized to come from 3 main sources, (1) sharing, (2) matching, (3) and learning (discussed in detail in Duranton and Puga, 2004). My data is well-suited to understanding the role of employer (firm) and occupational matching, so after establishing the importance of firm effects in understanding the urban wage premium, I then move to estimating the role that matching plays in agglomeration economies. A worker might be particularly productive at a certain firm or occupation, and the likelihood of finding a good match may be higher in cities because the labor market is thicker

and there are more options. I find that these firm match effects account for 76% of agglomeration effects. I then use occupational match effects rather than firm match effects, and find that occupational matching accounts for around 39% of the agglomeration effects. Finally, I include both match effects simultaneously and find that firm and occupational matching together account for 87% of agglomeration effects. In the context of the main microfoundational theories for agglomeration economies described by Duranton and Puga (2004), these results suggest that matching plays a large role in explaining agglomeration economies in Brazil, while sharing and learning play less of a role.

In the first part of this paper I establish that the selection of high-wage firms into larger cities accounts for a great deal of the urban wage premium. But what are high-wage firms? Recent work by Sorkin (2018) has shown that a majority of variation in firm premia is due to compensating wage differentials. So does this mean that cities have high wages because jobs there have bad characteristics that are being compensated for? To investigate this, I estimate Sorkin's model using the Brazilian data to get workers' revealed-preference valuation of jobs, and I find that jobs in larger cities have better non-wage characteristics rather than worse. Thus, high wages in cities are not due to compensating wage differentials.

I first contribute to the literature studying the urban wage premium. This literature has documented significant pay and productivity differences between urban and rural places, particularly in the developing world (Hicks, Kleemans, Li, Miguel 2017; Gollin, Lagakos, Waugh 2013, Lagakos 2020). Recent work has sought to understand why such differences persist, specifically studying moving costs (Bryan and Morten 2019, Lagakos, Marshall, Mobarak, Vernot, Waugh 2020), risk & uncertainty (Bryan, Chowdhury, Mobarak 2014), and compensating wage differentials (Gollin, Kirchberger, and Lagakos 2020). I contribute to this literature by showing that firm sorting plays a key role in

wage heterogeneity, that firm matching may play a role in the persistence of these gaps, and that compensating wage differentials (measured quite broadly) do not explain the urban wage premium.

This paper also contributes to the extensive literature that measures agglomeration economies, and seeks to understand their sources (see Bryan, Glaeser, and Tsivanidis 2020 for a recent review). Starting with Glaeser and Maré (2001), this literature has sought to eliminate worker-selection bias in the estimation of aggolomeration effects by using panel data and including worker-level fixed effects, and thus identifying agglomeration effects by the change in wages or productivity that a worker experiences when they move to a different city. Duranton and Puga (2004) lay out the key theoretical microfoundations that may explain agglomeration economies, and more recently, empirical studies have begun to measure the extent to which different theories matter, such as learning (De la Roca and Puga 2017), occupational matching (Papageorgiou 2020), and assortative matching (Dauth et al. 2018). I contribute to this literature by being the first (to my knowledge) to control for *firm* selection in estimating agglomeration effects, and by showing that occupation and firm match effects account for the majority of agglomeration effects.

2.2 Data

My main source of data is the RAIS dataset (*Relação Anual de Informações Sociais*) maintained by the Brazilian Ministry of Labor. The RAIS dataset is an employer-employee matched dataset containing essentially the entirety of the formal sector of Brazil (the formal sector being a bit over 60% of the workforce, depending on the year, according to International Labor Organization estimates). The only workers/entities that are not reported in RAIS are the informally employed, self-employed, elected officials, interns/trainees, domestic employees, and cooperatives. Employees have a personal stake in making sure

that they are reported in RAIS, because for many it guarantees that they recieve their yearly bonus (*Abono Salarial*) paid by the government. RAIS contains data for each worker on monthly wage, education level, a 6-digit occupation code, a 5-digit industry code, employer, type of employment contract, location of work, hiring date, and both date of and reason for end of employment.

For this paper I use data from 2009 to 2018, and for computational reasons I take a random sample of 20% of people who worked during this time period. I further restrict the sample by dropping all public sector workers (as their wages are highly regulated), dropping all 1-person firms, and following Card et al. (2013), I drop any part time jobs and restrict the sample to only male workers. For each worker I keep only the 'main job' of each worker for each year, where the 'main job' is defined as the job from which the worker earned the most money that year. Finally, I restrict the sample to only the largest interconnected set of firms and workers[1] (which covers 99.89% of the sample), and because identification comes only from 'switchers', I restrict the sample to be only firm-switchers (as is common in this literature, again following Card et al, 2013).

Beyond the RAIS data, the main data of importance is the definition of agglomerations. Local Brazilian jurisdictions are known as municipalities, but all agglomerations include many municipalities, so I need a way to aggregate municipalities into distinct agglomerations. Brazil does define metropolitan regions, similar to MSAs in the US, however these definitions are done at the state level, and thus are not consistent across the country. Thus, I follow Chauvin et al. (2017) and use *microregions*. Microregions are defined by the Brazilian Institute for Geography and Statistics (IBGE), which groups municipalities together that are economically integrated and have a common labor market. There are 558 microregions in Brazil, and on average each microregion contains 9.89 municipalities. The largest microregion (in terms of population) is São

[1] I use the algorithm described in Abowd et al. (2002).

Paulo, which contains 8 municipalities and had a population of 13.6 million at the beginning of my sample in 2009. The smallest microregion is Japurá in the Brazilian Amazon, which contains 2 municipalities and had a combined population of 22,000 in 2009[2]. To measure the populations of each microregion, I use the IBGE's 2009 municipal population estimates and sum across all municipalities in each microregion. In what follows, I will use the term 'city' and 'microregion' interchangably.

In my sample there are 909,649 distinct firms, and 42,262 of these (about 4.7% of them) have establishments in multiple microregions, while 404 firms have establishments in at least 20 microregions. The highest amount of microregions any one firm is in is 390, so no one firm is in all 558 microregions. I implement the same algorithm as above to find the largest interconnected set of firms and microregions, and find that these 42,262 multi-city firms together span 555 of the 558 microregions in Brazil. Thus, I also drop the 3 unconnected microregions from my sample.

2.3 Econometric Model

2.3.1 The Baseline AKM Model

In order to better understand the urban wage premium and agglomeration economies, I follow previous literature and estimate the premium associated with working in each city. To do this, I employ an AKM-style fixed effects panel model. In my baseline specification, I follow previous literature and estimate city premia (δ_c), controlling only for an individual worker effect (α_i), along with occupation, industry, experience and tenure (contained in X_{it}), as shown in equation 2.3.1. The outcome variable, y_{it}, is log wages for individual i during year t.

[2]Technically, the smallest microregion is Fernando de Noronha, but this is an archipelago 200 miles off the northeastern coast of Brazil, and contains no municipalities.

$$y_{it} = \beta_0 + \beta_1 X_{it} + \delta_c + \alpha_i + \gamma_t + \varepsilon_{it} \tag{2.3.1}$$

The AKM literature has established that firms are an important source of wage heterogeneity (Abowd, Kramarz, Margolis, 1999; Card, Heining, Kline, 2013), and if high-wage firms tend to select into large cities, this could bias estimates of agglomeration effects. Thus, I next introduce to the specification a firm effect (ϕ_j), to account for high-wage firms selecting into cities, as shown in equation 2.3.2. Define δ_c as a city effect, α_i is a worker effect, ϕ_j is a firm effect, and θ_o is an occupation effect. The functions $c(i,t)$, $j(i,t)$, and $o(i,t)$ respectively give the city, firm, and occupation of individual i at time t.

$$y_{it} = \beta_0 + \beta_1 X_{it} + \delta_{c(i,t)} + \alpha_i + \gamma_t + \phi_{j(i,t)} + \theta_{o(i,t)} + \varepsilon_{it} \tag{2.3.2}$$

The goal of this equation is to get an unbiased estimate for δ_c, the city premium. With this estimate of δ_c, we can estimate the elasticity of the city element of pay with respect to city size, which measures agglomeration economies.

An important note about this specification is that ϕ_j is a *firm* effect rather than an establishment effect. This allows me to disentangle city premia from firm premia, and as far as I know this is the first paper to do so. Disentangling city premia from firm premia enables me to show that firm premia actually account for a significant fraction of the city premium that is typically estimated when firm effects are not considered. Estimation of this high-dimensional fixed-effects model is done using the method described by Correia (2017).

2.3.2 Testing the Validity of AKM

According to the AKM literature, a key condition to get unbiased estimates of δ_c is the 'no endogenous mobility' assumption. One way that this assumption

could be violated is if workers or firms decide which city to work or operate in based on a city-worker- or city-firm-specific idiosyncratic shock. I evaluate this concern by creating a 'symmetry graphs' as done by Adhvaryu et al. (2020).

I start first evaluating the significance of selection based on city-worker shocks. The intuition behind this procedure is to see if the wage losses of people moving from high-wage cities to low-wage cities are roughly symmetric to the gains of people making the opposite transition, moving from low-wage cities to high-wage cities. If they are not— for example, if people moving from high-wage cities to low-wage cities don't experience wage declines, while people making the opposite transition see wage increases— then this is evidence of endogenous mobility, in other words, people choosing location based on a city-worker specific idiosyncratic shock.

To create these symmetry graphs, I take each worker and their wage, and group them into quartiles based on the average wage of the city they are in, and for workers who then move to a different city, I get the wage quartile of the new city, as well as the worker's wage in the new city. The symmetry graph is then made by graphing the average wage change from moving up in the city distribution by the average wage change of workers making the opposite move (moving down in the city distribution), for moves between each of the 4 quartiles of the city distribution. The wages I use are log wages, projected on firm dummies, in order to control for the fact that high-wage cities tend to have high-wage firms.

The graph resulting from this procedure is shown in figure 2.2 in the appendix. As can be seen, wage gains and losses are symmetric along a -45 degree line, indicating that selection based on city-worker idiosyncratic shocks is unlikely to be an issue.

Next I evaluate the significance of selection on city-firm shocks. This procedure is similar to the procedure above, except that because firms don't often

move from city to city, instead I look at firms that have multiple establishments in different cities, and I look to see if the differences in wages between the headquarters and satellite establishments are symmetric between firms where the satellite establishment is in a lower-wage city versus firms where the satellite establishment is in a higher-wage city[3]. The idea here is to examine if firms systematicaly choose to expand in to cities where they might have some positive idiosyncratic shock.

To create this graph I take the average wage and the city quartile of each non-headquarter establishment, and then I get the average wage and city quartile of that firm's headquarters establishment. I graph the difference between satellite establishment average wages and headquarters average wages for establishments in a higher quartile than their headquarters by the same difference for opposite quartile difference. The wages I use here are log wages, projected on occupation dummies, in order to control for the fact that firms may have different occupations working in different cities.

The graph resulting from this procedure is shown in figure 2.3 in the appendix. Again, as can be seen, wage differences are symmetric, accounting for the fact that satellite establishments always pay a constant amount less than the headquarters, whether the headquarters is in a lower- or higher-wage city.

One additional way to test for the importance of city-worker or city-firm idiosyncratic shocks is to directly include city-worker and city-firm dummies in the regression, and see if this improves the fit of the model. In table 2.5 I show the R^2 and Adjusted R^2 for the baseline estimating equation (equation 2.3.2), and then a version of this equation with city-worker interaction effects added in, one with city-firm effects added in, and lastly a specification with both city-worker and city-firm fixed effects added in. As can be seen, while the R^2 does increase slightly with the addition of these variables (as is natural when

[3]Crucially, my data separately indicates headquarter establishments from secondary establishments.

adding in a huge vector of variables), the adjusted R^2 actually *decreases* with the addition of these variables, suggesting that these city-worker and city-firm effects do not fit the data well, thus once again implying that they are not an issue biasing results.

2.3.3 Decomposing Agglomeration Effects

While it is of interest to estimate $\delta_{c(i,t)}$ in order to measure agglomeration effects, it is also of interested to break down $\delta_{c(i,t)}$ into some component parts, in order to understand where the agglomeration economies are coming from. Thus, let's write a richer model of $\delta_{c(i,t)}$, which accounts variation in the quality of firm and occupational matching by city, as well as a residual city effect that measures all city effects that are not accounted for by the match effects. The goal now is to understand how much of the city premium is due to firm and occupational matching versus other factors.

We can separate match effects into two sources. There is *within-city* variation in match effects, as well as *across-city* variation in match effects. We can attribute across-city variation in match effects to agglomeration effects (controlling for firm and individual effects, if there are on average better matches in one place, then this is part of the city effect). Define $\xi_{i,j(i,t)}$ to be a firm-worker match effect for a match between worker i and firm j, and $\zeta_{i,o(i,t)}$ to be an occupation-worker match effect for match of worker i with occupation o. I want to capture the across-city variation in both firm and occupation matches, which is given by the average match effect in the city. So assuming that these match effects are additively seperable, we can write

$$\delta_{c(i,t)} = \overline{\xi}_{c(i,t)} + \overline{\zeta}_{c(i,t)} + \lambda_{c(i,t)}.$$

where $\overline{\xi}_{c(i,t)} = \frac{\sum_{(i,t)\in c} \xi_{i,j(i,t)}}{\sum_{(i,t)\in c} 1}$ and $\overline{\zeta}_{c(i,t)} = \frac{\sum_{(i,t)\in c} \zeta_{i,o(i,t)}}{\sum_{(i,t)\in c} 1}$ and $(i,t) \in c$ denotes

observations where worker i at time t is in city c, and then λ_c is the residual city effect, capturing the city premium that is left over after accounting for firm and occupational match effects.

The terms $\overline{\xi}_{c(i,t)}$ and $\overline{\zeta}_{c(i,t)}$ are both averages of the firm and occupation match effects in a given city across individuals and time. In order to get these averages, we need to estimate the actual match effects. What we can do now is write a more nuanced version of the firm and occupation effects, where instead of firm and occupation effects being constant across all workers, they vary by worker, thus becoming 'match effects'. For each individual i, these match effects are only identified for firms j and occupations o that individual i works in for some time periods, and not others. Since the mean firm and occupation match effects (mean by city, that is) are already included in δ_c, then these worker-firm and worker-occupation match effects will be demeaned by city average, like so: $(\xi_{i,j(i,t)} - \overline{\xi}_{c(i,t)})$ and $(\zeta_{i,o(i,t)} - \overline{\zeta}_{c(i,t)})$. In other words, these terms capture the within-city variation in matches. Replacing the simple firm and occupation effects with these firm and occupation match effects, equation 2.3.1 becomes

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 X_{it} + \delta_{c(i,t)} + \alpha_i + \gamma_t \\ + \underbrace{(\xi_{i,j(i,t)} - \overline{\xi}_{c(i,t)})}_{\text{Demeaned firm match effects}} + \underbrace{(\zeta_{i,o(i,t)} - \overline{\zeta}_{c(i,t)})}_{\text{Demeaned occupation match effects}} + \varepsilon_{it} \end{aligned} \tag{2.3.3}$$

Plugging in for δ_c, we can write the equation like this,

$$\begin{aligned} y_{it} = \beta_0 + \beta_1 X_{it} + \underbrace{(\overline{\xi}_{c(i,t)} + \overline{\zeta}_{c(i,t)} + \lambda_{c(i,t)})}_{\delta_{c(i,t)}} + \alpha_i + \gamma_t \\ + \underbrace{(\xi_{i,j(i,t)} - \overline{\xi}_{c(i,t)})}_{\text{Demeaned firm match effects}} + \underbrace{(\zeta_{i,o(i,t)} - \overline{\zeta}_{c(i,t)})}_{\text{Demeaned occupation match effects}} + \varepsilon_{it} \end{aligned} \tag{2.3.4}$$

The positive and negative versions of both $\overline{\xi}_{c(i,t)}$ and $\overline{\zeta}_{c(i,t)}$ that are part of

δ_c and the match effect, respectively, cancel each other out. So this equation can just be estimated as

$$y_{it} = \beta_0 + \beta_1 X_{it} + \lambda_{c(i,t)} + \alpha_i + \gamma_t + \xi_{i,j(i,t)} + \zeta_{i,o(i,t)} + \varepsilon_{it} \tag{2.3.5}$$

And this equation will give us an estimate for λ_c, which tells us the residual city premium that is left over after controlling for match effects. We can also calculate $\overline{\xi}_{c(i,t)}$ and $\overline{\zeta}_{c(i,t)}$ after estimation.

Following this 'first stage' which estimates the city premia, I estimate the elasticity of city premium with respect to city population with the regression in equation 2.3.6, where β is the coefficient of interest.

$$\delta_c = \alpha + \beta Log(Population_{c,2009}) + \nu_c \tag{2.3.6}$$

The elasticity estimated by this equation is commonly used as a measure of agglomeration effects, as it shows how wages increase with density. In my empirical estimates, I will use the city premium of each city, δ_c as the outcome in the elasticity estimation above, and I will also estimate elasticities with the residual city premium λ_c (which captures what is left over of the city premium after firm and occupation matching are accounted for) and various intermediate residual city premiums (where only firm matching or only occupation matching are accounted for) as outcomes as well. These 'intermediate' city premia are simply the coefficients on the city dummies in various specifications of my model.

2.4 Results

First I estimate equation 2.3.2 and plot the estimated city effects (microregion premium) against the log population of each city in figure 2.1. Because from the

scatterplot it is clear that agglomeration economies only exist after a certain level of agglomeration is reached, I also graph the line from a nonparametric kernel regression in order to see more clearly how the city premium changes with population. As can be seen, the slope of the line is flat for smaller agglomerations, and only becomes positive at agglomerations with log population of at least 12.5, which translates into a population of 268,337. This suggests that agglomeration effects may only kick in after a certain level of agglomeration. For my estimates of the elasticity of city effects with respect to city population I use only cities above this population level, similar to restrictions made by previous researchers (Chauvin et al. 2017; De la Roca and Puga 2017).

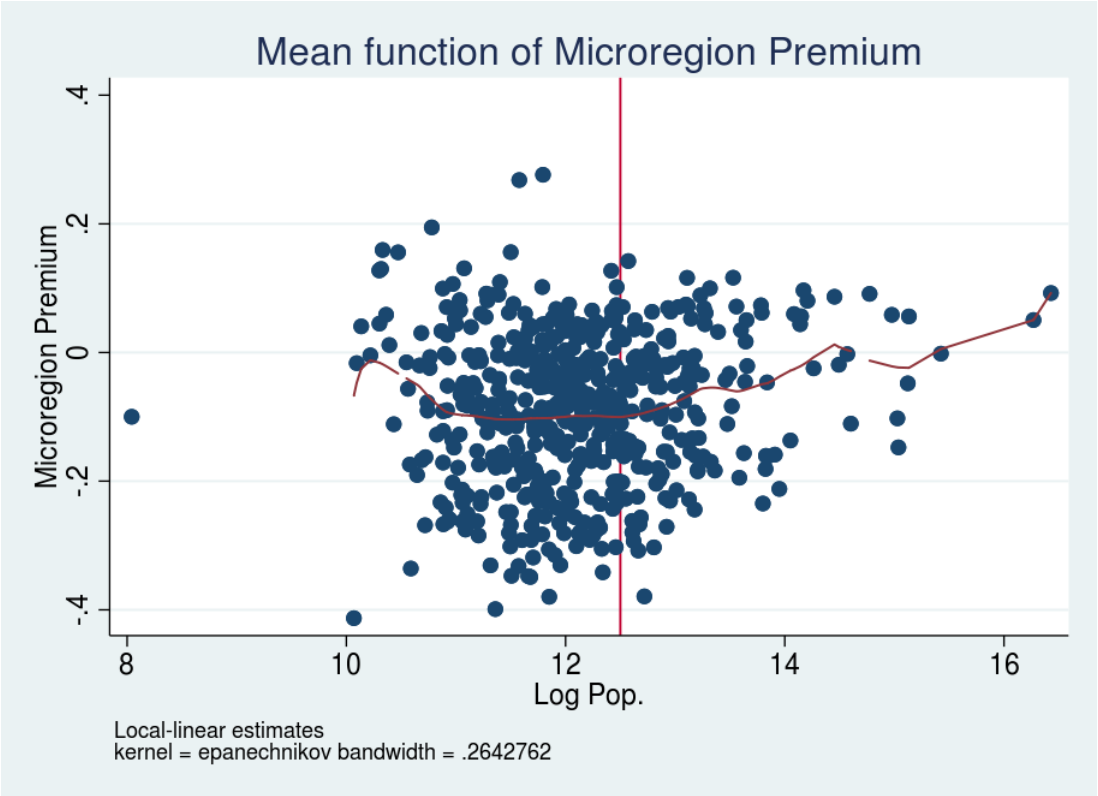


Figure 2.1: This graph uses the city premium estimates from column (2) of table 2.1. The vertical line is at 12.5, which implies a population of 268,337.

In Panel A of table 2.1 I estimate increasingly nuanced versions of the fixed effect model shown in section 2.3, which gives me the city premium and residual city premiums. For each specification in columns 1 through 7, in panel B I take

the coefficient associated with the city dummy variable estimated in panel A of the same column, and I estimate the elasticity of these coefficients with respect to population, as has previously been done in the literature. Thus, Column(3) gives the elasticity of δ_c with respect to city size, and column (7) gives the elasticity of λ_c with respect to city size, with the columns in between giving intermediate residual city premia.

Table 2.1: Elasticity of City Premium with respect to Population

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Regression equation		2.3.1	2.3.2				2.3.5
Panel A: 'First Stage'— estimate microregion premium							
Experience	0.0428***	0.00750***	0.00421***	0.00329***	0.00887***	0.00584***	-0.000826
	(0.00128)	(0.000768)	(0.000679)	(0.00112)	(0.000831)	(0.000696)	(0.00121)
Tenure	0.0332***	0.0187***	0.0166***	0.00613***	0.00954***	0.00900***	0.00585***
	(0.00107)	(0.000749)	(0.000535)	(0.000314)	(0.000463)	(0.000319)	(0.000288)
Observations	26767300	26612592	26587212	20083395	21344483	21210114	18631618
Microregions	555	555	555	555	555	555	555
City Effects	Y	Y	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y	Y	Y
Occupation Effects	Y	Y	Y	Y	Y	Y	Y
Industry Effects	Y	Y	Y	Y	Y	Y	Y
Individual Effects	N	Y	Y	Y	Y	Y	Y
Firm Fixed Effects	N	N	Y	Y	N	Y	Y
Firm Match Fixed Effects	N	N	N	Y	N	N	Y
Occupation Match Fixed Effects	N	N	N	N	Y	Y	Y
Panel B: Elasticity of Estimated City Premium with respect to City Size							
Log Pop.	0.0632***	0.0452***	0.0251***	0.00625	0.0330***	0.0153**	0.00317
	(0.0154)	(0.0116)	(0.00742)	(0.00393)	(0.00939)	(0.00613)	(0.00359)
Observations	158	158	158	158	158	158	158
Panel C: Percent Change in Elasticity from Baseline							
Equation 2.3.1 as baseline	-	(Baseline)	44.47%	86.17%	26.99%	66.15%	92.99%
Equation 2.3.2 as baseline	-	-	(Baseline)	76.10%	-	39.04%	87.37%

In Panel A of this table I estimate the fixed effect models described in section 2.3, and then in Panels B I estimate the elasticity of city premium with respect to city population implied by the corresponding microregion premiums ('city effects') estimated in the corresponding column of Panel A. In panel B, the sample is restricted to agglomerations of over 268,337 people. Standard errors in parenthesis

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For completeness, I first estimate a pooled OLS specification in column (1) where I leave out the individual effect. The estimate for the elasticity is about .0632, which is a bit larger than previous estimates of similar specifications that don't control for individual effects, as the estimates in Combes et al (2010) for France, Glaeser and Resseger (2010) for the US, Keisuke (2017) for Japan, and De la Roca and Puga (2017) for Spain. Next in column (2) I estimate the model shown in 2.3.1, which I consider to be my baseline model. When I include individual fixed effects, the elasticity becomes .0452, which is significantly larger than estimates from more wealthy countries. For example, De la Roca and Puga (2017) estimate an elasticity of .0241 for Spain. This difference may simply come from differing definitions of what an urban area is, and so I don't focus on comparing elasticity estimates across countries. Rather, I will focus on comparing elasticity estimates within Brazil, comparing the elasticity of city premia with respect to city size to the elasticity of various residual city premia with respect to city size.

Next, I begin to incorporate controls for firms. In the literature studying agglomeration economies, controlling for individual effects has long been a common practice, as it controls for high-ability workers selecting into cities. However, the issue of selection of highly-productive (or high-wage) firms into cities has only barely begun to be considered (Dauth et al. 2018). In order to disentangle firm effects from city effects, I utilize the existence of 'overlapping firms'— that is, an interconnected set of multi-city firms. The intuition behind this is similar to the identification of firm and individual effects in the AKM model, with identification coming from 'movers'— workers who move from one firm to another— but in this case since firms rarely move, I use firms that simultaneously exist in multiple cities via different establishments. Once I control for firm selection in column (3), I find that the urban wage premium drops significantly. The elasticity of city premium with respect to population drops

from .0452 to .0251, a 44% drop[4]. This estimate— .0251— is my preferred estimate of the elasticity of wages with respect to population, because it properly controls for the selection of high-wage firms into large urban areas. This result suggests that previous work that does not consider the selection of high-wage firms into urban areas may significantly overestimate agglomeration effects.

This drop in apparent agglomeration economies that results from controlling for firm premia is quite large, and so next I investigate the reason for this drop. There are 2 possible reasons that controlling for firm effects changes the estimated agglomeration effect: (1) firms with larger pay premia tend to locate themselves in larger cities, or (2) migrants tend to match differently in the distribution of firms after they migrate to a city.

I first examine the possibility that high-wage firms tend to locate themselves in larger cities. To see how large of a factor this difference might be in explaining the drop in estimated agglomeration effects, I estimate the elasticity of average firm premia received by workers with respect to population. To do this I first take the exponential of estimated firm effects for each person-year in the data, and then I find the average of this by microregion. After I have this microregion average of firm effects, and then take its log again[5]. A similar procedure is followed for getting the average person effect by city.

Table 2.2: Elasticity of Firm and Person Effects with respect to Population

	Avg. Firm Effect	Avg. Person Effect
Log Pop.	0.0266***	0.0483***
	(0.00796)	(0.00763)
Observations	158	158

Outcome variables are the average firm and person effects received by workers in each city. Standard errors in parenthesis

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In table 2.2 I find that the elasticity of firm premia with respect to city size

[4] 1 - 0.0251/ 0.0452 = 44.47%

[5] The reason for taking the exponential and then taking the log is of course that the log function is not a linear function, and so the sum of logs is not the same is a log of sums.

is large enough that it suggests that rural-urban migrants may match *lower* in the firm distribution after they migrate, because the elasticity of firm premia with respect to city size (.0266) accounts for more than the decline in the elasticity the results from include the firm effects (.0201)[6]. In other words, if migrating workers matched to firms in the same part of the firm effect distribution in destinations as their origin, then we would expect to see a slightly bigger decrease in the elasticity of the city effect with respect to population when we control for firm effects.

As a comparison, I also estimate the elasticity of person effects with respect to city size, which measures selection of high-ability workers into larger cities, and I find that the elasticity is large (in comparison to the elasticity of city effects with respect to population), but only slightly larger than the elasticity of firm effects with respect to city size.

To see if there are such differences in how workers match in the distribution of firms, I examine assortative matching for migrants compared to the rest of workers in both urban and rural places. In table 2.3 I show the correlation between firm effects and person effects for all the workers in my sample, differentiated by urban vs rural (defining urban places as cities in the top 10% of population, and rural in the bottom 90% of population) and differentiated by rural-urban migrants vs. all workers. As can be seen in the first row, urban workers tend to do much more positive assortative matching than rural workers, in line with recent work done by Dauth et al. (2018) with German data. What is interesting here is that while migrants tend to be more positively matched than the average worker in rural places before migrating, they are less positively matched than the average worker in urban places that they go to (even while they themselves become more positively matched when the move to an urban area).

Thus it seems that the decrease in elasticity of city effect with respect to city

[6].0452-.0251 = .0201

Table 2.3: Correlation between firm effects and person effects

	Rural	Urban
All workers	.1578	.2691
Migrants	.2125	.2431

size is driven largely by the fact that high-wage firms tend to be located in cities, and is in small part *counteracted* by the fact that migrants tend to match lower in urban areas than they do in rural areas. This may help to explain why such large wage differences persist between urban and rural places— while migrants may move to a high-paying city, they may have a difficult time matching with high-wage firms.

Having established the importance of controlling for firm premia when estimating agglomeration effects, I now turn to further decomposing agglomeration effects. Agglomeration effects are thought to come from 3 main sources: (1) sharing, (2) matching, (3) and learning (see Duranton and Puga, 2004).

In order to study the importance of matching, I now estimate a more nuanced model that incorporates firm match effects to capture the increased opportunity in cities to match to a firm that may suit a person well, and where they may be particularly productive. These firm match effects are simply a separate indicator for each firm-worker pair. In Column (4) I find that once these match effects are accounted for, the elasticity of the city effects with respect to city size is quite small (.00625). These firm match effects account for 76%[7] of agglomeration economies, compared to the baseline specification estimated in column (3).

Now in column (5) I remove the firm effects and now focus on occupational match effects. These occupational match effects are an indicator for each unique worker-occupation pair, using 5-digit occupation codes. Then in column (6) I redo this analysis, but putting the firm effects back in. I find that occupational

[7] 1 - .00625/.0251 = 76.10%

match effects account for about 39%[8] of the urban wage premium, again using equation 2.3.2 as the base model. This estimate is similar to recent work by Papageorgiou (2020), who uses a structural model and US data, and finds that occupational matching accounts for 35% of the urban wage premium.

Finally, in column (7) I put both the firm and match effects in, which gives us the specification outlined in equation 2.3.5, and thus the elasticity in panel B is the elasticity of the residual city premium, λ_c with respect to city size. I find that compared to the baseline model (column 2), the firm and match effects combined account for a total of 87%[9] of the elasticity of city premium with respect to city size[10]. This result is consistent with cities in Brazil existing as huge labor markets, where high-wage firms and high-wage workers go to find good matches. Thus, in Brazil at least, sharing and learning effects seem to play a comparatively small role in the urban wage premium compared to matching effects.

2.5 High Wage Firms and Compensating Wage Differentials

We know from the above that a significant amount of the urban wage premium is explained by high-wage firms selecting into bigger cities. But what are high-wage firms anyway? Until recently, the AKM firm premium (the fact that there is a firm premium or discount paid to all employees) was well established empirically, but mostly a black box, with some possible explanations being firms sharing rents with workers or compensating differentials, but little quantitative evidence as to the relative importance of these factors.

This changed with Sorkin (2018) when Sorkin introduced a structural model

[8] 1 - 0.0153/ 0.0251 = 39.04%

[9] 1 - .00317/.0251 = 87.37

[10] The separate estimation of the firm effects and match effects does not add up to be the same as when I estimate them simultaneously because some people simultaneously switch both firms and occupations.

to estimate the revealed-preference valuation of employment at a given firm, and found that 70% or more of variation in the firm component of wages is due to compensating differentials. So coming back to the present study, does this imply that firms pay more in cities just because there are bad non-wage characteristics? While cities have amenities that people enjoy, such as restaurants, arts, and live entertainment ranging from sports to theater, cities are also associated with traffic, pollution, crime, and high cost of living. In order to answer this question I will estimate a modified version of Sorkin's model in order to get workers' revealed-preference valuation of employment at each establishment, to see if workers value jobs in bigger cities more or less than jobs in smaller cities.

The intuition behind Sorkin's model is to use 'endogenous' worker flows—flows where the worker chooses move jobs, and thus reveals their preferences—to see which are the best establishments to work for, according to the workers. In the structural analysis I will now use establishments instead of firms, for reasons that will be seen below. To see Sorkin's model mathematically, assume N workers are choosing between working for establishment k and establishment j. Workers at establishment k receive $\tilde{V_k}$ utility plus an idiosyncratic draw that is distributed type I extreme value with scale parameter 1. Then, according to discrete choice theory, the fraction of workers choosing firm k over firm j can be written

$$\frac{M_{kj}}{N} = \frac{exp(\tilde{V_k})}{exp(\tilde{V_k}) + exp(\tilde{V_j})}, \tag{2.5.1}$$

where M_{kj} is the flow of workers going from establishment j to establishment k. The main adaptation I make to Sorkin's model is that instead of estimating the likelihood that a given move is endogenous (as Sorkin does), I observe directly if a given flow results from a quit rather than a layoff, and I use only these flows that result from quits. Thus, the assumption is that these quits are endogenous (reveal workers' preferences). In some countries, quits may not

reveal preferences because a person may be pressured to resign or may quit pre-emptively when getting fired is on the horizon. However, this is not the case in Brazil. Brazilian labor law makes firing very costly to firms, and fired employees receive many benefits that quitting employees do not receive.[11] The incentive as an employee to get fired rather than quit is in fact so strong that employees wishing to terminate employment will sometimes *ask* to be fired.[12] Thus, we can be sure that quits are indeed revealing workers' preferences for one firm over another, and are generally not the result of some exogenous pressure to resign instead of being fired.

Now, if we take 2.5.1 and sum over all establishments j and then rearrange, we get a recursive definition of the value of employment at a given establishment.

$$\frac{\sum_j M_{kj} exp(\tilde{V_j})}{\sum_j M_{jk}} = exp(\tilde{V_k}) \tag{2.5.2}$$

This equation tells us that good establishments hire away from other good establishments, and have few people quit from them. However, to get the true value of working at the establishment, we need to correct for establishment size and the job offer distribution,

$$exp(V_j^e) = exp(\tilde{V_j}) \frac{g_j(1-\delta_j)(1-\rho_j)}{f_j}, \tag{2.5.3}$$

where g_j represents the relative size of establishment j, $(1-\delta_j)(1-\rho_j)$ represents the probability of not facing an exogenous shock that sends you to either unemployment (δ_j) or a different job (ρ_j), and f_j represents the relative share of job offers extended by establishment j. The rest of Sorkin's method revolves around estimating these parameters, which I leave to the appendix in section 2.9.

[11] https://thebrazilbusiness.com/article/dismissing-in-brazil
[12] https://thebrazilbusiness.com/article/why-brazilians-ask-to-be-fired

Once I have the estimates of V_j^e, the utility of working for each establishment j, there are 3 outcomes of interest to understand the role of compensating wage differentials in explaining the urban wage premium. The first is the average V^e received by workers in each city. The second is the value of non-wage characteristics received by workers in each city. I calculate this by regressing V^e on ϕ (a vector of the establishment premia estimated from an AKM model), then calculating the residual.[13] Thus, this measurement of non-wage characteristics is any part of the average value of working at an establishment that is not explained by the establishment premium[14]. Third is a measure of the city-specific component of V^e. The city-specific component of V^e can be calculated by comparing the value of employment among different establishments of the same firm, but are in different cities. This can be done by regressing V^e on city and firm effects, and taking the coefficients on the city effects as the city-specific component of V^e. This procedure once again uses identification from multi-city firms. Intuitively, what this procedure does is it asks if workers prefer to work at Sam's Club in São Paulo vs. Sam's Club in Porto Alegre or Sam's Club in Goiânia.[15]

Once I have these 3 outcomes of interest, I estimate the elasticity of each of them with respect to population. These estimates are presented in table 2.4.

I find that jobs in bigger cities give workers much more utility V^e, and as column (2) tells us, much of this comes from non-wage characteristics, rather than just the higher wages. The coefficient in column (3) tells us that the higher utility received in larger cities is not due to favorable city characteristics, but rather the characteristics of the types of jobs that tend to be in larger cities.

[13]Concretely, run the regression $V_j^e = \alpha_0 + \alpha_1\phi_j + u_j$ and then the value of non-wage characteristics is defined as $\hat{u}_j = V_j^e - \hat{\alpha_0} - \hat{\alpha_1}\phi_j$

[14]The establishment premium is estimated from an equation like equation 2.3.2, but using establishment dummies instead of firm dummies, and dropping out the city dummies.

[15]Specifically, estimate the regression $V_j^e = \tau_{c(j)} + \gamma_{k(j)} + \nu_j$, where τ and γ are city and firm fixed effects, respectively, $c(j)$ is a function that tells what city establishment j is in, and $k(j)$ is a function that tells what firm establishment j is part of. The outcome of interest (the city-specific V^e) is then the vector of τ_c coefficients.

Table 2.4: Population and Revealed-Preference Value of Jobs

	Avg. V^e	Non-wage Chars.	City-Specific V^e
Log Pop.	0.0981***	0.0696***	0.0353
	(0.0256)	(0.0233)	(0.0273)
Observations	158	158	158

Standard errors in parenthesis

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In other words, the positive amenities of larger cities (restaurants, arts, entertainment, ect.) seem to be cancelled out by the negatives (traffic, pollution, high cost of living, ect.). This is consistent with theory by Rosen (1974) and Roback (1982).

So compensating wage differentials do not explain the high wages found in cities of Brazil, consistent with the findings of Gollin, Kirchberger, and Lagakos (2020), who study public goods, crime, and pollution (which are a subset of my measure of compensating wage differentials) for a set of sub-Saharan African countries. Rather, my findings imply that wages are high in parger cities because firms there share more rents with workers (presumably because they have more rents).

2.6 Conclusion

In this paper I use detailed panel data to study the urban wage premium and aggolomeration economies in Brazil. I show that firm selection— an element not previously considered by researchers estimating agglomeration economies— accounts for a significant portion of the estimated elasticity of city premium with respect to city population, implying that estimates of agglomeration economies that do not account for firm selection are significantly upward-biased. Then I show that when controlling for firm selection is combined with controls for firm and occupational matching, this accounts for the majority of estimated agglomeration effects, suggesting that cities (in Brazil at least) exist because

they provide a large labor market where high-wage firms and high-wage workers can go and find productive matches. Finally, I examine the possibility that compensating wage differentials explain the high wages in cities. I show that jobs in larger cities in fact have better non-wage characteristics rather than worse, and thus compensating wage differentials do not explain the high wages of cities.

The findings of this paper are particularly interesting in light of the COVID-19 pandemic, and suggest some avenues of new research. Due to the pandemic many people have been leaving cities as remote work becomes more widespread, leading some commentators to wonder if there will be decreased productivity due to decreased agglomeration effects. Because I find that a significant portion of agglomeration effects come from occupation and firm matching, this suggests that productivity need not decline much, because new online labor markets may still be able to provide the high-quality matching that cities also perform.

2.7 References

Abowd, John M., Robert H. Creecy, and Francis Kramarz. Computing person and firm effects using linked longitudinal employer-employee data. No. 2002-06. Center for Economic Studies, US Census Bureau, 2002.

Abowd, John M., Francis Kramarz, and David N. Margolis. "High wage workers and high wage firms." Econometrica 67.2 (1999): 251-333.

Adhvaryu, Achyuta, et al. No line left behind: Assortative matching inside the firm. No. w27006. National Bureau of Economic Research, 2020.

Ardington, Cally, Anne Case, and Victoria Hosegood. "Labor supply responses to large social transfers: Longitudinal evidence from South Africa." American economic journal: Applied economics 1.1 (2009): 22-48.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. "Under-

investment in a profitable technology: The case of seasonal migration in Bangladesh." Econometrica 82.5 (2014): 1671-1748.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis. "Cities in the developing world." Annual Review of Economics 12 (2020): 273-297.

Bryan, Gharad, and Melanie Morten. "The aggregate productivity effects of internal migration: Evidence from indonesia." Journal of Political Economy 127.5 (2019): 2229-2268.

Card, David, Jörg Heining, and Patrick Kline. "Workplace heterogeneity and the rise of West German wage inequality." The Quarterly journal of economics 128.3 (2013): 967-1015.

Chauvin, Juan Pablo, et al. "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States." Journal of Urban Economics 98 (2017): 17-49.

Combes, Pierre-Philippe, et al. "Estimating agglomeration economies with history, geology, and worker effects." Agglomeration economics. University of Chicago Press, 2010. 15-66.

Correia, Sergio. "Linear models with high-dimensional fixed effects: An efficient and feasible estimator." Unpublished manuscript (2017).

Dauth, Wolfgang, et al. Matching in cities. No. w25227. National Bureau of Economic Research, 2018.

Duranton, Gilles, and Diego Puga. "Micro-foundations of urban agglomeration economies." Handbook of regional and urban economics. Vol. 4. Elsevier, 2004. 2063-2117.

Glaeser, Edward L., and David C. Maré. "Cities and skills." Journal of labor economics 19.2 (2001): 316-342.

Glaeser, Edward L., and Matthew G. Resseger. "The complementarity between cities and skills." Journal of Regional Science 50.1 (2010): 221-244.

Gollin, Douglas, Martina Kirchberger, and David Lagakos. "Do urban wage

premia reflect lower amenities? Evidence from Africa." Journal of Urban Economics 121 (2020): 103301.

Gollin, Douglas, David Lagakos, and Michael E. Waugh. "The agricultural productivity gap." The Quarterly Journal of Economics 129.2 (2014): 939-993.

Hicks, Joan Hamory, et al. Reevaluating agricultural productivity gaps with longitudinal microdata. No. w23253. National Bureau of Economic Research, 2017.

Keisuke, Kondo. Urban Wage Premium Revisited: Evidence from Japanese matched employeremployee data. No. 17047. 2017.

Lagakos, David. "Urban-rural gaps in the developing world: Does internal migration offer opportunities?." Journal of Economic perspectives 34.3 (2020): 174-92.

Lagakos, David, et al. "Migration costs and observational returns to migration in the developing world." Journal of Monetary Economics (2020).

Papageorgiou, Theodore. Occupational Matching and Cities. No. 2020-049. 2020.

Roback, Jennifer. "Wages, rents, and the quality of life." Journal of political Economy 90.6 (1982): 1257-1278.

Rosen, Sherwin. "Hedonic prices and implicit markets: product differentiation in pure competition." Journal of political economy 82.1 (1974): 34-55.

Roca, Jorge De La, and Diego Puga. "Learning by working in big cities." The Review of Economic Studies 84.1 (2017): 106-142.

Sorkin, Isaac. "Ranking firms using revealed preference." The quarterly journal of economics 133.3 (2018): 1331-1393.

2.8 Appendix

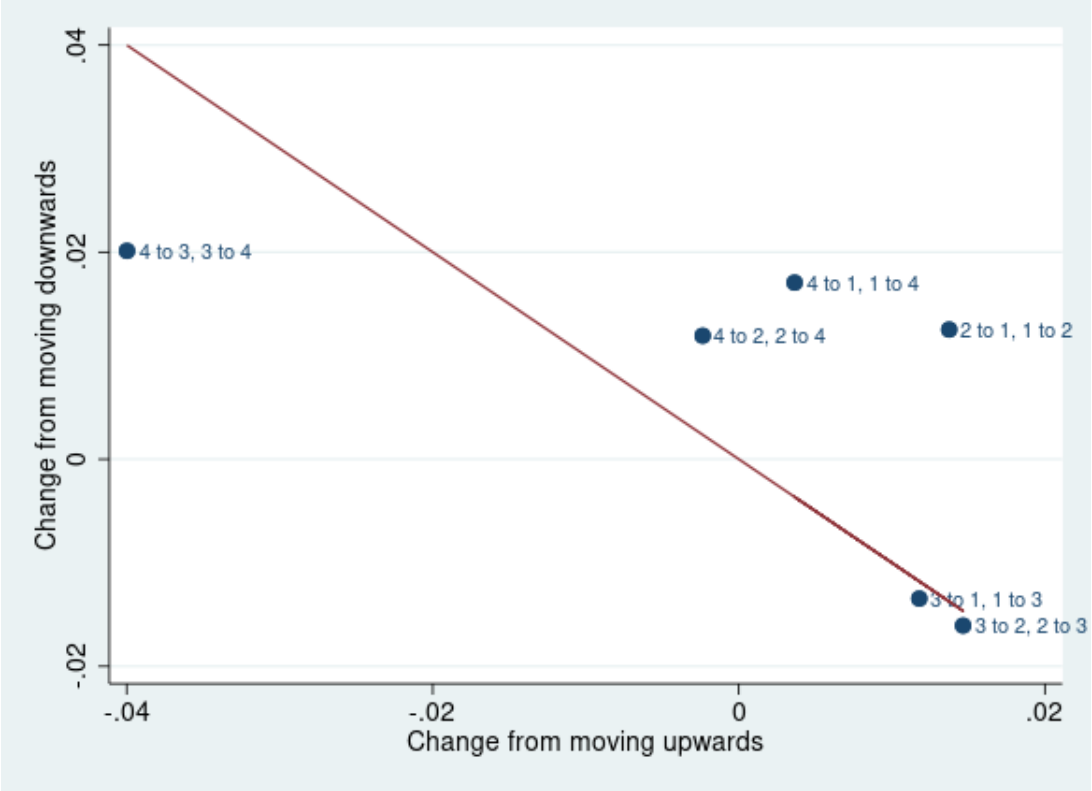


Figure 2.2: Testing for selection on city-worker shocks. This graph tests for endogenous mobility in the AKM model, following Adhvaryu et al. (2020). Each point shows the wage gains from moving up from city quartile X to city quartile Y, by the wage loss from moving down from quantile Y to quantile X, for all quartiles Y > X.

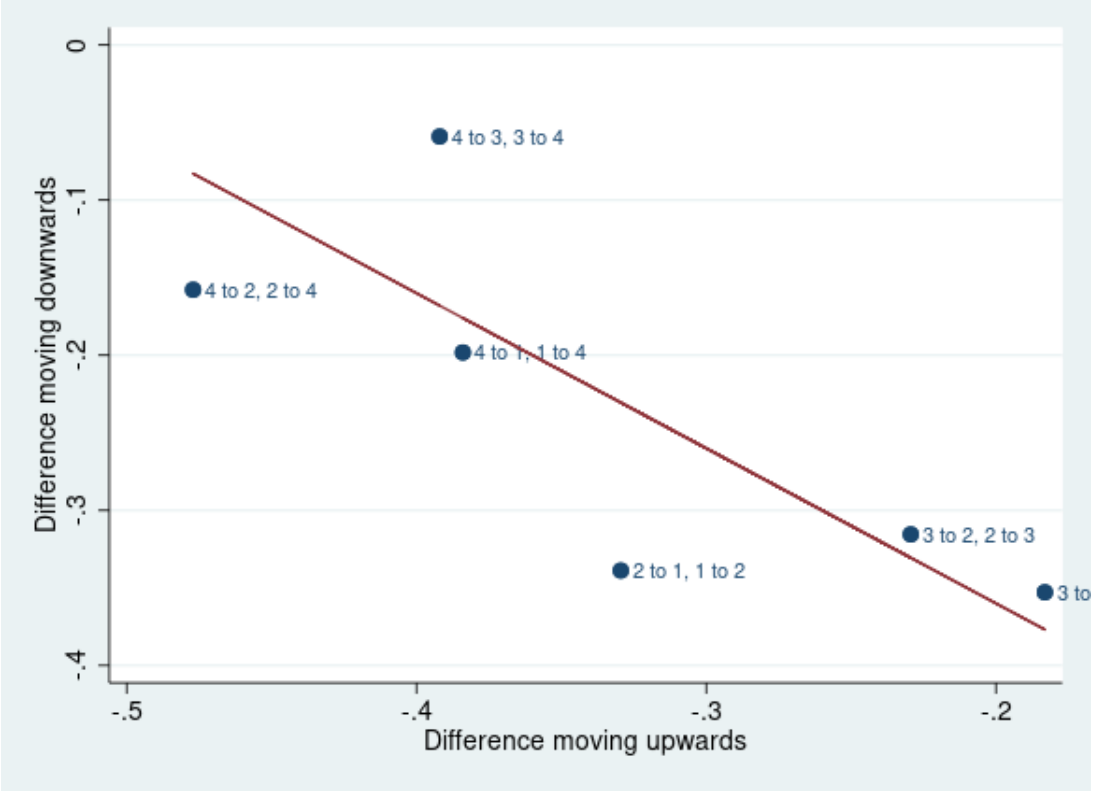


Figure 2.3: Testing for selection on city-firm shocks. This graph tests for endogenous mobility in the AKM model, following Adhvaryu et al. (2020). Each point shows the wage difference between satellite establishments in city quartile X and headquarters in city quartile Y, by the wage differences between satellite establishments in city quantile Y to and headquarters in quantile X, for all quartiles Y > X.

Table 2.5: Testing for endogenous mobility

	(1) Baseline (eq 2.3.2))	(2) City-worker effects	(3) City-firm effects	(4) Both
Observations	26587212	24453961	26584871	24444073
R^2	0.872	0.898	0.876	0.900
Adj. R^2	0.839	0.818	0.835	0.810

First column is the baseline (equation 2.3.2), second column ads a city-worker interaction effect, the third column instead ads a city-firm interaction effect, and the fourth column includes both of these.

2.9 Structural Model

I take Sorkin's (2018) framework, but I am able to simplify because I know when people quit. Then, the portion of workers N who choose employer k over

employer j in time t can be written

$$\frac{M_{kj}}{N} = \frac{exp(\tilde{V_k})}{exp(\tilde{V_k}) + exp(\tilde{V_j})}$$

Sorkin's equation (6) becomes

$$\frac{\sum_{j\in\varepsilon} M_{kj} exp(\tilde{V_j})}{\sum_{j\in\varepsilon} M_{jk}} = exp(\tilde{V_k})$$

and equation (7) is

$$S^{-1} M exp(\tilde{V}) = exp(\tilde{V})$$

And M must be filled with the 'strongly connected set' of employers

Equation (14), the flow ratios,

$$\frac{M_{jk}}{M_{kj}} = \frac{exp(\tilde{V_j})}{exp(\tilde{V_k})}$$

Equation (16) is

$$\sum_{j\in\varepsilon} M_{jn} exp(\tilde{V_n}) = \sum_{j\in\varepsilon} \lambda_0 U f_j \frac{exp(V_j^e)}{exp(V_j^e) + exp(V^n)} \frac{(1-\lambda_1) W exp(V^n)}{\lambda_0 U}$$

Equation (17)

$$\lambda_1 \sum_j g_j (1-\delta_j)(1-\rho_j) \sum_k f_k \frac{exp(V_k^e)}{exp(V_k^e) + exp(V_j^e)} = \frac{\sum_j \sum_k M_{jk}}{W \sum_j g_j (1-\delta_j)(1-\rho_j)} \tag{2.9.1}$$

Equations (18) and (19), which estimate the probability that a given separation is exogenous, is not used in my setup, because I directly observe quits and firings, and thus my M matrix is made up only of 'endogenous' separations—quits.

2.10 Model Estimation

Step 0 Initialize $\{V_j^e\}_{j\in\varepsilon}$ with guesses. Compute the relative size of employers (g_i), number of workers (W), exogenous separation probabilities (δ_j and ρ_j), and the share of people hired from nonemployment/informal employment by firm j, f_j^o from the data.

Step 1 Build the M matrix, where the (j,k) entry is the amount of people who quit at employer k and join employer j.

Step 2 Using equation (7), estimate the fixed point $exp(\tilde{V})$ (using a loop).

Step 3 Compute f, $exp(V_j^e)$ $exp(V^n)$, and λ_1, as detailed below

Step 4 Here Sorkin returns to step 1 in a loop to re-estimate exogenous/endogenous flow probabilities until things converge, but because I'm just using quits and not estimating the exogenous/endogenous flow probabilities, I'm done.

Details on step 3

Define C_1 as the share of offers accepted from nonemployment,

$$C_1 \equiv \sum_{j'\in\varepsilon} f_{j'} \frac{exp(V_j^e)}{exp(V_{j'}^e) + exp(V^n)} \tag{2.10.1}$$

then f_j^o, the share of people hired from nonemployment by firm j, can be written in terms of model parameters

$$f_j^o = \frac{f_j \frac{exp(V_j^e)}{exp(V_j^e)+exp(V^n)}}{C_1} \tag{2.10.2}$$

Now, take an initial guess for λ_1. Next I evaluate two equations. I will follow the convention that known quantities are on the left-hand side, and unknown quantities are on the right. The first equation follows from equation (13) and equation 2.3.1

$$\frac{g_j exp(\tilde{V_j})}{f_j^o}(1-\delta_j)(1-\rho_j) = g_j \frac{f_j exp(V_j^e)}{g_j(1-\delta_j)(1-\rho_j)} \frac{C_1}{f_i exp(V_j^e)}[exp(V_j^e)+exp(V^n)](1-\delta_j)(1-\rho_j)$$

$$= C_1[exp(V_j^e)+exp(V^n)] \tag{2.10.3}$$

The second equation comes from rewriting (16) in terms of C_1[16]

$$\frac{1}{1-\lambda_1}\frac{1}{W}\sum_{j\in\varepsilon} M_{jn} exp(\tilde{V_n}) = \frac{1}{1-\lambda_1}\frac{1}{W}\sum_{j\in\varepsilon} \lambda_0 U f_j \frac{exp(V_j^e)}{exp(V_j^e)+exp(V^n)} \frac{(1-\lambda_1)W exp(V^n)}{\lambda_0 U}$$

$$= exp(V^n)\sum_t \sum_{j\in\varepsilon} f_j \frac{exp(V_j^e)}{exp(V_j^e)+exp(V^n)}$$

$$= C_1 exp(V^n) \tag{2.10.4}$$

Now from equations 2.3.2 and 2.3.3, we know $C_1 exp(V_j^e)$ and $C_1 exp(V^n)$

Next rewrite equation 2.3.1 by multiplying by $\frac{C_1}{C_1}$ and rearranging

$$f_j^o = f_j \frac{exp(V_j^e)}{exp(V_j^e)+exp(V^n)} \frac{1}{C_1} \tag{2.10.5}$$

$$f_j^o = f_j \frac{C_1 exp(V_j^e)}{C_1 exp(V_j^e)+C_1 exp(V^n)} \frac{1}{C_1} \tag{2.10.6}$$

$$\frac{f_j^o}{\frac{C_1 exp(V_j^e)}{C_1 exp(V_j^e)+C_1 exp(V^n)}} = \frac{f_j}{C_1} \tag{2.10.7}$$

$$\tag{2.10.8}$$

Where the left and side is known because of equations 2.3.2 and 2.3.3.

Now we can solve for C_1 using the fact that $\sum_{j\in\varepsilon} f_j = 1$

$$\sum_{j\in\varepsilon} \frac{f_j}{C_1} = \frac{1}{C_1} \tag{2.10.9}$$

[16]Where $exp(\tilde{V_n})$ has been estimated from the equation in footnore 24 of Sorkin (2008), $exp(\tilde{V_n}) = \frac{1}{N}\sum_j \frac{exp(\tilde{V_j})M_{nj}}{M_{jn}}$.

Now that C_1 is known, we can get all of the f_j. Knowing C_1 also gives us $exp(V^n)$ and $exp(V_j^e)$ via equations 2.3.2 and 2.3.3

Given the parameters of the model, choose λ_1 so that the number of endogenous EE transitions implied by the model matches what we see in the data.

$$\lambda_1 \sum_j g_j(1-\delta_j)(1-\rho_j) \sum_k f_k \frac{exp(V_k^e)}{exp(V_k^e) + exp(V_j^e)} = \sum_t \frac{\sum_j \sum_k M_{jk}}{W \sum_j g_j(1-\delta_j)(1-\rho_j)} \tag{2.10.10}$$

Chapter 3

Teachers in Politics: Teacher-Politicians, Gender, and the Representation of Public Education

3.1 Introduction

In many developing countries, government is decentralized and local politicians can have a large effect on public service provision. Furthermore, most legislators on the city council (and even some mayors) are *moonlighting* politicians, who keep their day job, and work as a legislator only part time, forming a close connection between their occupation group and the municipal government.

In Brazil, a high number of these moonlighting politicians are school teachers[1]. At least one teacher runs for city council in 83% of municipalities, and at least one is elected in 36%. Politician-teachers may have a positive effect on schools or they may have a negative effect. City councillors have oversight responsibility that could include inspecting schools and bringing deficiencies to the attention of the mayor. Teachers have an intimate knowledge of the challenges that the city's schools face, and many of them run their campaigns emphasizing the fact that they are school teachers, a possible signal that they will give particular attention to the city's schools[2]. On the other hand, corrup-

[1]School teachers are the 3rd largest occupation group among city councillors in Brazil (after farmers and business operators), making up 5% of elected city councillors across the country.

[2]In campaigns in Brazil, candidates register their candidate name with the Superior Electoral Court (TSE), which is not usually their full legal name, but usually some variation on the first name. For school teachers, 30% of them run with some variation the word *professor* (teacher) in their campaign name.

tion in the Brazilian education system is well documented (Ferraz et al. 2012), and teachers who know how the system works may have more knowledge of how to extract illicit rents, and may have the credibility and capability to mask fraudulent projects as legitamate.

In order to empirically test the effect that teachers have on the city counsel, I use a regression discontinuity design that compares outcomes from municipalities where a teacher won a seat in the city council by a narrow margin to those where a teacher lost the seat by a narrow margin. As outcomes, I use employee data from the RAIS dataset (*Relação Anual de Informações Socias*) as well as principal survey data and student test score data from *Prova Brasil*. I find that when a female teacher is elected to the city council, the city hires more teachers, hires more *qualified* teachers, and pays them more.

Not only do female teachers on the city council improve the hiring of teachers, but they have a direct effect on quality of schools. Using principal survey data, I find that having a female teacher on the city council also increases the likelihood that the city's schools have necessary teaching resources, books, and financing, and possibly increases student test scores. No statistically significant effect is found for male teachers elected to the city council.

There are two (non-mutually exclusive) possible explanations for this large difference in effects. One is that women's connection to the education system amplifies existing gender policy preference differences for men versus women (Chattopadhyay and Duflo 2004; Schwindt-Bayer 2006; Brollo and Troiano 2016). Another explanation could be that differing career concerns influence the way men and women view political representation. Men who are elected to the city council may see themselves as transitioning to a political career, whereas women elected to the city council may simply see themselves as representing the interests of the schools for a time, and then leaving politics. Thus, the male teachers who were elected would not give particular attention to the city's

schools, but the women would. I provide some suggestive evidence for this second explanation.

This research contributes to a literature that studies how political factors can effect public services, particularly education (Akhtari et al. 2017; Colonelli et al. 2017). It also contributes the literature on gender differences in politics (Chattopadhyay and Duflo 2004; Schwindt-Bayer 2006; Brollo and Troiano 2016). Though Chattopadhyay and Duflo find that gender differences drive differences in policy choice, I show evidence that while simple gender differences do account for some of the effect, the majority of the effect is coming from the occupational connection, rather than gender. The work of Brollo and Troiano bears particular relevance to my reserach, as their findings suggest that female mayors may be less likely to engage in strategic political behavior.

This paper is also related to a large literature that estimates the effect of political connections on firms (Fisman (2001), Kwaja and Mian (2005), Adhikari (2006), Claessens et al. (2008), Ferguson and Voth (2008), Li et al. (2008), Chen et al. (2011), Asher and Novosad (2017), Akcigit et al. (2018)), which finds that firms who are politically connected are favored by the government. In this paper I find that this result holds similarly for the public sector. Branches of public services that have political connections are favored by politicians.

3.2 Institutional Background

In Brazil, cities (municipalities) are governed by a mayor and a city council. While the mayor has ultimate control over implementation of government programs, the city council plays a key role in that it brings suggestions for policies and programs to improve the municipality, passes city ordinances, and must approve the mayor's budget.

Most city councils have 9 to 11 seats (95% of municipalities), but some of the very large cities have huge city councils (up to 54 people). I focus my

analysis on city councils with 9 to 11 seats, since with the large cities, adding one teacher amounts to a very small change in overall composition.

Primary education in Brazil is split into two halves, with municipal government generally responsible for the first half (years 1-5), and the state government generally responsible for the second half (years 6-9). Schools are financed through a mixture of federal, state, and municipal taxes, with federal and state governments sending monthly transfers to municipal governments to help finance schools and other public services. The control that the municipal government has on the quality of its schools is significant. Municipal budgets, including money for municipal schools, are proposed by the mayor and must be approved by the city council. Some municipal governments even install school principals as political appointees, a practice that Akhtari et al. (2017) show has a negative effect on school outcomes in Brazil.

3.3 Data

I gather electoral data from the Brazilian Superior Electoral Court (TSE), school employment data from RAIS (*Relação Anual de Informações Socias*), and other school data from Brazil's nationwide school survey and standardized test, *Prova Brasil*. The electoral data from the TSE contains the candidate's name, gender, occupation, education level, votes recieved, and whether elected. I use data from the 2004, 2008, and 2012 elections.

For the RAIS data, I use the years 2003, 2007, and 2010[3], corresponding to the electoral terms of the electoral data, plus an earlier year to use as controls. In the RAIS data I observe all teachers employed by the municipal government, their wages, education level, and hours worked per week. From this I construct 6 variables of interest for each municipality-term: average teacher hours worked per week, median monthly teacher wage, number of teachers employed, number

[3]Due to some data technicalities, data from 2011 forward is not currently available to me.

of teachers employed who have a college degree, total teacher-hours worked in the municipality, and total salary spending on teachers per municipality. Summary statistics are shown in table 1[4].

Table 3.1: Teacher Employment Summary Statistics

	count	mean	sd	min	max
Teacher Hrs/Week	12443	30.54278	8.862533	1.404255	44
Median Teacher Wage	12443	848.9286	596.3409	0	28279.08
Teachers Employed	12163	53.94466	46.74538	0	1434.015
N Teachers w/ College Degree	12163	40.44019	38.75794	0	490.8566
Tot. Teacher-Hours	12163	85265.54	77991.43	0	1515940
Tot. Teacher Salary Spend	12163	610787.8	644177.5	0	3.55e+07

Table 3.2: Principal Survey Questions

Variable	Survey Questions
Council Meets	*School council is a group made up of representatives of the school and the community, with the objective of monitoring school activities. In this year, how many times did the council meet?*
Lack Finance	*Was there a lack of financial resources at the school?*
Lack Resources	*Was there a lack of teaching resources at the school?*
Lack Books	*In this year, was there a lack of books for the students?*

For the first, respondents can answer 0, 1, 2, or 3 or more times. I code this as a binary variable with '3 or more times' indicating that the council meets. For the latter 3 variables, respondents can anwer 'No', 'Yes, but it wasn't a serious problem', and 'Yes and it was a serious problem'. I code the first at '0' and the second two as '1'.

For other school performance measures, I use three rounds (2007, 2011, and 2015) of Brazil's nationwide school survey and standardized test, *Prova Brasil*. I observe survey responses for principals, and test scores for students (4th graders) for all public schools in Brazil with at least 20 students. In the *Prova Brasil* principal survey, the school principal responds to a set of questions, a subset of which are shown in table 2. Summary statistics are shown in table 3. The student score data is standardized to have mean 0 and a standard

[4]While there are some major outliers in the right tail of the distributions of these variables, results are robust to windsorization.

Table 3.3: Principal Survey Summary Statistics

	count	mean	sd	min	max
Council Meets	5221	.4788355	.4995997	0	1
Lack Finance	5246	.5278307	.4992724	0	1
Lack Resources	5268	.3982536	.4895847	0	1
Lack Books	5209	.6680745	.4709496	0	1

deviation of 1. The data tells which level of government administers the school (state or municipal), and I focus only on municipally administered schools.

3.4 Econometric Model

3.4.1 Constructing the Running Variable

In order to estimate the causal effect of a teacher being in the city council, I use a regression discontinuity design, which will compare municipalities where a teacher barely won a seat to those where a teacher barely did not win a seat. This empirical strategy is important because municipalities that have teachers on the city council may be significantly different from municipalities without teachers on the city council, with respect to the way citizens and politicians prioritize education. This strategy isolates exogenous variation in the amount of teacher representation on the city council, which allows for causal inference.

My identification strategy follows closely that used by Poulsen and Varjão (2021), which uses detailed knowledge of Brazil's electoral system to construct a measure for margin of win/loss. A key element to understanding the construction of the running variable is the D'Hondt method, which Brazil uses to allocate seats to coalitions of parties after all the votes have been cast. In order to understand the D'Hondt method, consider the following example, taken from Poulsen and Varjão (2021).

Imagine three different coalitions are competing for 6 seats in a

fictional city council. The coalition of parties A & B receives 100,000 votes, the coalition of parties C & D receives 80,000 votes, and party E, which is running as an isolated party, receives 20,000 votes. First, the 'electoral quotient' is calculated, which is the total amount of votes cast divided by the number of seats available. In our case, the electoral quotient is $(100,000 + 80,000 + 20,000)/6 = 33,333$. Only coalitions whose raw vote count exceeds the electoral quotient are eligible to be awarded seats. Thus, party E is already disqualified from winning seats, since it only received 20,000 votes. After this, a series of quotients is calculated for each coalitition, according to the formula

$$Q_s = \frac{V}{s+1}$$

where V is the total of votes the coalition received and s is the round of calculation (or number of seats already awarded to the coalition). In an election where n seats are available, coalitions are awarded 1 seat for each quotient they have among the highest n quotients.

This is illustrated in the table below, where the coalition of A & B and the coalition of C & D have both been awarded 3 seats, since both have 3 quotients among the top 6 quotients.

D'Hondt Method Example (6 seats available)

	Q_0	Q_1	Q_2	Q_3	Q_4	Seats Won
Parties A+B	$100,000^*$	$50,000^*$	$33,333^*$	25,000	20,000	3
Parties C+D	$80,000^*$	$40,000^*$	$26,666^*$	20,000	16,000	3
~~Party E~~	~~20,000~~	~~10,000~~	~~6,666~~	~~5,000~~	~~4,000~~	0
Note: Asterisks denote quotients in the top 6. Each quotient a coalition has in the top 6 of quotients earns them a seat.						

> Because of the proportional representation system, only the last seat filled in each municipality (city) can be considered as being 'on the margin' of being flipped from one coalition to another. In the example above, that is the seat awarded to the coalition of parties C & D for their quotient of 26,666 (which was the lowest quotient of the top 6). In this case, the coalition of parties A & B was also on the margin of winning this seat, because of their quotient of 25,000 (which was the highest quotient *not* in the top 6).[5] [I] consider these two quotients (26,666 and 25,000) as the 'marginal quotients'.

Seats in the legislature are filled according to a open party list system, meaning that candidates are put in order depending on how many votes they got, and the candidates with the most votes are given seats first.

The running variable for the sharp RD design is then constructed as the margin of victory (or loss) for the coalition where a teacher was the candidate associated with the 'marginal quotient'. Specifically, for each municipality i, the running variable is formally defined as

$$R_i = \frac{Q_{i,teacher} - Q_{i,non-teacher}}{TotalVotes_i},$$

where $Q_{i,teacher}$ is the marginal quotient for the coalition where a teacher was the associated candidate on the party list (who either got a seat because of that quotient, or would have if the quotient were higher), $Q_{i,non-teacher}$ is the marginal quotient for the coalition without a teacher as marginal candidate, and $TotalVotes_i$ is the total amount of votes cast for city councillors in the election. In the example shown above, if the coalition of parties A & B had no teacher candidates, and the coalition of parties C & D did have a teacher as the marginal candidate, the running variable would then be calculated as

[5]Note that parties A & B didn't miss winning that seat by 1,666 votes, they actually missed it by 4*1,666=6664 votes, since the total votes was scaled down for purposes of the quotient.

$R_i = \frac{26{,}666-25{,}000}{200{,}000} = .0083$. Municipalities where either both of the marginal quotients are associated with a teacher, or neither of them are, are dropped from the analysis (because there is no treatment being induced by the running variable at that point).

In figure 3.1 I show a 'first stage' effect, which shows the variation in teachers on the city council that the running variable induces at the cutuff. Some municipalities in the sample have multiple teachers on the city council, which is why the average amount of teaches on the city council goes from approximately .5 to 1.5 at the cutoff. Thus, in the terminology of Cattaneo et al. (2016), I use a pooled sharp RD with cumulative cutoffs.

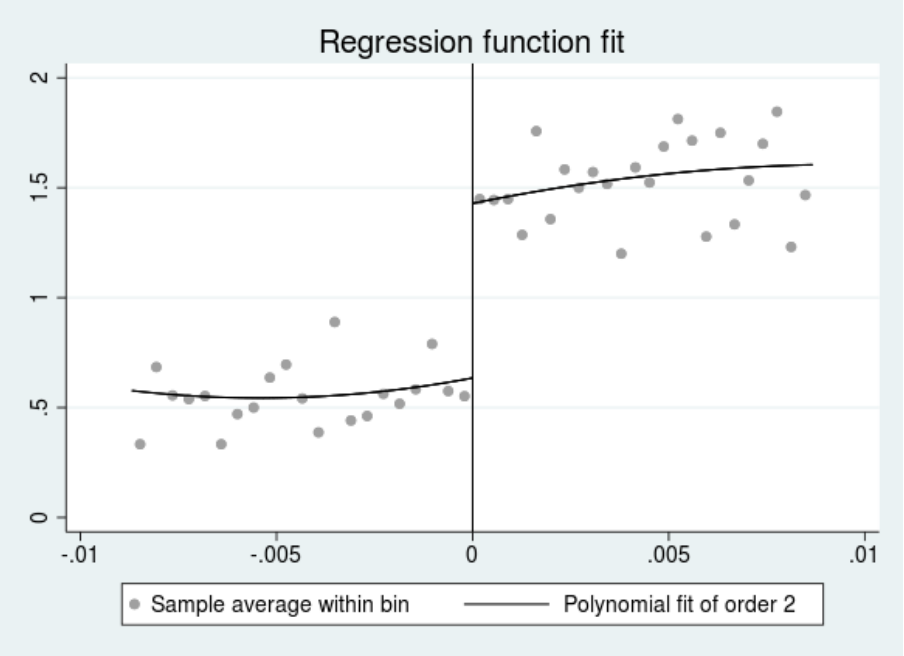


Figure 3.1: Teachers on the City Council

Notes: Some municipalities in the sample have multiple teachers on the city council, which is why the average amount of teachers on the city council goes from close to .5 to 1.5 at the cutoff.

3.4.2 Regression Discontinuity Model

With this running variable in hand, designed to measure the margin of victory or loss for teachers running for city council, I run a regression discontinuity design of the following form:

$$y_{m,t} = \beta_0 + \beta_1 T + f(R) + \beta_2 * f(R) + \beta_3 y_{m,t-1} + X_i + \varepsilon_{m,t}.$$

Where $y_{m,t}$[6] is the outcome of interest for municipality m at time t, T denotes treatment, and $f(R)$ is a flexible function of the running variable. The main coefficient of interest is then β_1. I also include in the regression a vector of municipality-characteristic controls, X_i, as well as a lagged version of the outcome variable. While not necessary for identification, these allow me to increase the precision of the estimates[7]. For all of my estimates, I use the bias-corrected and robust variance estimator from Calonico, Cattaneo, and Titiunik (2014). In my preferred specification, I use a triangular kernel and a quadratic function for $f(*)$.

3.4.3 RD Design: Smoothness and Balance Tests

Regression discontinuity designs are valuable econometric estimators because they can give nonparametric estimates of treatment effects under minimal assumptions. The key assumption for the regression discontinuity design to be valid is the 'no precise manipulation' or 'smoothness' assumption. I test this assumption using the typical McCrary (2008) density test, but using the new local polynomial density estimators put forth by Cattaneo et al. (2018). Using this test, I fail to reject the null hypothesis of no manipulation with a p-value of .60.

[6]For the *Prova Brasil* regressions, the outcome is $y_{m,t,s}$, since we observe data at the school level.

[7]Estimates are robust to the exclusion of these covariates.

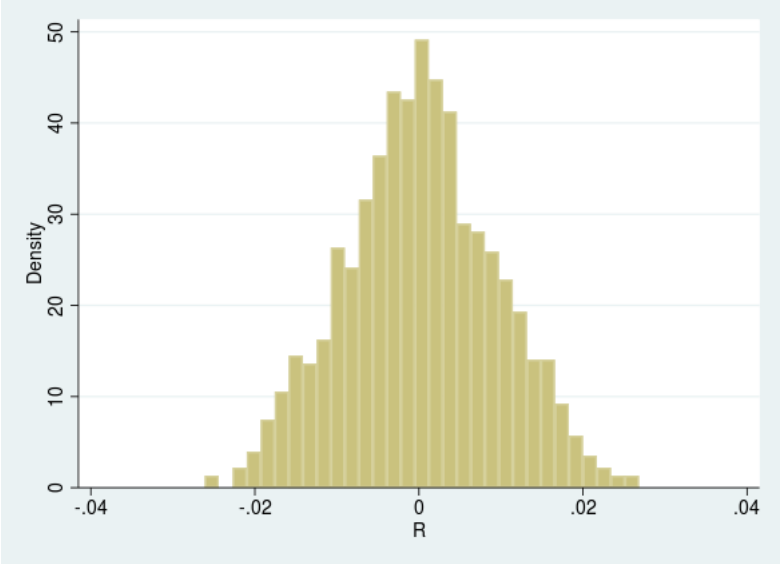


Figure 3.2: Density of the running variable (margin of victory/loss for the marginal legislator politically opposed to the mayor). Using McCrary's (2008) density test, I find no evidence of manipulation of the running variable.

I also test the smoothness assumption by doing 'placebo' tests, that is, doing the RD model with pre-determined variables as outcomes, to check covariate balance. Of 28 placebo tests, I reject the null hypothesis at the 5% level in only one case, which is roughly what one would expect from such a number of tests.

3.5 Results

3.5.1 Main RD Results

The main results of my analysis are presented in tables 5 through 7. Each table contains 3 panels, containing the estimates for the whole sample, just male teacher-politicians, and just female teacher-politicians, respectively. As can be seen, for the main variables of interest, estimates are sometimes statistically significant for the pooled sample, never statistically significant for the male sample, and always statistically significant for the female sample.

In table 5 I find that having a female teacher on the city council increases teachers weekly hours worked by 6.6 hours, increases their monthly wage by R$302, and increases teacher employment by 25 teachers per 10,000 residents.

Table 3.4: Placebo Tests

	Native State	Native Mun.	Mayor Ally	Num. Test Takers
Robust	0.0483	0.0891	0.0136	-8.035
	(0.0756)	(0.127)	(0.124)	(151.3)
Obs.	914	776	808	781
Clust.	828	716	743	721
	Coalition Size	2nd Term Mayor	Mayor Age	Metro
Robust	0.0250	-0.172	-3.320	0.0196
	(0.327)	(0.106)	(2.346)	(0.0309)
Obs.	796	799	715	778
Clust.	731	735	661	704
	AM Radio	TV Station	Internet Prov.	Pop. Growth Rate
Robust	-0.164	-0.164**	0.00788	0.0148
	(0.113)	(0.0774)	(0.149)	(0.0296)
Obs.	715	749	598	894
Clust.	661	692	561	801
	Judiciary District	Illiteracy Rate	Med. Income	Bookstore
Robust	-0.217	0.996	0.781	-0.0510
	(0.132)	(2.228)	(33.13)	(0.125)
Obs.	691	895	859	751
Clust.	644	809	781	693
	Urban Rate	Pct. < 25 yrs	Councillor Wage	Hrs Worked
Robust	0.961	1.247	25.47	-1.924
	(5.636)	(1.759)	(259.9)	(1.625)
Obs.	754	801	807	681
Clust.	696	734	738	634
	GDP Elect.	GDP Gwth, Elect.	GDP, Exam	GDP Gwth, Exam
Robust	2.696	0.0391	0.0310	-0.00217
	(2.728)	(0.0354)	(2.510)	(0.0373)
Obs.	659	955	657	612
Clust.	617	869	616	575
	Pop, Elect.	Pop. Growth, Elect.	Pop, Exam	Pop Growth, Exam
Robust	-1679.0	-0.00505	-1801.4	-0.0106
	(5780.8)	(0.0106)	(5673.5)	(0.0216)
Obs.	957	815	950	707
Clust.	870	749	865	656

Treatment effect using CCT robust confidence intervals and bias correction.
Standard errors clustered at the municipality level. All the GDP variables
are GDP per capita. Both the GDP and population variables are had for both the
election year and the year of the *Prova Brasil* survey and exam.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These amount to 21%, 36%, and 46% increases, respectively. These are large effects, and though the estimated treatment is not a large change in the composition city council, it is a large change in the representation of the city's public education system on the city council— often going from 0 to 1 representatives, and sometimes 1 to 2 representatives[8].

In table 6 I find that having a female teacher on the city council increases the number of teachers with a college degree, and also increases the total teacher-hours hired out by the municipal government, as well as the total salary spending on teachers.

For many of these variables, the coefficient for male teacher-politicians is large, but standard errors are often larger. While a failure to reject the null hypothesis is in practice not usually seen as a reason to suppose there is no effect, Abadie (2018) shows that null results often convey more information than rejection results, and thus that more weight should be given to these null results, especially when there is little prior probability of a point null. The fact there there are consistent effects for females but not males, combined with the fact that I do not find statistically significant results for even the most proximate variables a male teacher on the city council may effect (for example, teacher wage), casts doubt on any effect male teachers may have on these variables.

Moving to the outcomes from the principal survey, I find that having a female teacher on the city council increases the likelkihood that the local school council meets by 76 p.p., decreases the likelihood that a school has financial problems by 21 p.p, decreases the likelihood that the school lacks teaching resources by 49 p.p., and decreases the likelihood that the school lacks books by 29 p.p.. In this case, coefficients for men are much smaller than for women, and always statistically insiginificant.

Lastly, I test if these things improve test scores for students. While point

[8]Due to data restrictions, I cannot estimate if the specific school of the teacher-politician is favored.

Table 3.5: Teacher Employment Outcomes I

	Teacher Hrs/Week	Median Teacher Wage	Teachers Employed
Panel A: Male & Female Teacher-Politicians			
Robust	-0.331	771.7	11.10
	(2.070)	(663.6)	(11.50)
Observations	258	290	301
Clust.	247	278	288
Panel B: Male Teacher-Politicians			
Robust	-1.648	1060.2	17.51
	(2.523)	(1149.4)	(16.97)
Observations	145	174	169
Clust.	140	168	163
Panel C: Female Teacher-Politicians			
Robust	6.620**	302.2***	24.94*
	(2.723)	(88.11)	(12.86)
Observations	132	114	84
Clust.	130	112	83

Treatment effect using CCT robust confidence intervals and bias correction.
Standard errors clustered at the municipality level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Teacher Employment Outcomes II

	N Teachers w/ College Degree	Tot. Teacher-Hours	Tot. Teacher Salary Spend
Panel A: Male & Female Teacher-Politicians			
Robust	18.43*	17429.9	328752.3*
	(10.12)	(21037.5)	(173969.1)
Observations	216	309	258
Clust.	209	296	247
Panel B: Male Teacher-Politicians			
Robust	19.98	24766.6	318631.3
	(14.50)	(32629.5)	(235233.4)
Observations	145	157	177
Clust.	140	151	171
Panel C: Female Teacher-Politicians			
Robust	38.16***	47618.3**	697739.5***
	(9.902)	(20302.7)	(131113.2)
Observations	84	89	96
Clust.	83	87	94

Treatment effect using CCT robust confidence intervals and bias correction.
Standard errors clustered at the municipality level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: School Quality Outcomes

	Council Meets	Financial Problems	Lack Resources	Lack Books
Panel A: Male & Female Teacher-Politicians				
Robust	0.203	-0.141	-0.205**	0.0422
	(0.136)	(0.0984)	(0.0880)	(0.0932)
Observations	1361	1322	1650	1433
Clust.	357	345	422	370
Panel A: Male Teacher-Politicians				
Robust	0.0565	-0.0537	-0.0645	0.153
	(0.151)	(0.121)	(0.124)	(0.105)
Observations	783	824	675	750
Clust.	198	209	161	187
Panel A: Female Teacher-Politicians				
Robust	0.762***	-0.231*	-0.494***	-0.290**
	(0.115)	(0.124)	(0.107)	(0.141)
Observations	392	530	541	414
Clust.	111	152	154	123

Treatment effect using CCT robust confidence intervals and bias correction.
Standard errors clustered at the municipality level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

estimates are similar for the 3 different samples, the estimate has statistical significance only for the female regression, suggesting a .26 standard deviation improvement in language scores, significant at the 10% level.

These results shows that having a female teacher on the city council not only improves teacher hiring, but also directly improves conditions at school, and likely student learning outcomes as well.

3.5.2 Gender Effect Test

Previous economics research shows inherent differences in gender policy preferences (Chattopadhyay and Duflo 2004; Schwindt-Bayer 2006), thus, one might argue that the effects that I estimate in the tables above are merely gender effects rather than 'teacher' effects. I test this by rerunning my RD model with the main outcomes of interest, and rather than define treatment as having a teacher win a seat, I define treatment as having any woman win the marginal seat in the legislature. Results are reported in table 9. Of the 10 variables in

Table 3.8: Student Test Scores

	Language Grade	Math Grade
Panel A: Male & Female Teacher-Politicians		
Robust	0.149	0.122
	(0.177)	(0.178)
N	2762	2495
Clust.	543	484
Panel B: Male Teacher-Politicians		
Robust	0.230	0.191
	(0.261)	(0.261)
N	1435	1404
Clust.	261	251
Panel C: Female Teacher-Politicians		
Robust	0.260*	0.228
	(0.137)	(0.142)
N	929	892
Clust.	203	195

Treatment effect using CCT robust confidence intervals.
Standard errors clustered at the municipality level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

my main analysis where I find statistically significant effects for female teacher-politicians, only 2 are statistically significant at the 5% level, with 4 being statistically significant at the 10% level. Even for those estimates that are statistically significant, the point estimates are only a fraction of the magnitide of my main point estimates. Thus, while gender can explain part of the estimated effects above, it only explains a small part of it.

3.5.3 Explanation of Gender Heterogeneity

Next I attempt to offer an explanation of why results differ so much for men versus women. One explanation is that this is simply an amplification of existing gender effects. Women may prefer strengthening the education system over other public policies, and this would likely be amplified if that woman is also a teacher. Another possible explanation has to do with the political career prospects of incoming legislators on the city council, for men versus women. In table 10 I find that women are 3.6 p.p. less likely to rerun for office in the

Table 3.9: Gender Effect Test

	Teacher Hrs/Week	Median Teacher Wage	Teachers Employed	
Robust	1.421	40.85	21.95*	
	(1.819)	(65.90)	(11.60)	
Obs	610	678	630	
Clust.	586	648	603	
	Teach. w/ Degree	Tot. Teacher-Hours	Tot. Teacher Salary Spend	
Robust	13.78*	30529.9*	247564.1*	
	(7.280)	(16604.5)	(145539.7)	
Obs	630	620	681	
Clust.	603	594	651	
	Council Meets	Financial Problems	Lack Supplies	Lack Books
Robust	0.0681	-0.0623	-0.0968**	-0.141**
	(0.0725)	(0.0700)	(0.0456)	(0.0660)
Obs	3333	3444	3922	2887
Clust.	953	985	1102	822

Treatment effect using CCT robust confidence intervals and bias correction.
Standard errors clustered at the municipality level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

city council (a 6.6% difference) and 1.8 p.p. less likely to run for mayor (a 37% difference in probability). While this is only suggestive, it may be that men who are elected may see themselves as transitioning to a political career, leaving their education career behind, whereas women may still see themselves principally as school teachers, but also as temporary representatives of the education system in the municipal government. This explanation is also consistent with Brollo and Troiano's (2016) interpretation, that women are less likely to engage in politically strategic behavior— because they aren't as motivated by career concerns.

Table 3.10: City Counselor Career Prospects

	Rerun for City Council	Run For Mayor
Mean	.538	.048
Female	-0.0358***	-0.0179***
	(0.00833)	(0.00449)
Observations	9155	9155

For elected city counselors, this table shows how gender is correlated with future political career prospects.
Year of first election fixed effects included for both estimates.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6 Conclusion

This paper shows that there can be large benefits to public education when (female) teachers get involved in politics. Because the RD model compares municipalities with similar level of 'teacher-politician preference', we can conclude that having a female teacher on the city council causes significant improvements in the city's education system, ranging from teacher wages and qualifications to school supplies, finances, and test scores. Having a male teacher on the city council seems to have no such effects. The different effect by gender has a few possible explanations, but further work is required to fully understand them.

3.7 References

Abadie, Alberto. Statistical non-significance in empirical economics. No. w24403. National Bureau of Economic Research, 2018.

Akhtari, Mitra, Diana Moreira, and Laura Trucco. "Political Turnover, Bureaucratic Turnover, and the Quality of Public Services." (2017).

Brollo, Fernanda, and Ugo Troiano. "What happens when a woman wins an election? Evidence from close races in Brazil." Journal of Development Economics 122 (2016): 28-45.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. "Robust nonparametric confidence intervals for regression-discontinuity designs." Econometrica 82.6 (2014): 2295-2326.

Cattaneo, Matias D., et al. "Interpreting regression discontinuity designs with multiple cutoffs." The Journal of Politics 78.4 (2016): 1229-1248.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. "Simple local polynomial density estimators." arXiv preprint arXiv:1811.11512 (2018).

Chattopadhyay, Raghabendra, and Esther Duflo. "Women as policy makers:

Evidence from a randomized policy experiment in India." Econometrica 72.5 (2004): 1409-1443.

Colonnelli, Emanuele, Edoardo Teso, and Mounu Prem. "Patronage in the allocation of public sector jobs." (2017).

Ferraz, Claudio, Frederico Finan, and Diana B. Moreira. "Corrupting learning: Evidence from missing federal education funds in Brazil." Journal of Public Economics 96.9-10 (2012): 712-726.

McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test." Journal of econometrics 142.2 (2008): 698-714.

Poulsen, Alexander, and Carlos Varjão. "Political Opposition, Legislative Oversight, and the Performance of the Executive Branch" (2018).

Schwindt-Bayer, Leslie A. "Still supermadres? Gender and the policy priorities of Latin American legislators." American Journal of Political Science 50.3 (2006): 570-585.