The Local Political Economy of Conditional Cash Transfers in Brazil

by

Anderson Frey

B.A., University of São Paulo, 2002
M.A., New York University, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Economics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October, 2016

Abstract

Conditional cash transfers (CCT) are currently one of the most popular poverty reduction policies worldwide. Nevertheless, there is still limited evidence of their impact on the local political dynamics of developing nations plagued by corruption, clientelism, and vote buying. This thesis studies this issue using data from Brazil's *Bolsa Família*, the largest CCT program in the world.

Chapter 2 proposes a theoretical mechanism to illustrate how a CCT program could affect local politics in such institutional setting, when local politicians cannot manipulate program eligibility. In a nutshell, when mayors are able to buy votes by diverting public resources into private payments to swing voters, a CCT program works as an income shock that reduces the voter's marginal utility coming from vote buying.

Chapter 3 uses data from Brazil's *Bolsa Família* to test the theoretical predictions from the first Chapter. It shows that, when transfers are shielded from the influence of political intermediaries, they trigger a reduction in incumbency advantage, an increase in both political competition and the quality of candidates, and a reduction in the support for high-clientelism parties. Transfers also lead incumbents to shift spending toward redistributive health and education services. These results are estimated with a nonparametric multivariate regression discontinuity design.

Despite of the improvements brought out by the CCT program, when politicians can control access to the program, some of these political impacts might be negative in the short-term. Chapter 4 tells this other side of the story: what happens when politicians are able to manipulate program enrollment. Using administrative data from the *Bolsa Família* registry, and a regression discontinuity design in close elections, this Chapter shows that mayors with reelection incentives are more likely to promote income underreporting fraud by households, for the purpose of eligibility to CCT. This fraud is rewarded by voters, as corrupt mayors have a higher reelection probability. Finally, the Chapter also shows the need for disciplining devices to reduce this type of corruption. A higher risk of audits by the federal government is shown to drastically reduce the effects of reelection incentives on fraud.

Preface

This dissertation is original, unpublished, independent work by the author, Anderson Frey.

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I am very grateful to Francesco Trebbi for the advice, encouragement, and constructive criticism at all stages of this project. I would also like to thank my other committee members, Siwan Anderson, Patrick Francois and Thomas Lemieux, for their invaluable contributions.

My research also benefited from conversations with numerous faculty members and seminar participants at the UBC economics department. Although they are too many to list here, I would like to single out David Green and Thorsten Rogall. I am also grateful to my fellow PhD students for their help and advice at multiple stages of this project, and their companionship during this journey, especially Oscar Becerra, Nathan Canen, Thomas Cornwall, Hugo Jales, Chad Kendall, and Rogério Santarrosa.

I am indebted to my family, for their longstanding support, and most of all, I am grateful to my wife Michelle, who always provided unconditional love, support and encouragement, and never once doubted our decision to embark on this path together, even when I did.

Soli Deo Gloria.

Chapter 1

Introduction

Conditional cash transfers (CCT) programs, present in more than 40 countries, are now one of the leading global poverty alleviation policies. These programs have been successful in providing a stable income for the very poor, provided that they comply with conditions related to school attendance and regular doctor visits (Fiszbein et al., 2009). What is often an overlooked but important feature of CCT programs, is that many were designed to prevent capture by political intermediaries in both the distribution of resources and selection of beneficiaries.[1]

Local political leaders in many developing democracies have historically been able to reap electoral rewards from the capture of public resources. The problem is exacerbated when the practice is fueled by funds transferred from foreign or higher levels of government, as in the case of CCT programs. Capture can take many forms, from undeserved credit claiming to clientelism and vote buying.[2] While shielding transfers from the influence of local political manipulation is at the core of most CCT designs, there is still limited evidence of the long-term impact of these programs on local political dynamics.[3] If this feature of the CCT design is successful, would it

[1]De La O (2015) provides a qualitative discussion on the potential impact of CCT program design on clientelism.

[2]For the purpose of this study, the terms clientelism and vote buying are used interchangeably to mean the promise of a private transfer to a select group of voters, using public resources, with the expectation that it will influence electoral support. Stokes (2005) also uses the term machine politics to refer to this political exchange. The relationship between clientelism and economic development is discussed in the following reviews: Boix and Stokes (2009) and Hicken (2011). It is also a constant in the literature examining clientelism, in both political sciences and economics, such as: Brusco, Nazareno, and Stokes (2004); Vicente and Wantchekon (2009); Fujiwara and Wantchekon (2013); Anderson, Francois, and Kotwal (2015); Cruz, Keefer, and Labonne (2015).

[3]At the national level, it is reasonable to expect that such programs would deliver positive

reduce the power and influence of local clientelistic networks? Would it make the poor less vulnerable to vote buying? How would it change the electoral process, the profile of successful politicians, and the local public good distribution? would politicians try to manipulate the program resources using fraud? Would this be effective?

This thesis answers such questions by examining the political impacts of a CCT program in local (municipal) Brazilian governments. The country provides an unique context for this analysis, for a few reasons. First, Brazil runs the largest CCT program in the world, *Bolsa Família* (BF), reaching 13 million families. Second, local politics are characterized by clientelism and pork barrel spending across all levels (Alston and Mueller, 2006; Fried, 2012), and vote buying is abundant.[4] Third, different from most other CCT experiences, BF has simple eligibility rules and self-enrollment. In addition to enrolling themselves, households know exactly why they were eligible and where the funds come from,[5] and they are paid without intermediaries. Fourth, the program enrollment is conducted by local offices. This allows politicians to defraud the income reporting process, and to claim credit for CCT funds transferred to some households.

Within this institutional context, this thesis analyzes the local political impact of the CCT program from two different angles. In the first two Chapters, it examines the side of the story in which the CCT design succeeds, and the arrival of transfers is effectively shielded from both local political manipulation and credit claiming.

electoral rewards to federal governments that implemented. This is a topic that has been more extensively explored in the literature. For example, Manacorda, Miguel, and Vigorito (2011); Baez et al. (2012); De La O (2013) and Zucco (2013) show that this is the case in Uruguay, Colombia, Mexico, and Brazil, respectively.

[4]More than 450 mayors were impeached due to the practice between 2000 and 2009. Sugiyama and Hunter (2013) show that, in three poor Northeastern municipalities, 66% of the survey respondents were aware of quid pro quo offers for votes, and 28% reported having received recent offers.

[5]The BF program has been aggressively marketed as a federal program, thereby limiting the ability of local politicians to claim credit for its implementation. This is the most recognized welfare program in the country.

These Chapters show that transfers have a positive and long-lasting impact on the local political dynamic. Higher CCT coverage reduces the incumbency advantage for mayors, increases both electoral competition and the quality of the candidates, weakens support for parties associated with clientelism, and prompts an increase in redistributive (health care and education) spending.[6] These effects are observed for up to nine years after the start of the program.

In Chapter 2, I propose a probabilistic voting model of vote buying to illustrate the mechanism by which the CCT program affects local politics in the long term. In this model, incumbent mayors are able to buy votes by diverting public resources into private payments to impoverished swing voters. The CCT program, in this context, is a positive income shock that reduces the voter's marginal utility coming from these private payments. This leads to a reduction in the incumbency advantage, and a consequent increase in electoral competition. As vote buying becomes less attractive, incumbents try to recover the lost votes by funneling resources back into public goods.

Because of the manipulation opportunities arising from the CCT design in Brazil, coverage can not be considered exogenous to local politicians. Chapter 3 tests the model predictions empirically, using a multivariate regression discontinuity design (MRDD) to ensure the exogeneity of the municipal CCT coverage. It relies on the part of the cross-municipal variation in coverage generated by the Family Health Program (FHP), a household-based health program run by municipalities, and funded by the federal government. The intuition is simple: the FHP funding is discontinuous, and exogenous to mayors.[7] The FHP health-care teams had capillarity and penetration among the poor, and were able to better spread information

[6]In Brazil, public health and education are available to all, but are consumed primarily by the poor. The wealthy and middle class are much more likely to use private alternatives. Given this context, a shift in spending toward these categories can be seen as redistributive (Fujiwara, 2015).

[7]Municipalities with populations below 30,000, and a human development index (HDI) below 0.7, are eligible to receive 50% additional federal funding for the FHP.

about BF in the early stages of the CCT program. More FHP funding lead to better information, which in turn created differentials in CCT enrollment that were able to persist over time due to the program rules.

Together these Chapters find the following: the additional FHP funding generates a 7.9 percentage points (pp) higher CCT coverage.[8] A 10pp increase in CCT coverage would trigger a 8.2pp vote loss for the incumbent, which is mostly driven by the effect on politicians that belong to high-clientelism parties. Higher CCT coverage also prompts a 6.3pp fall in the margin of victory, 0.4 more candidates in the mayoral race, 50% less private campaign donations to incumbents, and a 24pp increase in the share of competitive challengers having at least a high school education. With respect to budget allocation, there are increase of 2.3pp and 2.7pp in the shares of budget allocated to education and health care; along with a 1.5pp decrease in the share allotted to capital investment.

There is, however, another side to this story. Chapter 4 explores one way in which mayors can manipulate and claim credit for the CCT resources: income reporting fraud, which is measured using the administrative registry of the BF program. I estimate that at least 8 million households are currently receiving BF benefits by means of underreporting their income.

In a nutshell, Chapter 4 shows three things. First, using a regression discontinuity design on close elections, it demonstrates that the possibility of reelection for mayor creates incentives for fraud, generating expenses with the program that are nearly 7% higher in these municipalities. This gives mayors ample ability to claim credit over the distribution of resources that are five times the average campaign spending in local elections. Second, using random program audits by the federal government, it shows that a higher expectation of being audited significantly reduces the likelihood of mayors promoting fraud. Third, using matching over 40

[8]For municipalities having at least 25% of poor families

observed characteristics of municipalities, parties and mayors, it shows that fraud helps incumbents to be reelected, especially in more competitive political environments.

Both the approach of the first two Chapters, and the approach of Chapter 4, estimate the same thing: the effect of higher CCT coverage on local politics. However, while Chapter 2 and 3 show that CCT coverage reduces incumbency advantage and increases electoral competition, Chapter 4 shows that incumbents can benefit from the program. These results can be re-conciled after considering the nature and the persistence of each effect, in the context of Brazil. The negative effects shown in Chapter 4 are characteristic of the period when beneficiaries are joining the program, and the manipulation of eligibility by politicians can create credit claiming opportunities. Similar results already exist in the literature (De Janvry, Finan, and Sadoulet, 2012; Labonne, 2013). However, given the program characteristics (long permanence, easy access to the central government office for complaints), and the local political environment (two-term limit for mayors), these effects are likely short-lived.

In Chapter 2 and 3, the effects are not only estimated for a longer period, but they also represent the lasting impact of the CCT program. They only exist if beneficiaries see the CCT income as permanent, in a way that would change their consumption decisions, and utility. Accordingly, these effects should last as long as the CCT benefit is stable, and should survive through different political cycles and incumbents (they do not concern the relationship between voters and one specific politician, i.e., the one that might have given them access to CCT). In other words, after a short-period in which credit can be claimed by politicians for getting the program benefits to some voters (Chapter 4), the long-term effects of the CCT program that come through the impact of the permanent income increase on vote buying should be positive (Chapters 2 and 3). Although the different empirical techniques employed in this study do not all me to compare the magnitude of the opposing effects estimated in Chapters 3 and 4, a back of the envelop calculation indicates that the long-term (positive) effects are at least four time larger than the short-lived (negative) effects of CCT on local politics.

This thesis contributes to five main strands of the literature pertaining to both economics and political science. First, it takes into account the extensive literature that looks at various aspects of vote buying around the world.[9] A constant throughout this literature is the association between vote buying and poverty.[10] Chapter 2 and 3 propose and test a plausible channel through which income transfers can reduce vote buying. Chapter 2 also addresses the literature concerned with the mechanisms that support vote buying where there is imperfect monitoring of voters. Stokes (2005) argues that politicians can partially monitor voters with the help of social networks, and both Nichter (2008) and Hidalgo and Nichter (2015) show that vote buying can actually be turnout buying under weak monitoring. The model proposed here provides theoretical support for vote buying of swing voters in secret-ballot elections.

Second, the literature on the political effects of CCT is generally concerned with national politics (the level at which the program is implemented). A number of papers point out that the electoral support for the incumbent increases following a CCT implementation.[11] Labonne (2013) and Rodriguez-Chamussy (2015)[12] ex-

[9]Finan and Schechter (2012) show the role of reciprocity in the way voters are targeted for vote buying in Peru. Anderson, Francois, and Kotwal (2015) discuss the role of land ownership and cast relations in clientelism in India. Brusco, Nazareno, and Stokes (2004) discuss the role of discount rate in vote buying in Argentina. Sugiyama and Hunter (2013) show evidence that public goods are used in clientelistic exchanges in Brazil. Fujiwara and Wantchekon (2013) and Cruz, Keefer, and Labonne (2015) discuss the role of information in the dynamics of vote buying in Benin and the Philippines, respectively. Larreguy, Marshall, and Trucco (2015) show how institutional changes in a urban land tilting program in Mexico reduced clientelism.

[10]The causes and nature of this relationship are reviewed in Hicken (2011).

[11]In Uruguay (Manacorda, Miguel, and Vigorito, 2011), in Colombia (Baez et al., 2012), in Mexico (De La O, 2013), and in Brazil (Zucco, 2013)

[12]Unlike this thesis, Rodriguez-Chamussy (2015) focuses on party incumbency rather than candidate incumbency, given that mayors cannot be reelected in Mexico.

amine the impact of CCT on incumbency advantage using randomized program roll-outs in the Philippines and Mexico, respectively. Unlike the results presented in Chapter 3, both studies find that incumbents benefit from higher coverage, most likely because mayors are able to claim part of the credit for the introduction of CCT. De Janvry, Finan, and Sadoulet (2012) examine an older Brazilian CCT program (*Bolsa Escola*), in which mayors could claim credit for the policy (they selected the beneficiaries). The authors show that better program performance was associated with higher incumbency advantage. This thesis provides a comprehensive picture of the way in which cash transfers affect a variety of local political outcomes, in a context where the ability of mayors to claim credit for the program is limited.

Third, by extending the multivariate regression discontinuity design (MRDD) method proposed by Zajonc (2012), and by using a formal approach to bandwidth selection, Chapter 3 departs from the strategies commonly used for the estimation of regression discontinuity designs based on more than one running variable. Also, to my knowledge, this is the first study that uses the heterogeneity of the average treatment effect generated by a RDD to inform an instrumental variable regression.

Fourth, this thesis extends the literature on the relationship between policy, political institutions and corruption in developing countries, with an unique and comprehensive contribution. It shows, both theoretically and empirically, the multiple impacts of a policy design (the CCT program) on local politics from two different perspectives. The one in which the policy succeeds in its intent to reduce clientelism and political manipulation, and the other in which it fails, thereby increasing fraud and the electoral rewards of corruption.[13] Moreover, by showing evidence of the impact of disciplining devices on corruption, it contributes to a still sparse literature

[13] Specifically in the case of Brazil, these are a few examples of related articles that examine the relationship between corruption, institutions and elections: Brollo et al. (2013) show that corruption increases after a resource windfall to municipalities; Ferraz and Finan (2011) show how reelection incentives determine corruption; and Fujiwara (2015) shows that a change in voting technology increased electoral turnout among the poor and prompted more redistributive health care policies.

(Pande and Olken, 2012).[14]

Finally, Chapter 4 contributes to the political accountability literature by showing how reelection incentives can increase corruption in some cases, and how corruption can, in turn, help incumbents to be reelected. These outcomes are opposite to the prediction given by the agency models of rent seeking,[15] because this specific type of fraud is aligned with both the preferences of voters and the politician's objective. Camacho and Conover (2011) also examine politically-motivated fraud in a CCT program (Colombia), and in line with this Chapter, they find that fraud is higher in environments of more political competition. But given that mayors have a one-term limit in Colombia, that paper focuses on the mechanics of program manipulation, as opposed to the incentives for fraud (e.g. reelection and audit risk), or the electoral rewards of fraud.

[14]Again in the case of Brazil: Ferraz and Finan (2008) show that the information about corruption audits, when disseminated, reduced incumbency advantage; and Lichand, Lopes, and Medeiros (2016) show that the expectation of future audits reduced corruption in health services.

[15]See Barro (1973); Ferejohn (1986), and more recently Duggan and Martinelli (2016), for the theory. Ferraz and Finan (2011) provide an empirical application in Brazil.

Chapter 2

A Probabilistic Voting Model of Vote Buying

The motivation for starting this study with a theoretical framework is threefold. First, the literature currently lacks a proper theory that illustrates how a CCT program could affect multiple local political outcomes in a context of clientelism and vote buying, and where the program design successfully shields the transfers from local political capture.[16] Accordingly, this model provides a well structured narrative to connect all the empirical results to be presented in Chapter 3.

Second, it incorporates in the theory some features of vote buying in the local context. Vote buying in Brazil characteristically involves the use of public resources, the targeting of the poor, the payment made in more than one installment, and the high risk of impeachment for mayors.[17] It is also abundant, even in the presence of secret-ballot elections. Sugiyama and Hunter (2013) show the results from a survey where 28% of respondents received recent vote buying offers (Northeastern Brazil). They also show that government programs are highly associated with vote buying. A survey conducted by a local polling institute after the 2004 elections (IBOPE, 2005) shows that vote buying is more common among the poor and, in at least 67% of the cases, includes offers of public goods and services.

[16] In the probabilistic voting models proposed by both Labonne (2013) and Camacho and Conover (2011), local incumbents are thought to fully claim credit for CCT transfers coming from the central government, and vote buying is not included.

[17] The following examples of media coverage (in Portuguese) provide insights on the mechanics of this practice: *http://goo.gl/xnF3LV*; *http://glo.bo/1urEBhD*; and *http://goo.gl/r0lWzm*.

Third, it also provides one plausible solution for what I call the secret-ballot puzzle. The literature on vote buying has tried to understand how and why the practice is so widespread in the context of poverty, even in the apparent absence of a commitment device, i.e., when monitoring of voters is limited by secret-ballot elections. Several explanations exist for this phenomenon, such as the partial monitoring of voters using polling station data (Querubin and Snyder, 2013), or social networks in poor communities (Finan and Schechter, 2012; Boix and Stokes, 2009).

Also, part of the literature has claimed that the prevailing practice under secret-ballot elections is turnout buying. In this case, instead of buying the support of voters that would otherwise vote for another candidate, politicians pay their own supporters to go vote (Nichter, 2008). In Brazil, however, voting is mandatory and turnout is already high (more than 80%), so even the manipulation of turnout is made difficult.[18] In this proposed model, incumbents signal their promise of post-election benefits to targeted voters, if reelected, by paying a first installment before the election. They will commit to honor their promises for fear of being exposed (and prosecuted) by the illegal practice of vote buying. Targeted voters, in turn, have incentives to vote for the incumbent for fear of not being targeted for private transfers in the future election cycle (by the new candidates).

Accordingly, the model intuition is as follows. Incumbent mayors can reap electoral rewards from converting public resources intended for the poor into private payments (vote buying) to swing voters.[19] Vote buying here consists of a cash payment[20] accompanied by a promise that, on the condition that the incumbent is reelected, the payment will be repeated. Because vote buying is illegal and punish-

[18](Hidalgo and Nichter, 2015) shows that local politicians can indeed buy turnout in Brazil by buying voters form other municipalities (paying them to migrate).

[19]The targeting of swing voters is in keeping with the vote buying model proposed by Stokes (2005), and the framework for redistributive policy proposed by Dixit and Londregan (1996).

[20]These payments may also come in the form of goods and services redirected from public resources (e.g. medicine, food, cement). These resources are perceived as close substitutes for cash given their immediate necessity to the poor voters, who would at any rate have used cash to buy such goods and services.

able with impeachment, incumbents have incentives to pay the second installment if reelected. The swing voters targeted today, might not be targeted in a future election by the new candidates, so they have an incentive to maximize the reelection probability of the incumbent in order to receive their second installment. As long as a group of swing voters can be identified in the population, this mechanism makes it is optimal for the incumbent to pay voters without monitoring.

In this context, cash transfers represent a positive income shock to the poor that reduces the marginal utility of these vote buying payments.[21] With CCT, vote buying becomes a less effective electoral strategy, and it falls. Accordingly, the incumbency advantage is also reduced, given that it comes from the incumbent's unique ability to target swing voters with public resources. As a consequence, incumbents channel resources back into pro-poor public goods. I also show that, if politicians have a type that makes them more effective in vote buying than in public good distribution, then higher-type politicians (i.e. relatively better at vote buying) will lose more votes, and be less able to shift resources towards public goods.

2.1 Basic Model Setup

Consider a two-period model, where there is an incumbent politician and a challenger, indexed by $P = \{I, C\}$. In period 1, the incumbent carries out policy, and faces a challenger in the election at the end of the period. There is a two-term limit, so only the challenger could potentially be reelected in the election taking place at the end of period 2.

[21]The mechanism here relies on CCT programs having a significant impact on the consumption standards of the targeted populations. In the case of Brazil, this is expected due to the significant increase in income provided by the program (~50%), and its long term existence. Fiszbein et al. (2009) also show that this is the case for most CCT programs around the world. They also show that, in many cases, CCT beneficiaries tend to treat the transfer as permanent income in their consumption decisions.

There are three groups of voters, indexed by $J = \{H, L, S\}$.[22] The groups differ in size, income, and in the distribution of political preferences. The share of group J in the population is given by α^J. Group H is comprised of wealthy voters with income w. Groups L and S are comprised of poor voters, with income $y < w$. The parameter σ^{Ji} denotes the preference of voter i in group J for the challenger. This parameter has a group-specific, uniform distribution on $(-\frac{1}{2\phi^J}, \frac{1}{2\phi^J})$, with density ϕ^J. Group S is the "swing" group. When compared to L, S has a higher density of political preferences ($\phi^S > \phi^L$), and is smaller ($\alpha^S < \alpha^L$). Define $\phi^S = z\phi^L$, with $z > 1$. The intuition here is that, within the larger group of poor voters, incumbents can observe a group of voters that have close-to-neutral political preference for the current candidates, and are an attractive target for vote buying. Finally, the challenger's relative popularity in the entire population is given by λ, with an uniform distribution on $(-\frac{1}{2\psi}, \frac{1}{2\psi})$.

Political preferences here can reflect any dimension in which the candidates differ from each other, apart from policy. These characteristics of each candidate are observable and cannot be altered. While group income, size, and density remain constant, individual political preferences (σ^{Ji}) are redrawn for every electoral cycle, given the change in candidates. The poor voters belonging to groups L and S are therefore different in each race. As an example, an individual with strong preference for candidate A over candidate B is allowed to have a strong preference for candidate C over A in a different race. Also, a voter with $\sigma^{Li} > (1/2\phi^S)$ in period 1 would not be eligible to be part of group S. However, in period 2 every individual voter has an equal probability being in the swing group, which is given by $\alpha^S/(\alpha^L + \alpha^S)$. All the group parameters are exogenous and, while politicians observe these parameters, they do not observe individual preferences.

[22] This is in the spirit of the probabilistic voting model of Persson and Tabellini (2000)

2.1.1 Probability of Reelection

Voters have utility, under politician P, coming from public goods (g_P), and private consumption (c_P), which includes their income and private payments from politicians. I will use a log utility function, separable on the consumption of public and private goods, as $U(g_P, c_P) = \delta \log(g_P) + (1-\delta)\log(c_P)$. This functional form for the utility was chosen so the model can be solved explicitly. The main assumption behind this choice is that public and private consumption enter the utility independently. I provide an appendix with the results for a generic concave utility function to show how a more flexible functional form would affect the main model predictions.

Voters choose the candidate that is expected to provide the highest utility in period 2, taking into account their political preference, and the challenger's relative popularity. An individual i will vote for the incumbent if: $U(g_I, c_I) - U(g_C, c_C) - \lambda - \sigma^{Ji} > 0$. In which case, every voter with $\sigma^{Ji} < U(g_I, c_I) - U(g_C, c_C) - \lambda$, votes for the incumbent. The incumbent's vote share is:

$$\pi_I = \sum_J \alpha^J \phi^J [U^J(g_I, c_I) - U^J(g_C, c_C) - \lambda + \frac{1}{2\phi^J}] \tag{2.1}$$

With $\phi = \sum_J \alpha^J \phi^J$, and under the model's distributional assumptions, the probability that the incumbent is reelected is:

$$Prob.[\pi_I > \frac{1}{2}] = \frac{1}{2} + \frac{\psi}{\phi} \sum_J \alpha^J \phi^J [U^J(g_I, c_I) - U^J(g_C, c_C)] \tag{2.2}$$

This is the objective function to be maximized by the politicians. I assume that they do not extract rents from office other than ego rents, in which case they care only about the reelection probability. Later in this Chapter, I discuss the implications of including rent seeking in the model.

2.1.2 Budget Allocation

The budget b is exogenous, and fixed across time and between politicians. It can be allocated across two types of public goods: one that is consumed by the poor groups (t_P), and one that is consumed by the wealthy group ($b - t_P$).[23] Politicians can also allocate resources to vote buying, in the form of cash payments to the swing group. Because they are unable to observe individual political preferences, the amount of vote buying resources (v_P) is divided equally among all voters in the group. These funds are diverted from pre-existing allocations to pro-poor and pro-wealthy public goods, with exogenous shares s and $1 - s$. If N is the total number of voters, and $\eta = 1/(\alpha^S N)$, the utilities of the groups of voters under politician P are:

$$H : U(g_P^H, c_P^H) = \delta \log(b - t_P - (1 - s)v_P) + (1 - \delta) \log(w)$$
$$L : U(g_P^L, c_P^L) = \delta \log(t_p - sv_p) + (1 - \delta) \log(y)$$
$$S : U(g_P^S, c_P^S) = \delta \log(t_p - sv_p) + (1 - \delta) \log(y + \eta\theta_P v_P)$$

Candidates also differ in their ability to deliver vote buying payments, denoted by θ_P. While t_P and v_P are decision variables, this ability is exogenous, fixed over time, and observable by voters. It enters the model as a multiplier to v_P. If $\theta_P < 1$, the politician is relatively less efficient at delivering vote buying payments when compared to public goods, and vice-versa (the implicit multiplier on public goods is equal to one.). Finally, vote buying offers are only made with public resources.

[23] These assumptions reflect the Brazilian context, where municipalities receive most of their revenues from higher levels of government, and where public health and education services are available to all but are only consumed by the poor.

2.2 Vote Buying and Incumbency Advantage

Given that voters are prospective, the incumbent's policy in period 1 is only useful to them as an indication of the expected policy for period 2. Policy here is defined as the allocation of the budget across groups of voters, and types of public goods and vote buying. If the incumbent wins, he replicates the policy from period 1 due to the lack of reelection incentives.[24] The public goods allocation is maintained,[25] and the same vote buying payments are made to the same voters in period 2. Here, the period 1 payments are the first installment of a full transfer to be completed upon reelection. However, since the incumbent is not eligible in period 2, the question is: why would he honor his promises?

Vote buying is an electoral crime in Brazil.[26] The law prohibits candidates to offer goods or services to voters in exchange of votes. The candidate does not even need to explicitly demand the vote in the exchange, if there is enough evidence that the practice falls in the category of ”vote buying”. Punishment for this crime includes losing the post, and the political rights for a period of 8 years. In 2000-2009, more than 600 politicians were prosecuted and lost their tenures due to practice in Brazil. By paying the swing group the first installment, the incumbent undertook an illegal activity. Accordingly, if voters do not receive the promised amount, they may expose the practice to the electoral courts, and thereby trigger an impeachment process.

Voters will still face a competing utility given by the challenger, but challengers are not able to make credible promises without access to budget resources. In the absence of a first installment, challengers are not at risk, and are therefore in a position where they might not honor their promises. However, if they win, they will

[24]This assumption would be slightly different if rent extraction was allowed, with no change to the results.
[25]I assume that there is a very small cost of shifting policy.
[26]Article 41 of the Law 9.504/1997.

seek reelection. Thus, voters infer their expected utility coming from challengers from their allocation of public goods and vote buying in period 2, which follows the incumbent's optimization problem. The use of the illegality of vote buying in this model simplifies the commitment problem between voters and politicians, and clearly illustrates the difference in the ability of incumbent and challenger to effectively buy votes. These feature of the local electoral rules, however, might not present in other contexts, and yet vote buying is still observed around the world under similar characteristics. Modeling the specific and complex features of the bargaining problem between voters, politicians, and possibly brokers (Stokes et al., 2013), is beyond the scope of this study.

Poor voters in groups L and S expect to have the following private consumption under the incumbent: $c_I^L = y$ and $c_I^L = c_C^S = y + \frac{\theta_I v_I}{\alpha^S N}$, respectively. Because the composition of groups in period 2 is not yet known, all poor voters expect from the challenger the same private consumption: $c_C^L = c_C^S = y + \frac{\theta_C v_C}{(\alpha^S + \alpha^L)N}$. While the incumbent's allocation is more advantageous to the swing voters, the challenger's policies benefit the non-targeted poor voters. In this context, incumbency advantage exists because the incumbent is the only politician that can target the swing voters of the current electoral cycle, and their votes are more likely to define the current election than those of non-targeted poor voters.

If politicians do not differ in type, the values of t_P and v_P are the same under both candidates. For the sake of simplicity, I assume that $\theta = \theta_I = \theta_C$,[27] $t_I = t_C = t$, and $v_I = v_C = v$. In this case, it is the private consumption alone that determines the incumbency advantage. Combining the utility of private consumption under each candidate with equation 2.1, with $\epsilon = 1/((\alpha^L + \alpha^S)N)$, the vote share of the incumbent is now:

[27] This assumption will be relaxed in Proposition 6.

$$\pi_I = \alpha^L \phi^L \delta \log(\frac{y}{y + v\epsilon\theta}) + \alpha^S \phi^S \delta \log(\frac{y + v\eta\theta}{y + v\epsilon\theta}) - \lambda + \frac{1}{2} \tag{2.3}$$

According to the equation above, for high enough values of $(\eta - \epsilon)$ and $\alpha^S \phi^S$, there will exist incumbency advantage (i.e.$\pi_I > \frac{1}{2}$). The equilibrium values of pro-poor public good (t) and vote buying (v) are explicitly calculated in the appendix to this Chapter, and they show under which conditions vote buying will exist.

2.3 Model Analysis

This model aims to provide a theoretical framework for the analysis of the effect of a CCT program on incumbency advantage, budget allocation, the profile of politicians, and vote buying. Conditional cash transfers enter the model as a positive income shock to the poor. Such shock is assumed to be exogenous to local politicians, i.e., they can neither claim credit for its arrival, or manipulate its allocation, magnitude or timing. In this section, I show how CCT would change the model equilibrium, and affect the political outcomes. To respect the intent of the program, I assume that $w > y$ even after the income shock has occurred. All proofs for the following propositions are in the appendix.

Proposition 1. *The implementation of a CCT program reduces the amount of resources allocated to vote buying ($\frac{dv}{dy} < 0$).*

The marginal utility of private consumption of the poor decreases with the income shock. Given that, by assumption, there is no change in the marginal utility of consumption of public goods, it is optimal for the incumbent to reduce the amount allocated to vote buying.

Proposition 2. *The implementation of a CCT program shifts the budget allocation toward the pro-poor public good ($\frac{dt}{dy} > 0$), and away from the pro-wealthy* *public good. This happens for a low enough value of the parameter s, and only if vote buying exists.*

An income shock does not affect the marginal utility of public good consumption, so without vote buying, the incumbent would not change the policy allocation. If there is vote buying, it follows from Proposition 1 that the incumbent will channel resources back into public goods. The magnitude of the reallocation (between t and $b - t$) depends on the size and density of the poor and wealthy groups, and on s, which is the exogenous share of vote buying resources taken from the pro-poor public good allocation.[28] Incumbents increase their allocation to pro-poor goods for a low enough size (α^H) and density (ϕ^H) of the wealthy group, and for a low s. If s is too high, then the increase in the pro-poor spending will be low (or even negative), due to the direct effect of a reduction in v on the allocation to the pro-poor good.

Proposition 3. *The implementation of a CCT program reduces the vote share of the incumbent ($\frac{d\pi_I}{dy} < 0$). This happens if and only if there is vote buying, and if the swing group has a large enough density of political preferences around the neutral level.*

Without vote buying, incumbency advantage would not exist, and the CCT program would not affect vote shares. With vote buying, voters face different utilities coming from the incumbent and the challenger (equation 2.3). Given that the incumbency advantage depends on v, it will also decline after the income shock. The necessary condition is that the ratio of the densities of groups S and L ($z = \frac{\phi^S}{\phi^L} > 1$) is at least equal to the ratio of the marginal utilities of private consumption for the groups ($z \geq \frac{y+v\eta\theta}{y}$).

Proposition 4. *Both the decline in the incumbent's vote share and the shift to*

[28]The value of s is not observed in the data and cannot be estimated empirically in this study. However, anecdotal evidence points to a value that is neither close to one or zero, as surveys indicate that vote buying offers can come in various and flexible forms such as doctor visits (e.g. taken from t) or bags of cement or water (e.g. taken for $b - t$).

pro-poor public good, are larger when the political preferences of the poor are more neutral ($\frac{d\pi_I}{dyd\phi^L} < 0$ *and* $\frac{dt}{dyd\phi^L} > 0$ *). This happens if the poor group is large enough, or if it has a more neutral density of political preferences when compared to the wealthy group.*

The density of the political preferences of the poor population is given by ϕ^L ($\phi^S = z\phi^L$). A higher ϕ^L makes the poor a more attractive target for both public goods and vote buying. If group L is an attractive target for public good distribution (i.e., if it either has a large size or a high density of political preferences around zero), then the effect on the attractiveness of public goods dominates. In this case, for a higher value of ϕ^L, more resources are moved into pro-poor public goods, and the incumbent's vote loss is larger.

Proposition 5. *Both the decline in the incumbent's vote share and the shift to pro-poor public goods are larger when the poor population is large (* $\frac{d\pi_I}{dyd\alpha^L} < 0$ *and* $\frac{dt}{dyd\alpha^L} > 0$ *). This happens if the distance between the densities of the wealthy and poor groups (*$\phi^H$and ϕ^L*) is relatively small.*

A larger α^L relative to α^S increases the uncertainty of poor voters about receiving a vote buying offer from the challenger, improving the incumbency advantage. However, it also makes the poor a more attractive target for public good distribution, which works against the incumbent. When the latter effect dominates the former, then the decline in the incumbent's vote share is higher for a larger poor population ($\frac{d\pi_I}{dyd\alpha^L} < 0$). This depends on the levels of ϕ^L and ϕ^H. It suffices for the result that these densities are close in magnitude, independent of the direction. The intuition is that both wealthy and poor groups must be relatively attractive to receive the incoming resources from vote buying.

Proposition 6. *Incumbent politicians with a higher type will face a larger decline in vote shares (*$\frac{d\pi_I}{dyd\theta_P} < 0$*), and will carry out a smaller shift to pro-poor goods (*$\frac{dt}{dyd\theta_P} < 0$*).*

To test this Proposition, I allow politicians to differ in type (θ_P). A higher type means that the politician is more efficient when it comes to buying votes rather than delivering public goods. Given that transfers reduce vote buying, high-type incumbents reduce the resources invested into an electoral strategy where they are relatively more efficient, which generates the larger vote loss. Also, given that high-type incumbents are less efficient in delivering public goods, their budget shift will be smaller.

2.4 Incorporating Rent Seeking

The appropriation of public resources by politicians is a common feature of developing democracies (Pande and Olken, 2012). In this section, I allow for politicians to extract rent from the public budget b. Based on the previous model, an amount r of rents can be extracted from v and, instead of being paid to swing voters, are appropriated by the incumbent (r is a third decision variable). Assume that $\bar{r}$ is the maximum rent that can be extracted without legal consequences, or the politician being exposed for vote buying.

It is easy to see that winning incumbents always extract $\bar{r}$ in period 2, due to the absence of a commitment device after reelection. Because the utility of voters only considers the expected performance of candidates in period 2, incumbents also have no incentive to extract less rent in period 1. Thus, for both candidates, it is optimal to always have $\bar{r}$ is all periods. The new allocations for v and t are shown in the appendix. Under these assumptions, there is always a positive amount of rent being extracted, but there is no change in the model predictions for $\frac{dv}{dy}$, $\frac{dt}{dy}$ or $\frac{d\pi_I}{dy}$.

A further refinement could be made in the spirit of the agency models, where politicians have an incentive to extract less than maximum rents if it helps them to get reelected. This is the approach used in the existing corruption literature in

Brazil (Brollo et al., 2013; Ferraz and Finan, 2011). The intuition is that candidates differ in their willingness or ability to seek rents, and high-rent politicians might find it optimal to extract less in period 1, so they can hide their true type. Instead of providing a completely new theoretical framework, I illustrate this mechanism in the context of this model by assuming that politicians have an ability to extract rents uniformly distributed in $(\underline{r}, \bar{r}\ )$, unobserved by voters.

They cannot extract more rent than their true type allows. As before, reelected incumbents always extract their true rent type in period 2. When incumbents extract r in period 1, voters then update their expectation for r_I in period 2 as follows: $r_I = 0.5(\bar{r}+r)$. For simplicity, I assume that the politician's utility is linear in r. Denote the probability of reelection shown in equation 2.2 by p, and the true type of the incumbent by R. With $0 < \gamma < 1$, the utility of the incumbent is defined as $V_I = \gamma r_I + pR$. This the new equation to be maximized by the incumbent.

In the appendix I show the new equilibrium values of $\frac{dv}{dy}$, $\frac{dt}{dy}$ and $\frac{dr}{dy}$. Here, rent seeking works as another resource reallocation alternative for an incumbent reducing v. The attractiveness of this alternative is given by γ, and it is easy to see that, when $\gamma = 0$, the equilibrium conditions become similar to the ones in the base case model. As before, an income shock will reduce vote buying ($\frac{dv}{dy} < 0$), and increase the allocation to pro-poor goods for low enough values of s ($\frac{dt}{dy} < 0$). Also, the income shock causes an increase in rent extraction. This result comes from the specific assumption made regarding the utility function. If public good and private consumption were complementary in the utility, there is an equilibrium in which rents would be reduced as a result of a CCT program.

For the reelection probability, the intuition is that a higher maximum rent ($\bar{r}$) reduces the incumbent's vote share. In this case, the expected utility of voters from the incumbent in period 2 is reduced due to rent extraction (which is not the case in the utility coming from the challenger). Accordingly, the income shock would also have a lower effect on the fall in vote shares ($\frac{d\pi_I}{dy}$), as the incumbency advantage created by vote buying is already less effective with the existence of rents. All in, although the intuition does not change, the result from Proposition 3 now depends on stronger conditions on the exogenous parameters to hold (shown in the appendix).

Chapter 3

Cash Transfers, Clientelism, and Political Enfranchisement

Motivated by the previous theoretical predictions, this Chapter shows that transfers that are shielded from the influence of local politicians improve the electoral process and contribute to the political enfranchisement of the poor. As mentioned before, the *Bolsa Família* (BF) program in Brazil was aggressively marketed as a federal program, limiting the ability of local politicians to claim credit for its implementation. There are, however, two main challenges to identifying the exogenous effects of this program in the local political environment: the implementation was not randomized, and the program is jointly run by the central and municipal governments, giving mayors some leeway to manipulate the enrollment of beneficiaries.

Accordingly, this Chapter estimates the political effects of the CCT program using a novel identification strategy. It relies on the part of the cross-municipal variation in CCT coverage generated by the Family Health Program (FHP), a household-based health program run by municipalities, and funded by the central government. For this identification strategy to be effective, two things are required. First, there must exist exogenous assignment of FHP funding or coverage. Accordingly, since August 2004, municipalities with population below 30,000, and a human development index (HDI) below 0.7, are eligible to receive 50% additional federal funding for the FHP. This discontinuous assignment of program funding allows this study to instrument local CCT coverage with this funding differential, using a multivariate regression discontinuity design (MRDD).

Second, it is required that the FHP funding has a strong impact on the CCT coverage (first stage), and no meaningful direct effect on political outcomes (exclusion restriction). With respect to the first stage, the impact of the FHP funding on CCT coverage comes from two institutional features of the BF program: The form of dissemination of information about BF, and program permanence rules. Access to information about BF was key to the enrollment of potential beneficiaries in the early years (2004-06). Where local city administrations had few resources to promote enrollment, the FHP teams had sufficient capillarity and penetration among the poor to disseminate the information, therefore supporting higher enrollment rates where the FHP was better funded. Moreover, BF program rules allow households to continuously receive the benefit for a minimum of four years once they are enrolled (much more in many cases), which allowed the coverage differentials created between 2004 and 2006 to persist over the long term (2012+).

With respect to the exclusion restriction, there is strong evidence that the political effects of the FHP operate almost exclusively through the channel of CCT coverage. This evidence emerges from the use of a MRDD for the estimation. MRDDs differ from single-variable cases on the basis of their implementation methods and interpretive framework. For example, treatment effects are identified along a frontier of points on a two-variable space, rather than on a single point (for the variables for which the discontinuous assignment exist, in this case: population and HDI). This allows the researcher to observe the heterogeneity of these treatment effects for municipalities with different values of population and HDI along this frontier. In this context, the exclusion restriction is supported by the fact that, although the FHP funding differential is applied to the entire frontier, effects on political outcomes are only observed for the same group of municipalities for which differentials is CCT coverage are also observed.

I estimate the MRDD nonparametrically, and most results are robust to the choice of kernel, bandwidth, and the inclusion of control variables. This Chapter provides estimates based on a data-driven plug-in algorithm for bandwidth selection in the multivariate case, adapted to the specific demands of this study, and fully described in the appendix.

3.1 Institutional Background

In this Section I briefly describe the main institutional features of the Brazilian CCT program, the Family Health Program, and the local political environment in Brazil.

3.1.1 Brazilian CCT program

Bolsa Família (BF) is the largest CCT program in the world, covering 13 million households (nearly 20% of the population in Dec 2012). It was created in 2003 and unified other smaller CCT programs previously run by different government ministries.[29] In a nutshell, households with per capita income below a certain threshold are eligible to receive a monthly BF grant, which varies according to the number of children in the family under 18 years of age. For example, a family of two adults and two school-age children with per capita income at the lower threshold (R$70), would receive R$134, roughly a 50% income increase.

Eligibility is based on self-declared income, but households are subject to audits run by both the local and federal program offices. Program permanence is subject to compliance with conditionalities, particularly school attendance and, for targeted populations, health check-ups. The BF operations are run jointly by the central and local governments.[30] The former determines the major guidelines (the

[29]See Fried (2012) and Zucco (2013) for a historical account of the program
[30]For a more extensive analysis, see Lindert et al. (2007)

annual budget, the total cap on the benefits, the eligibility thresholds and the municipal coverage targets), controls the approval and cancellation of benefits, and pays the beneficiaries directly through a cash-card system. Local offices are responsible for the enrollment process, household data collection, requesting cancellations and additions, keeping the registry updated, and checking whether the conditionalities are being met by the families.

3.1.2 CCT and Local Politics

The Brazilian political system has a low level of ideological identification (Ames and Smith, 2010). The party system is highly fragmented and local political alliances tend to span the entire ideological spectrum. Mayors have a two-term limit and elections are held every four years in one round, under majority rule.[31] Between 2000 and 2012, 70% of eligible incumbents ran for reelection, and nearly 60% were reelected. Voting is mandatory and the average turnout in the 2008 and 2012 elections was 83%. Accordingly, this study will focus on the reelection of candidates and not parties.

The implementation of most government policies in Brazil is decentralized and programs are often financed jointly by federal, state, and municipal administrations. The majority of revenues for small municipalities come from transfers from higher levels of government; and clientelism and pork barrel spending are abundant at all levels (Alston and Mueller, 2006; Fried, 2012). Given that mayors have significant control over the budget allocation (Ferraz and Finan, 2011), vote buying offers that are financed through public funds and services are endemic to the political culture. Many of these exchanges also include bestowing administrative favors, such as access to health services or redirecting supplies from public construction projects.

The innovative BF program was specifically designed to reduce local political

[31] The runoff system exists for larger municipalities in Brazil, which are not included in this sample.

interference and promote the central government brand. Surveys indicate that its beneficiaries perceive BF as being more resistant to local political manipulation than other government programs (Rego and Pinzani, 2013; Sugiyama and Hunter, 2013). BF funds now represent the second most important source of federal government transfers to municipalities, comprising more than 12% of the total transfers, and these funds represent a disproportionately important source of revenues in less populated areas. Where BF total spending represents ~0.5% of Brazil's GDP, the BF transfers to small municipalities represent nearly 5% of the local budget.

3.1.3 The Family Health Program (FHP)

The FHP was created by the Ministry of Health (MH) in 1994. The program provides teams of health professionals that regularly visit households to provide basic health care; teams include a minimum of one family doctor, one nurse, one assistant nurse, and six health agents. Each team is responsible for a geographic area within the municipality, and serves a population of up to 4,000 by keeping a registry of clients, providing home visits, and functioning as the first point of access point to the broader health care system. Given that the majority of the middle and high income population use only the private health care system (Alves and Timmins, 2003), the FHP is, in practice, providing services to the poor population.

The identification strategy in this study uses a discontinuity in funding for the FHP program to instrument the cross-municipality variation in CCT coverage. Basic health care in municipalities is co-financed by federal, state, and local resources. The federal transfers used to finance the FHP teams are paid monthly as a fixed amount per team.[32] These payments were uniform across municipalities until August 2004, when municipalities with population below 30,000, and an HDI below 0.70,[33] started

[32]The federal FHP transfers represent roughly two-thirds of the basic attention funds, which in turn represent 6% of all the direct transfers to municipalities, including CCT.

[33]The population limit is 50,000 for the states that form the legal Amazon, including the entire North region (7 states), and the states of Maranhao and Mato Grosso. This region is excluded from

to receive an extra 50% funding per team (see the timeline of events in Figure 3.1). The HDI for eligibility was calculated based on the 2000 census, and the population was referenced using the 2003 estimates of the government's statistics department (IBGE). Both these values, and the list of eligible municipalities, have not been updated since.[34]

3.2 Using a Discontinuity in the FHP Funding as Research Design

The *Bolsa Família* program was first implemented in late 2003 and has had full, or nearly full global coverage since 2006. Nevertheless, there is significant variation in coverage between municipalities, already taking into account the number of eligible poor families. In 2012, the average local coverage exceeded the estimated number of eligible families by 10%, with a standard deviation of 16% (Figure 3.3).[35] This study uses the cross-municipality coverage variation to identify the effects of CCT on political variables. Given that the joint administration of the program between central and local governments allows for local political manipulation of eligibility, this paper will not treat this variation as exogenous, in contrast to previous attempts in the literature.[36]

I use a discontinuity in the funding for the FHP (described above) as an instrument for CCT coverage. I argue that the allocation of FHP funds prompted a

our sample.

[34]The list of locations eligible for the benefit that was released in 2004 has not changed even with the publication of new population estimates and a new census in 2010. This means that the eligibility for treatment could not have been manipulated by local political authorities. The original list is constant of the following decree: PORTARIA Nº 1.434/GM, July 14, 2004.

[35]The global coverage target for BF is the sum of local targets and is binding. The local coverage targets however are not binding. The sample has a large number of municipalities with coverage higher than 100%.

[36]Zucco (2013) argues that this differential can be treated as a quasi-random distribution within groups of similar cities, and Fried (2012) shows that this differential is not determined by national politics.

differential in program enrollment across municipalities in the early years of the BF program. While the smaller, older CCT programs had at most 6 million beneficiaries in 2002 (Zucco, 2013),[37] BF's coverage target was 11.1 million families (Figure 3.3). In order to achieve this, a massive migration from older programs had to be undertaken, along with a concerted effort to reach new entrants. Smaller municipalities lacked the resources necessary to reach all of the eligible households, given that they did not receive federal resources for program administration prior to 2006. At this time, FHP teams provided one of the most important sources of information about government benefits, given the extent of their capillarity and penetration within poor communities.

Evidence supporting this mechanism comes from a 2009 survey[38] including more than 10,000 CCT-eligible households. The survey shows that health care agents asked the household about their coverage status in 50% of the visits. In 12% of the households surveyed, the information about BF came first from a health professional.[39] Another 9% of the respondents stated that they would speak to a health professional when it came to questions about the program, in preference to other local officials.

There are two main reasons that account for early CCT coverage differentials persisting in the long run. First, the target of 11.1 million households was reached in 2006 and capped until 2009, which meant that new beneficiaries could only join when someone left BF.[40] Second, although beneficiaries are required to update their information every two years, under penalty of losing the benefit, this rule was not

[37]This number is an overestimation, given that hundreds of thousands of households which were benefiting from more than one program at the same time were double counted.

[38]AIBF - *Avaliação de Impacto do Bolsa Família*.

[39]This percentage is calculated on the basis of including only respondents that learned about the program from a first-hand source. This excludes information coming from family and friends, which in turn may also have been acquired from a first-hand source such as the media, schools, health professionals, etc.

[40]The global coverage target for BF based on the 2004 PNAD survey was 11.1 million families, effective between 2004 and 2009. The target was changed in 2009 and in 2011, in order to incorporate data from the 2006 PNAD survey and the 2010 census, respectively.

properly enforced prior to 2009. When the government started to increase the number of audits, it also created a permanence rule,[41] which allowed households that surpassed the income threshold to receive the benefit for two additional years. This long term persistence of coverage differentials allows me to examine the political impact of the program up to 9 years after its implementation.

This identification strategy also relies on the assumption that the FHP funding differentials do not have a significant direct effect on political outcomes (exclusion restriction). Supporting evidence emerges from the use of a MRDD for the estimation, which allows me to observe the heterogeneity of the FHP funding effects on both CCT coverage and the political outcomes, over different values of population and HDI (details in Section 3.4 and Figure 3.4). The idea is that, although the FHP funding differential is applied to the sample, effects on political outcomes are only observed for the same group of municipalities for which differentials is CCT coverage are also observed.

The intuition from the existence of heterogeneous effects on CCT coverage is the following: CCT coverage post-2004 is only likely to respond to FHP funding in municipalities that are neither very poor nor very wealthy. Very poor municipalities (low HDI) were the main targets of the old CCT programs, which means that the information about BF was widely available in these locations prior to 2004. On the other hand, wealthy municipalities (high HDI), especially the ones with low population, are likely to have a higher budget, more employees, and better coverage of health services. In these places, it was easier for city administrators to boost BF enrollment in the early years, severely limiting the impact of the additional FHP funding on information dissemination. If CCT coverage differentials are only observed in these municipalities, the same should happen to the political effects.

Before proceeding to the empirical strategy, there are two other issues to address.

[41] Legislation: Portaria MDS No 617 from August 11, 2010.

First, given that the FHP funding discontinuity was introduced after the CCT program (Figure 3.1), the exclusion restriction cannot be directly tested. Section 3.5.3 provides detailed evidence to support a valid exclusion restriction.

Second, the identification relies on the assumption that no other relevant variable follows the same pattern as the FHP funding at the discontinuity. The most important source of transfers from the Brazilian central government to municipalities is the *Fundo de Participacao dos Municipios* (FPM). The FPM is distributed in a discontinuous form across several population thresholds, where larger municipalities receive more funding. One of the population thresholds, at 30,564, is close to the population frontier of 30,000 in this study. Nevertheless, Section 3.5.4 shows strong evidence that the FPM does not drive the results. Two other alternative tests are performed to rule out the existence of potential omitted variable biases. First, the effects are estimated for all outcome variables in two periods, pre-treatment (see Table 3.4). Second, the paper shows the balance of fixed and pre-determined municipality characteristics, also before treatment (Table 3.1).

3.3 Data Sources and Description of Variables

Data on municipal CCT coverage comes on a monthly basis from the MDS, since January 2004. CCT coverage is measured as the percentage difference between the number of households receiving the benefit, and the number of eligible households.[42] In addition to *Bolsa Famiília*, I also include households receiving the older CCT programs from the Ministry of Education (*Bolsa Escola*), the Ministry

[42]The number of eligible families was estimated three times by the MDS, based on the PNAD surveys from 2004 and 2006, and the 2010 census. The 2006 and 2010 estimates are an improvement over the previous one on the basis of including a coefficient for income volatility. For the coverage in 2008 and 2012, I use the estimates of poor families from 2006 and 2010, respectively. For the pre-treatment CCT coverage calculate in June 2004, I use the 2004 estimate.

of Health (*Bolsa Alimentação*) and the MDS (*Cartão Alimentação*).[43] Nearly all the beneficiaries of these programs migrated to BF between 2003 and 2006, which means that their number is not meaningful after 2008.

The federal government provides monthly data with respect to the basic health transfers to municipalities, including the FHP funding. The MH makes data available with respect to coverage by health teams, as well as some health outcomes. These outcomes are measured within the scope of the public health system; while they might not be a good proxy for the overall quality of health services in a given municipality, they are a good proxy for the quality of the services provided to the poor share of the population.

Annual budget allocation data has been obtained from the National Treasury database (FINBRA), which breaks public expenses into two main categories: first, in terms of capital investment, personnel expenses, and other expenses; and second, in terms of function (e.g., education and health).[44] Not all municipalities release the data every year. I only use municipalities that released four years of data in at least one of the mayoral tenures of interest here (2005-08 and 2009-12), as well as for the base period of 1997-00, which is used as a control.[45] Thus, the sample used to estimate the effects on budget shares is a subset of the main sample.

Election data comes from the Federal Electoral Authority (TSE). For the four municipal elections held between 2000 and 2012, I extract the following variables: the

[43] Although households could only migrate to BF if they no longer received other benefits, there was a portion of recipients of the older programs that had more than one benefit. They would be double-counted in the sample. Since the number of beneficiaries in this older programs is not discontinuous at our thresholds of interest, and their number is 0.1% of the sample in 2008, and 0.0% in 2012, this should not be cause for bias.

[44] I exclude from the sample all the municipalities that report a zero share of budget in either personnel or capital expenses, and also in education or health expenses. This is most likely a reporting error.

[45] Health expenses were only reported as a separate category after 2000. Previously they were aggregated with spending in sanitation. Thus, all regressions including health expenses for the mayoral tenure of 1997-2000 also include sanitation expenses. They are not fully comparable to the ones using the later tenures (although sanitation was on average only 6% of the aggregated expenses in 2004-2012). For the year of 2001, I simply assume that health and sanitation expenses had the same ratio as in 2002-04 (same mayoral tenure), and I adjust the data accordingly.

incumbent's vote share, as a percentage of valid votes; the margin of victory, as the difference between the winner and the runner-up in percentage points; the number of candidates; the education level of the challengers, as the number of challengers with high school education among the ones that ranked first, or second in a competitive election;[46] campaign donations from the private sector;[47] and turnout. Data from the 2008 and 2012 mayoral elections is used. The 2004 election happened two months after the introduction of the discontinuity, so any effects are unlikely to be observed. For robustness, the paper shows the estimation results for the 2000-2004 period.

I classify the main Brazilian political parties according to their level of clientelism, on the basis of data from the Democratic Accountability and Linkages Project (DALP).[48] The variable labeled "challenger's party" is defined as the number of challengers in clientelistic parties, among the ones that ranked first, or second in a competitive election. In keeping with the proposed model, the sample includes only municipalities where the mayor has reelection incentives. The subset is determined using the results from the previous election.[49] Cases in which an eligible mayor did not run for reelection are not excluded from the main sample, as this decision is likely to be endogenous. However, for the estimation of the election outcomes, only

[46]Competitive election is one where the margin of victory was less than 10pp.

[47]I only use campaign donations from corporations and individuals (which excludes donations from parties and candidates). I will also aggregate all direct donations to the mayoral candidate, to the municipal election committee, and to the municipal party branch (noting that the last two can also be spent on campaigning for assembly members). Donations data are not available for the 2000 election nor for all incumbents.

[48]The DALP (Democratic Accountability and Linkages Project) is a survey from 2008 where political experts from several countries respond to questions about the political behavior of local parties. The project is supported and made available by the Political Science Department at Duke University. I use the scores from the four questions related to the intensity with which parties use clientelistic exchanges to gather votes. The parties with an average score above 3 (out of 4 in an increasing scale) are identified as clientelistic parties, with value = 1. All other parties are identified with value = 0. All small parties that were not evaluated by DALP are identified as non-clientelistic. Between 2008 and 12, 95% of municipalities were governed by a party that was represented in the DALP survey.

[49]In some cases the municipal election was ruled illegal by the electoral courts and a new, supplementary election was called. In this case, the results of the supplementary election were used to appoint the incumbent.

municipalities where the incumbent is actually running can be used.[50]

The following pre-determined variables come from the 2000 census: age profile, as the share of population aged 20-50; income inequality, as the population share of the top 10% in income divided by the share of the bottom 40%; share of urban population; share of males; and schooling, calculated as the share of household heads having completed high school. The GDP per capita is the average from 2000 to 2002. For the purposes of this study, municipalities with a small number of poor households are likely just generating noise. Thus, the main specification only uses municipalities with at least 25% of poor households, roughly 60% of the full sample.[51] The sample only includes municipalities created prior to 2000, and all data in R$ was converted to real values based on the CPI from Dec 2012 (IPCA).

3.4 Empirical Strategy

In this Section I briefly describe the use of a multivariate regression discontinuity design as an estimation strategy, and provide a link between the model of Chapter 2 and the statistical parameters estimated in this Chapter.

3.4.1 The Multivariate Regression Discontinuity Design (MRDD)

Single score RD designs have been widely explored in recent economic applications, and are generally seen as one of the most credible identification strategies (Lee and Lemieux, 2010; Keele and Titiunik, 2014). An extension of the RD approach is the case where the treatment eligibility is determined by two running

[50] I do not include mayors that did not achieve their post by way of an election (e.g. vice mayors who may have inherited the position following a resignation), given that they did not have a vote share in the previous election. Also, the timing of such event may occur be too close to the forthcoming election, which suggests that the reelection incentives for these mayors may be insignificant to budget allocation. Although most of these non-elected incumbents can be identified in the data, adding them to the sample does not alter the results.

[51] Results for the estimation using the full sample are provided in the appendix.

variables, e.g., latitude and longitude (Gerber, Kessler, and Meredith, 2011; Dell, 2010; Keele and Titiunik, 2014) or test scores (Jacob and Lefgren, 2004; Papay, Willett, and Murnane, 2011; Zajonc, 2012; Clark and Martorell, 2014). In the two-score case (MRDD), the average treatment effect (ATE) is identified for a frontier of points, in contrast to a single point in the one-score case. In this study, a municipality m with population p_m and HDI h_m, with respective treatment cut-offs at 30 (thousands) and 0.70, has the ATE defined over the following frontier: $F = (p_m, h_m) : (p_m <= 30, h_m = 0.7) \cup (p_m = 30, h_m <= 0.7)$ (Figure 3.4).

This changes the estimation and interpretation of the treatment effects within the RD framework, mainly due to potential heterogeneity of these effects along the frontier. The literature on identification and estimation of MRDDs is sparse and lacks consensus on a definitive strategy. Papay, Willett, and Murnane (2011) propose a framework to estimate the ATEs nonparametrically when there are multiple treatments, which is not the case for this project. Although Reardon and Robinson (2010) and Wong, Steiner, and Cook (2013) review potential estimation strategies, they focus on the average effects, without emphasis on the heterogeneity along the frontier.

In several applications, researchers approached the problem by reducing it to a single-score RDD. This can be accomplished by estimating two separate ATEs for the two running variables (Reardon and Robinson, 2010; Wong, Steiner, and Cook, 2013), or by collapsing the variables into one single score. This single score is usually defined as the minimum distance to the frontier, among the values of the multiple scores (Jacob and Lefgren, 2004; Dell, 2010; Clark and Martorell, 2014). This latter approach, however, is more compelling when the variables are on the same scale, as test scores. Another noteworthy alternative is to assume a parametric[52] function

[52] In this parametric approach, the ATE along the frontier can only be inferred if the polynomial on the scores does not include interactions with the treatment dummy (spline). Under the inclusion of such interactions, the coefficient measuring the treatment dummy will have a different interpretation. It will reflect the conditional ATE at the point where the running variables equal

over the two score variables, as suggested by Reardon and Robinson (2010). Dell (2010) provides an application of this strategy by using a cubic polynomial in a geographical MRDD.[53] In both these cases, if there are enough observations on both sides of the entire frontier, the heterogeneity of the ATE can be consistently estimated using fixed effects for frontier segments, interacted with the treatment dummy. This is not, however, the case of the sample here (Figure 3.4).

There are two good reasons to explore the heterogeneity of the effects here. First, the two scores have a distinct nature. The sub-populations being compared along the frontier, by way of either a minimum distance approach or a parametric approach, might differ considerably. This would defeat the spirit of the RDD.[54] Second, the estimation of the effect of CCT on political outcomes depends on having a strong instrument for CCT coverage. The former approaches would mask areas of the frontier where the instrument is weak or ineffective. This paper follows the general estimation approach in Zajonc (2012), in which the conditional ATE (CATE) is estimated for several points of the treatment frontier, and the average effect for any frontier segment is derived by averaging those CATEs. This approach makes the treatment heterogeneity along the frontier fully observable, and ensures that the observations being used are in the same neighborhood with respect to the scores.

For scores p_m and h_m, the CATE (τ_{Con}) at a point (p, h) in the frontier is given by equation 3.1 below. $N_\epsilon^+(p, h)$ and $N_\epsilon^-(p, h)$ are neighborhoods of radius ϵ around point (p, h), comprised of treated and non-treated observations, respectively. Zajonc (2012) shows that this effect can be identified by way of assumptions similar to the single-score problem. Namely, the orthogonality of treatment assignment to

the cutt-offs, in contrast to the average treatment effect for the entire frontier.

[53]More precisely, Dell (2010) adopts a semi-parametric approach as the local polynomial is estimated for different bandwidths in distance to the treatment border

[54]A municipality with a 30,000 population with HDI below 0.60 might be compared to one with a 3,000 population and a 0.7 HDI. The spirit of the RDD is to match observations that are in the same neighborhood in regards to the score variables, which is not necessarily the case under these approaches.

the outcome variable; the positivity of the frontier, to assure that points near the frontier do exist; and the continuity of both the conditional regression functions $\mathbb{E}[y_m(1) \mid p_m = p, h_m = h]$ and $\mathbb{E}[y_m(0) \mid P_m = p, H_m = h]$, and the marginal joint density of the scores along the frontier.

$$\tau_{Con} = \lim_{\epsilon \to 0} \mathbb{E}[y_m \mid (p_m, h_m) \in N_\epsilon^+(p,h)] \; - \; \lim_{\epsilon \to 0} \mathbb{E}[y_m \mid (p_m, h_m) \in N_\epsilon^-(p,h)] \quad (3.1)$$

Following the recommendation for the similar, single-score RDD (Imbens and Lemieux, 2008; Imbens and Kalyanaraman, 2012), the CATE for the point (p, h) can be estimated nonparametrically by local linear regression, using the two scores as dependent variables. In municipality m at period t,[55] CCT coverage is defined as cct_{mt}. The first stage of the instrumental variable estimation is shown in equation 3.2 below.

$$cct_{mt} = \alpha_0 + \alpha_1\delta_m + \alpha_2 p_m^c + \alpha_3 h_m^c + \alpha_4\delta_m p_m^c + \alpha_5\delta_m h_m^c + \eta_t + \gamma_s + \theta_m + \mu_{mt} \quad (3.2)$$

The treatment effect is denoted by α_1, where $\delta_m = 1[(p_m <= 30, h_m <= 0.7]$. The values of population and HDI centered around point (p, h) are denoted by (p_m^c, h_m^c). I will also include state effects (γ_s), a period dummy (η_t), and a vector of municipal controls (θ_m).[56] This is usual in RDDs to reduce the sample variability

[55] A period is defined as the 4-year electoral term, and the CCT coverage is measured at the end of the period. In this case, the estimation is done for the values of coverage in 2008 and 2012 (December).

[56] Unless otherwise noted, the following variables are included as controls: latitude, longitude, their interaction, area, pre-treatment CCT coverage, pre-treatment FHP coverage, share of Old CCT beneficiaries in the population, GDP per capita (log) and the share of males in the population. The regressions for the incumbent's vote share and campaign donations also include the share of votes in the last election and a dummy indicating if the candidate belongs to the federal party. For variables that measure the education and clientelistic party affiliation of the challengers, dummies are included for both the federal party, and the clientelistic party affiliation of the incumbent in that election. Finally, for the budget variables, the past share of the budget (1997-2000 tenure) is

(Lee and Lemieux, 2010). The local linear regression will be weighted by the edge kernel in the main specification, but the robustness to the choice of kernel will also be tested. All kernels are two-dimensional, defined as the following product kernel: $K(u_1, u_2) = [K(u_1) \cdot K(u_2)]$.

This equation can be estimated for any point of the frontier, by centering the scores at the desired point. The effects are estimated for 19 points, limiting the data to a bandwidth defined over the two score variables.[57] For a simpler interpretation of the bandwidths, I will normalize the score variables to the same scale, according to their standard deviation. The average effect for any frontier segment (τ_{Avg}) can be estimated as the average of CATEs for k points along the frontier, weighted by the joint density $\lambda(p_k, h_k)$ of each point (equation 3.3).

$$\hat{\tau}_{Avg} = \frac{\sum_{k=1}^{K} \tau_{Con}(p_k, h_k)\hat{\lambda}(p_k, h_k)}{\sum_{k=1}^{K} \hat{\lambda}(p_k, h_k)} \tag{3.3}$$

Given that the subsample used in the estimation of each CATE might overlap, the standard errors of the averaged coefficients are bootstrapped. Confidence intervals are calculated using the bias corrected and accelerated bootstrap method (Efron, 1979).

Another advantage of using a nonparametric strategy is the possibility of using a formal process for bandwidth selection. I present results estimated under optimal bandwidths, calculated with a data-driven plug-in algorithm for bandwidth selection in two dimensions. This algorithm is based on Zajonc (2012), in the spirit of Imbens and Kalyanaraman (2012). In contrast to the procedure proposed by Zajonc (2012), this algorithm allows the use of a different optimal bandwidth for each frontier point for which the CATE is estimated. It also allows the use of different

also included as a control.

[57] As an example, for the point $p_m = 25$ and $h_m = 0.7$, the centered values are $p_m^c = -5$ and $h_m^c = 0$. For bandwidths of 10,000 in population and 0.1 in HDI, the data used in the estimation is $D = (p_m, h_m) : (15 \leq p_m \leq 35, 0.6 \leq h_m \leq 0.8)$

bandwidths for the two score variables, reducing the estimated mean squared error of the coefficients. Excessively wide bandwidths are a common problem in outputs of plug-in algorithms. This procedure puts a cap on the maximum bandwidth, effectively limiting the amount of bias in the estimation. Finally, the algorithm is expanded to estimate the bandwidth for different kernels. I describe the construction of this algorithm in the appendix (Section B.2). As a robustness check, I also present results for constant bandwidths of 0.90 and 0.75 standard deviations.

For any political outcome out_{mt}, equation 3.4 below shows the second stage of the 2SLS estimation. Here, the effect of CCT on political outcomes is estimated using the predicted values of coverage from the first stage. The interaction between treatment δ_m and the score variables is also included in the second stage.[58] The coefficient β_1 represents the conditional ATE of CCT coverage on political outcomes, at the specific frontier point for which it was calculated. Average effects for any frontier segment, defined here as $\hat{\tau}_{IVAvg}$, can be calculated using equation 3.5 below.

$$out_{mt} = \beta_0 + \beta_1 c\hat{ct}_{mt} + \beta_2 p_m^c + \beta_3 h_m^c + \beta_4 \delta_m p_m^c + \beta_5 \delta_m h_m^c + \eta_t + \gamma_s + \theta_m + \epsilon_{mt} \quad (3.4)$$

$$\hat{\tau}_{IVAvg} = \frac{\sum_{k=1}^{K} \beta_{1k}(p_k, h_k)\hat{\lambda}(p_k, h_k)}{\sum_{k=1}^{K} \hat{\lambda}(p_k, h_k)} \quad (3.5)$$

Although I run the regression for a total of 19 points along the frontier, the main specification in this paper uses the frontier segment where the instrument is deemed strong.[59] Section 3.5.1, considers the reasons why the instrument is strong in only

[58] The single instrument used in the estimation is the ATE for CCT coverage at each discontinuity point.

[59] I use the segment that aggregates a continuum of 6 points where the average of the t-tests for the coefficient of the CCT coverage variable is at least 3.2, which corresponds to the rule of thumb statistic for the instrument to be considered strong in a just identified RDD. The segment selected using this strategy corresponds to the following: $Segment = (p_m, h_m) : (p_m >= 27.5, h_m = 0.7) \cup (p_m = 30, h_m >= 0.65)$

selected parts of the sample. The average optimal bandwidths for each outcome variable in this segment are shown in Table B.1.

3.4.2 Mapping the Theory to the Data

The model in the last Chapter presented six predictions for the effect of the CCT program on budget allocation, incumbency advantage, quality of politicians, and vote buying. These predictions are tested empirically using the cross-municipal variation in CCT coverage. This variation is instrumented by the discontinuity in the FHP funding, as described in Section 3.2, for municipalities at the treatment frontier. The effect of the CCT program on any political outcome is always calculated, for a given segment of the treatment frontier, by equation 3.5 defined in Section 3.4.

Proposition 2 predicts that transfers would increase pro-poor spending by incumbents, and it is tested using the local budget allocation. The data categorizes the spending on the basis of type and function. By type, the three observed categories are capital investment, personnel expenses, and other expenses. By function, the paper uses six categories, as follows: education, health, administration, urbanization and housing, social security, and transportation. They represent nearly 90% of the total spending. I define the pro-poor public good in terms of spending in education and health services. The effects of CCT coverage on the pro-poor and pro-wealthy spending are used test the signal of the terms $\frac{dt}{dy}$ and $\frac{d(b-t)}{dy}$, respectively.

The variable that maps the data to Proposition 3 (transfers reduce an incumbent's vote share) is the share of votes of the incumbent. The effect of CCT on this variable is a proxy for the value of $\frac{d\pi_I}{dy}$. In addition, I will examine secondary political outcomes that might lend support for this result (e.g. campaign donations, margin of victory in elections, number of candidates in elections).

Proposition 6 states that incumbents that are less efficient in public good distribution lose more votes and conduct a smaller shift towards pro-poor goods with the arrival of CCTs. This proposition is tested by mapping the candidate's type to his level of education, under the assumption that less educated politicians are likely to be relatively less efficient at public good distribution (and more efficient at vote buying, with a higher type θ).[60] The education levels of the competitive challengers are used as an outcome (competitive challengers are the ones that win, or become the runner-up in a close election). The effect of CCT on this variable corresponds to $\frac{d\pi_I}{dyd\theta_P}$. Additional tests for the signal of the incumbent's vote loss and budget shift to pro-poor goods by politician's quality (i.e.,$\frac{d\pi_I}{dyd\theta_P}$ and $\frac{dt}{dyd\theta_P}$) are performed by splitting the data according to the education level of the incumbent. The effect of the CCT program on the incumbent's vote share and budget allocation is then observed for the different subsamples.[61]

Although Propositions 4 and 5 do not provide new insights on political outcomes, the alignment of the empirical results with the predictions provides support for the mechanism under examination (they predict how the main results should vary with both the size of the poor population, and the political preferences of voters). Proposition 4 is tested, for both budget allocation and the incumbent's vote share, by splitting the data into municipalities according to the density of political preference around the neutral level,[62] and by examining the coefficients for each subsample. In

[60]This assumption is largely based on the following fact regarding candidates for mayor in 2000-12 in Brazil: Candidates from parties with high-clientelism score are on average less educated than candidates from parties with low-clientelism score (e.g. 12% less likely to have a college degree).

[61]Low education refers to politicians having less than a college degree, and high education refers to politicians having at least some post-secondary education (more than high school).

[62]Although local Brazilian politics are largely candidate-centered, as opposed to party-centered, federal elections are more ideological, and political platforms of congressional candidates are more aligned with the coalitions at the federal level. Based on this, I identify a group of "swing parties" in Brazil (PMDB, PP and PTB). These are the large parties that between1994 and 2012 always participated in the national government coalition, independent of the party holding power. Between 1994 and 2002, they composed the center-right coalition with PSDB and PFL. Between 2002 and 2012, they composed the center-left coalition led by PT. These parties had nearly 40% of the votes in the congressional elections of 1998 and 2002. Their share of votes in these congressional elections is used to proxy the density of the political preferences of municipalities, pre-treatment. Municipalities with an above-median share, are identified as "high density". Unfortunately, I do not observe the vote shares by level of income in these municipalities. The model prediction is based on the level of political preference of the poor group. Therefore, in the empirical estimation, I implicitly assume that this split, based on the political preferences of the entire municipality, is a

a similar fashion, Proposition 5 is tested, for both outcomes, by splitting the data into municipalities according to the share of poor families. In fact, this paper uses the subsample of municipalities with a high share (at least 25%) of poor families as the paper's main specification.

Proposition 1 states that the arrival of the CCT program reduces vote buying. Although the existence of vote buying cannot be directly tested within the scope of this study, its inclusion here is key to reconcile the mechanism that connects the other results. One alternative test can be performed using the DALP survey, which classifies Brazilian parties according to the level of clientelism. With $\frac{dv}{dy} < 0$, the model predicts that the loss in vote shares should be larger in municipalities governed by high-clientelism parties. The construction of all variables referenced above is fully described in Section 3.3.

3.5 Results and Interpretation

Before proceeding with the discussion of the results, note that Table 3.1 shows that the municipalities on the two sides of the treatment frontier are comparable. The coefficients are shown for the preferred frontier segment. The coefficients are neither statistically significant at the optimal bandwidth (column 3), nor at the alternative bandwidths (columns 1 and 2).

3.5.1 First Stage: Regression Discontinuity Results

Table 3.2 shows the first stage results for the health funding and CCT coverage, for the preferred frontier segment. The coefficients estimated along the entire frontier can be seen in Figure 3.6. Column (1) is the preferred specification. It shows that a municipality at the discontinuity would have annual basic health transfers of R$1.8bn (pre-treatment), with the treatment triggering an increase of ~25%

good proxy for the poor share of the population.

(R$0.44bn). CCT coverage is 7.9pp (percentage points) higher at that segment. The additional amount of resources received by voters in a treated municipality is then R$1.2bn, which is nearly 3x higher than the differential in health funding generated by the discontinuity.

The remainder of the table shows the robustness of these results to the choice of bandwidth, kernel, and the use of controls and state effects. There is no significant variation in the magnitude of the coefficients under most specifications. Finally, as the bandwidth increases, the magnitude of the CCT coverage coefficient become weaker. This means that the potential bias coming from widening the bandwidth, although seemingly small, would work in favor of the results.

Heterogeneity of the ATE

Figure 3.5 shows the heterogeneity of the coefficient estimated for CCT coverage along the treatment frontier. The strength of the effect varies significantly across municipal characteristics. The first plot shows the CCT coverage coefficients for each one of the 19 bins in the frontier. The instrument is only strong at the filled blue circles (2 points on the left and 5 points on the right). Wealthier and smaller municipalities (left side of the figure) have a coefficient that is barely statistically significant, and weak in magnitude. The instrument is also weak for that frontier segment. For very low-HDI municipalities (extreme right side), the coefficient is weak in magnitude and statistically insignificant, which renders the instrument ineffective.

The second and third plots shows that, in line with the expectation, CCT coverage only responds to the FHP discontinuity in locations with the following characteristics: a relatively low infrastructure of public services, and a relatively low coverage of old CCT programs (see the discussion in Section 3.2). Accordingly, the second plot shows the pre-treatment, average coverage of health programs for the same points as the first plot. The strength of the CCT coefficient (first plot) coincides with low coverage, either by family health teams or health agents.

The intuition is that, in area well-covered by health services before treatment, the FHP extra funding would not have a significant impact on how well information about CCT was disseminated in those municipalities (i.e. health teams had already enough funds to cover most poor population with frequent visits). This is supported by the coefficients estimated for FHP visits along the treatment frontier, shown in Figure B.3 and Figure B.4 in the appendix. They show that both the total number of visits, and the rate of visits per family, are only significantly higher for treated municipalities on the right side of the plot.

The third plot shows the quality of infrastructure before treatment (per capita budget and per capita number of public employees), and the coverage of old CCT programs. Again, the instrument is weak where the infrastructure for public services is better. This indicates that the funding discontinuity had a limited role on the information channel for CCT enrollment in those areas. The low-HDI areas (extreme right) show a much higher coverage from older CCT programs, indicating that information about BF would likely have been well disseminated in these locations prior to 2004 (beneficiaries of old CCT programs only had to migrate to BF).

3.5.2 Second Stage: Political Outcomes

Table 3.3 shows the main results for the political outcomes, for the paper's main specification. Column (1) shows the pre-treatment value; column (2) shows the reduced form coefficients from the regression discontinuity; column (3) shows the results of an OLS regression; and column (4) shows the results of the IV regression. Table 3.4 shows the robustness of the MRDD results to kernel and bandwidth choice, and the inclusion of municipal controls. Table 3.4 also contains a placebo test, with the coefficients estimated for the past values of the variables.[63] All results are estimated for the preferred frontier segment. The results for the entire treatment frontier are shown in Figure 3.6 and Figure 3.7. Table 3.6 shows the coefficients for subsamples of the data, according to the following categories: education of the incumbent, party of the incumbent, density of political preference, and the number of poor families in the municipality. All results are for a sample of municipalities with at least 25% of poor households. An exception is column (7) in Table 3.6, where the results are shown for a sample of municipalities with a small share of poor households (less than 40%). For reference, Table B.2 in the appendix presents the estimation results for electoral turnout, as well as all the remaining budget variables not included here.

Election Outcomes: Proposition 3 The main result of interest here is the loss of support by the incumbent, which is a direct model prediction. From Table 3.3, the pre-treatment average vote share for incumbents is 51% in the elections of 2008 and 2012. In line with the theory, columns (2) and (4) show that the overall effect of CCT on the incumbent's vote share is negative. From the IV regression, for a 10pp increase in CCT coverage, there is an 8.2pp vote loss for the incumbent. This result is robust to the choice of kernel, bandwidth, and the exclusion of municipal controls (Table 3.4). Column (8) in Table 3.4 shows that this effect was not present in the elections of 2000 and 2004, when in fact the coefficient was statistically insignificant and positive, at 1.1pp.

The loss in vote shares by the incumbent is supported by the effects found for other election outcomes that are not direct predictions of the model.[64] These

[63] Column (8) with the past coefficients includes pooled data from the elections in 2000 and 2004, or the 1997-2000 and 2001-2004 mayoral tenures when available, unless otherwise indicated.

[64] Although the results for the loss of vote shares by incumbents is significant, the results using the probability of victory for incumbents are not statistically significant. In other words, incumbents lose votes, but most of them are still reelected. This is mostly because elections were not very competitive before treatment (the average margin of victory was very high at ~17pp). If we look

variables measure different dimensions of electoral competition (margin of victory and number of candidates in elections), and support for the incumbent (campaign donations). From the IV results in column (4) of Table 3.3, for a 10pp increase in CCT coverage, there are 0.4 more candidates running for mayor,[65] a 6.3pp lower margin of victory, and 50% less private campaign donations. These results are also robust to the choice of kernel, bandwidth, and the exclusion of controls (Table 3.4). As before, no significant effect on those variables are observed in past elections. Finally, the voter turnout does not significantly change at the discontinuity (Table B.2).

Election Outcomes: Proposition 6 I also test the prediction that less educated candidates would lose more support after the CCT arrival. From the IV results in column (4) of Table 3.3, for a 10pp increase in municipal CCT coverage, there is a 24pp higher share of competitive challengers that have at least a high school education. This result is robust to the choice of kernel, bandwidth, the exclusion of municipal controls, and it is also absent from the pre-treatment period. This prediction can also be tested by splitting incumbents into more and less educated, as shown in Table 3.6, in columns (1) and (2).[66] As expected, the magnitude of the negative coefficient for the incumbent's vote share, for the group of less educated incumbents, is almost double that of the group of more educated incumbents.[67]

at the results by sub-sample, we see that in the case of less educated incumbents, there is also a significant reduction in the probability of victory (in addition to the loss in vote shares, which is also much higher for this sub-sample).

[65]The fall in vote shares of the incumbent is not a mechanic consequence of the reduction in the increase in the number of candidates. Even when using only 2-candidate races, the effect on vote shares are still statistically significant.

[66]High education is defined here as having more than 12 years of formal schooling, i.e., some post-secondary education; and low education is defined here as up to and including high school.

[67]Note that the confidence intervals for the two coefficients overlap. This is the case with most of the results in this Table, since the subsamples have a much lower number of observations the estimation has high variance. Nevertheless, it is important to notice that most deviations in magnitudes respect the theoretical propositions and the coefficients in the sub-groups where stronger results are expected remain significant is most cases.

Election Outcomes: Propositions 4 and 5 The test of predictions from Propositions 4 and 5, for the incumbent's vote share, are shown in Table 3.6. First, the sample is split into municipalities with low and high density of political preference (see Section 3.4.2). The results are shown in columns (3) and (4). In line with the theory, in areas of less dense political preferences, the coefficient for the incumbent's vote loss becomes insignificant. It has roughly half the magnitude of the coefficient estimated for the high-density group. The results for Proposition 5 are shown in columns (7) and (8), and are also in keeping with the theory. The vote loss for the incumbent is higher in municipalities with a larger poor population.

Election Outcomes: Proposition 1 Although the existence of clientelism cannot be directly tested in the scope of this study, one alternative test can be performed using the DALP survey. The sample is split into groups of parties that have a high or low clientelism score, as defined by the survey. Columns (5) and (6) show that the incumbent's vote loss is higher when they belong to a high-clientelism party, indicating that the mechanism currently in place is, in fact, affecting vote buying practices. In addition, Table 3.4 shows that, for a 10pp increase in municipal CCT coverage, there is a 20pp lower share of competitive challengers for high-clientelism parties. This result is robust to the exclusion of municipal controls, narrower bandwidths, it is not present in the elections pre-treatment, but loses significance under under-smoothing kernels. For all variables, under narrower bandwidths, the magnitude of the results increases. As in the case of CCT coverage, increases in bandwidth represent a conservative approach and offer no reason for concern with respect to bias.

Election Outcomes: Discussion The overall direction shown by the results lines up well with the theoretical predictions. As for the magnitude, the loss in the incumbent's vote share is more than one-to-one in relation to the additional number of households receiving the benefit. This suggests some form of propagation of the voting effects of the CCT. If there are positive economic spillovers from higher CCT coverage, the electoral effects are expected to be higher than the ones restricted to the households receiving the benefits. The same goes for the increase in the number of candidates, given that even wealthier households will face a larger number of choices. As far as campaign donations go, it is unlikely that the reduction comes from donations from poor voters. Donations probably follow the reduction in the allocation to the pro-wealthy public good (see below), given that this group provides the bulk of financial support to candidates. Accordingly, lower campaign spending could also feed into further vote loss. All in all, this study's identification strategy supports the hypothesis that the loss of votes by the incumbent goes beyond the affected poor families, but it cannot identify which of the propagating effects is more relevant.

Budget Outcomes: Proposition 2 The main result of interest here is the budget allocation in health care and education. From Table 3.3, the pre-treatment average budget share of these services together is 52%. In line with the theory, columns (2) and (4) show an increase in the budget allocation to these areas. From the IV regression, for a 10pp increase in CCT coverage, there are increases of 2.7pp and 2.3pp in the shares allocated to education and health care, respectively. The health results are robust to the choice of kernel, bandwidth, and the exclusion of municipal controls. For education, the coefficients are less robust, losing significance both in the IV regression and without the municipal controls (Table 3.4). Although the research design predicts a small automatic increase in health spending as a direct result of the FHP funding gap, this increase would be at most 0.2pp. This is 5 times smaller than the lower limit of the confidence interval (1.1pp), and 10 times smaller than the point estimate.

As a placebo test, the positive effect for both variables was not present in the electoral tenures prior to 2004, when the aggregate result was negative and insignificant (see column (8) of Table 3.4).[68] For all other categories of budget allocation by function, the results were insignificant (Table B.2). The corresponding reduction in budget shares was likely spread across several functions. Two categories shown in Table B.2 warrant a comment. First, the effect in urbanization spending is positive at 2.1pp in 1997-04, despite being low and insignificant in 2005-2012. This indicates that urbanization spending could have provided the main source for funds reallocated to health and education. Second, although transportation spending[69] (2.7% of the total spending) was negative and significant, this result has to be treated with skepticism. The effect is not robust to the IV estimation, and it also pre-dates the treatment, given that it was present in between 1997 and 2004.

The spending by type provides a better picture of the mechanism in place. From Table 3.3, columns (2) and (4) show a decrease in the budget allocation to capital investment. From the IV regression, for a 10pp increase in CCT coverage, there is a 1.5pp decrease in the share allocated to this category. Between 1997 and 2004, this effect was positive and insignificant. The result is robust to kernel choice, but it loses power with the exclusion of the controls, or at narrower bandwidths. This suggests that incumbents reallocate resources to redistributive spending from infra-structure investment. In line with this dynamic, the expenses with personnel were higher and robust at the discontinuity, most likely due to the shift from capital investment to labor-intensive spending in education and health.[70] While the potential implications of this budget shift in the profile of public employees is outside the focus of this study,

[68]These effects were also insignificant when estimated for these two variables using only data from 2004.

[69]Transportation spending in small municipalities is mostly infra-structure spending in road transport, i.e., construction of roads and bridges. Most small municipalities have little spending on a public system of transportation.

[70]Although treated locations have a slightly lower annual budget, and a slightly higher number of employees in 2008-12, these effects are not statistically significant (Tables B.2 and B.7). Nevertheless, their combination allows for treated locations having a higher share of budget spent with personnel, but no significant difference in average wages.

it is briefly discussed in the appendix (Section B.1).

Budget Outcomes: Propositions 4, 5 and 6 The tests of the remaining theoretical predictions for the budget allocation variables are shown in Table 3.6. Columns (1) and (2) show that a less educated incumbent will also be less effective in accomplishing the budget shift, in keeping with the theoretical prediction from Proposition 6. Columns (3) and (4) show that the budget shift is stronger in municipalities with high-density of political preference. Columns (7) and (8) show that the results are all weaker and statistically insignificant for a low-poverty sample.

Finally, columns (5) and (6) show the budget effects by clientelism scores. The expectation here is less straightforward than the negative effects predicted for the incumbent's support. Following the result found for less educated incumbents, the budget shift should be smaller in high-clientelism parties if the politicians in these parties possess a personal competitive advantage in vote buying. However, the budget shift should be larger in high-clientelism parties if the local strength of such a party is mainly supported by a specific municipal characteristic (e.g. a high number of voters with low political preference), following the results from Proposition 4. Accordingly, the only significant result here was a larger increase in health spending for the high-clientelism group.

3.5.3 Exclusion Restriction

The direct effects of the program's funding cannot be tested separately from the effects happening through CCT, given that the discontinuity was introduced at the early stages of the BF program (Figure 3.1). Instead, there are four pieces of evidence indicating that the political outcomes are predominantly generated by the CCT variation, and not by the FHP funding.

First, the funding discontinuity is small and does not significantly change total budget (Table B.2). Even if these funds were used for purposes other than health, it is unlikely that this funding was relevant. Second, a channel through which the FHP could affect politics is that of health outcomes. However, Table B.6 shows that the funding differential did not have a significant effect on the primary health outcomes measured by the basic health system (e.g., rate of vaccinated children, infant mortality, and rate of pre-natal exams).

Third, the heterogeneity of the average treatment effects of the FHP funding on CCT coverage along the treatment frontier provide strong evidence in favor of the exclusion restriction. The coefficients for political outcomes show a weak response in the same areas of the frontier where the instrument is weak (Figures 3.6 and 3.7). To further investigate this, I selected two equal size ranges of the frontier (6 bins): the first was the preferred segment for this study, where the instrument is statistically strong, and the second was the segment with population from 7,500 to 17,500, and HDI of 0.7 (weak instrument).[71] The coefficients for these segments are shown in Table B.5. In the weak-IV segment (Segment A), all variables that were significant in segment B become insignificant, with the exception of health spending. This variable was still significant but with a lower power and magnitude.

Fourth, I attempt to identify a change in the direct effect of the FHP in politics after the arrival of the BF program using a different strategy. I run a regression using the political outcomes (y_{mt}) as dependent variables, explained by a dummy that reflects the presence of the FHP in municipality m at time t (FH_{mt}). This regression suffers from omitted variable bias. Political outcomes could be affected by the same factors influencing the decision to implement the FHP. Nevertheless, if the omitted variable bias is constant over time, then the interaction between a

[71] I selected these bins (0.7 HDI and low population) over the 6 bins on the other extreme of the frontier (low HDI and 30,000 population) for two reasons. First, the low population side provides a much larger sample. Second, the low HDI side is not balanced with respect to the number of poor families (i.e. treated municipalities tend to have more poor families), whereas the other two selected areas are balanced in all variables from Table 1.

time dummy for post-2004 ($\eta_{2008-12}$), and the FHP presence (FH_{mt}), should not be significant unless there is another variable changing the program's impact on politics.[72] This effect is measured by α_1 in the equation below.

$$y_{mt} = \alpha_0 + \alpha_1 FH_{mt} * \eta_{2008-12} + \alpha_2 FH_{mt} + \eta_{2008-12} + \eta_{2004} + \gamma_{micro} + \theta_m + \mu_{mt} \quad (3.6)$$

I also include municipality-specific controls (see the Table B.4 for details), denoted by θ_m, and dummies for micro regions[73], denoted by γ_{micro}. Columns (1-2) in Table B.4 show the coefficient α_2, i.e. the direct effect of the FHP presence in the political outcomes. This coefficient cannot be trusted for inference, due to the omitted variable bias. Columns (3-4) show the value of α_1. After 2004, there is a statistically significant change in the program effect when it comes to the incumbent's vote share, the education of the challenger, health spending, capital investment, and personnel expenses. This change is always in the same direction (positive or negative) as the one pointed out by the main results of the MRDD in this study. I cannot affirm that the estimates in this paper are an upper or lower bound for the true effect of CCT on political outcomes without making assumptions about the signal of a potential direct impact of the FHP. This evidence, however, supports the hypothesis that another variable is driving these political outcomes, and confirms the findings with respect to the direction of these effects.

Finally, potential direct impacts are not necessarily intuitive. The FHP is the best rated health program in Brazil (IPEA, 2011). It is unlikely that it triggers a massive negative effect on the support for the mayors that implement it. As for

[72] I also include a dummy for 2004 (η_{2004}), so the effect in α_1 is compared to the effect in 2000. Although the 2004 electoral cycle started after the BF program, there insufficient time for the appearance significant changes in the political outcomes.

[73] Micro regions are not administrative regions, but they are used by IBGE for statistical purposes. They are comprised of contiguous municipalities with similar social and economic characteristics. There are 558 micro-regions in Brazil, each one with an average of 10 municipalities.

budget allocation, even if the entire FHP extra funding was actually spent in health services,[74] the predicted mechanical increase in the budget share of health services is only 0.2pp, nearly 10 times lower than the point estimate found in the paper.

3.5.4 Alternative Policy Discontinuities

The main potential source of omitted variables bias would be a policy with discontinuous implementation around the same thresholds of population and/or HDI used in this study. The *Fundo de Participacao dos Municipios* (FPM) is the most important source of federal transfers, and the main source of revenues for small municipalities. The FPM is distributed in a discontinuous form across several population thresholds, where larger municipalities receive a higher amount.[75] Although one of the FPM thresholds, at 30,564, is close to the 30,000 threshold used in this design, there is strong evidence that the FPM is not the cause of the political effects observed in this study.

First, the methodology of fund allocation in this design differs from the FPM methodology. The population threshold that determines the eligibility to a higher FHP funding was fixed in 2003, while the thresholds change every year for the FPM. This difference creates confounding effects in cases where municipalities crossed the FHP threshold at any time between 2003 and 2012, especially in the estimation under the edge kernel. Second, the absence of a significant effect on the total budget is evidence that the FPM is not generating a funding gap at this threshold (Table B.2). This is not surprising, given that the theoretical FPM differential is much higher at lower population thresholds.[76]

Third, the FPM dates back to the 1980s, and the population thresholds have

[74]The FHP is jointly financed by central, state and municipal governments. Thus, although the federal funds have to be spent in the program, municipalities would have been allowed to reduce their own contribution to the program and spend in other budget areas as they saw fit.

[75]This variation was recently explored in the political economy literature (Brollo et al., 2013).

[76]The difference in funding at the first 7 population thresholds for the FPM is: 33% for 10,188; 25% for 13,584; 20% for 16.980; 17% for 23,772; 14% for 30,564; 13% for 37,356; and 11% for 44,148.

remained the same since 2000. Thus, any direct effects of the FPM on political outcomes would have been observed in past elections (2000 and 2004), which is not the case (Table 3.4). Fourth, Table B.3 shows the reduced form coefficients for political outcomes in 2008 and 2012, setting the population cut-off for different FPM thresholds. I use populations of 23,773 and 37,356 (one threshold lower and one higher than 30,564). None of the variables had a significant result in the same direction as the results revealed by this study.

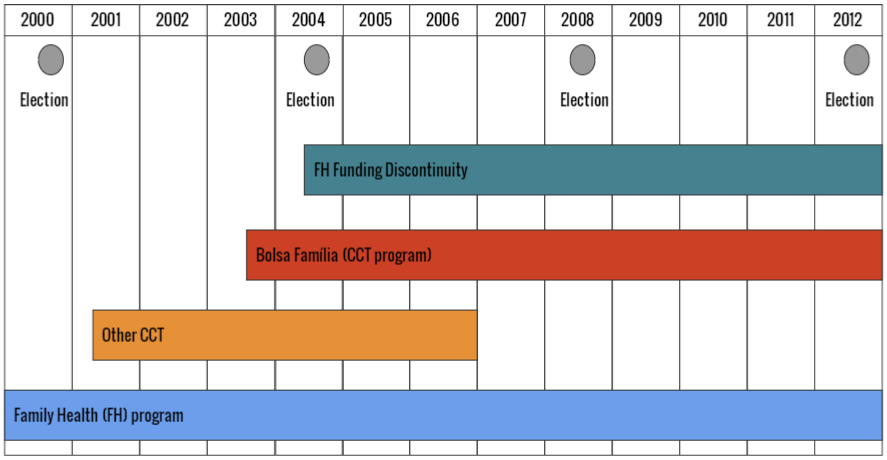
3.5.5 Alternative Explanations for the Mechanism

In this Section I briefly discuss two potential alternative explanations for the mechanism driving the empirical results. The main novelty in the framework presented in these Chapters is the inclusion of clientelistic exchanges in the canonical probabilistic voting model, as the cause of incumbency advantage. Although clientelism cannot be directly observed, there are two main reasons for this inclusion. First, the traditional model would not be able to explain why incumbents lose votes with changes in the utility of voters (this is shown in the Appendix). Second, clientelism as the channel for the political effects of CCT is supported both by the context, and by the results in this study (e.g. the fall in vote share of high-clientelism parties, or the results of the secondary propositions). It is possible that these (or other) alternative mechanisms co-exist with the one being tested in this thesis. These alternative stories, however, are neither tested empirically in this study, nor developed theoretically. Here I simply provide a few arguments on why I believe the proposed model is the best fit for the results in the data.

Change in Preferences. There is a class of models (e.g. median-voter theorem) that treats voters preferences as single-peaked over a certain policy dimension. In this context, the permanent income boost generated by the CCT program could shift their preferences, which would lead them to vote away from the politician that represented their old preference set (e.g. the incumbent, the less educated), and vote for the challenger. The main difficulty would be to conciliate the fact that voters are receiving more of the goods targeted to the poor, when they are becoming less poor, and their preferences should be moving move away from public services of education and health care. This model would also fail to explain why the only party effects observed are based on the clientelism score (there are no significant effects for party by coalition or ideology).

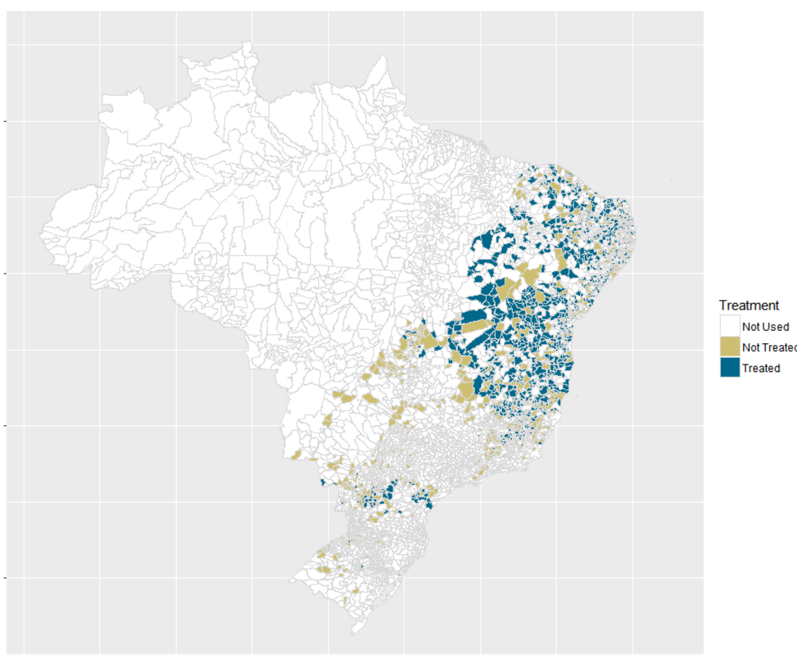
Increase in Demand for Public Services. The CCT benefit comes attached to the condition that beneficiaries attend school and have regular health check-ups. This could increase the demand for these services, which would lead politicians to provide more of them, and voters to replace politicians that provide little, with the ones that could be ”better” at delivering public goods (thereby voting the incumbents out). This story, however, fails to account for the fact that the poor population in Brazil already lacked proper access to schooling and health care before BF, and the demand for these services was likely already there before the program came. If providing these public goods was an effective electoral strategy, incumbents would have done it before the program arrived. Moreover, mayors had no obligation attached to the CCT program to increase spending in these areas.

Figure 3.1: Timeline of Events



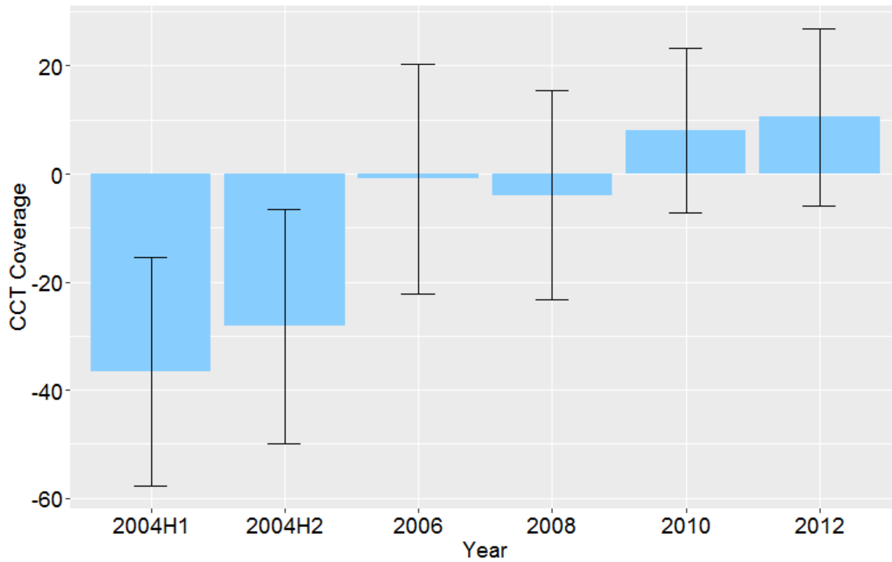
The FHP started in 1995. The Bolsa Família program started in Oct 2003 and the FHP funding discontinuity in Aug 2004. All elections happened in early October.

Figure 3.2: Map of Municipalities in the Sample



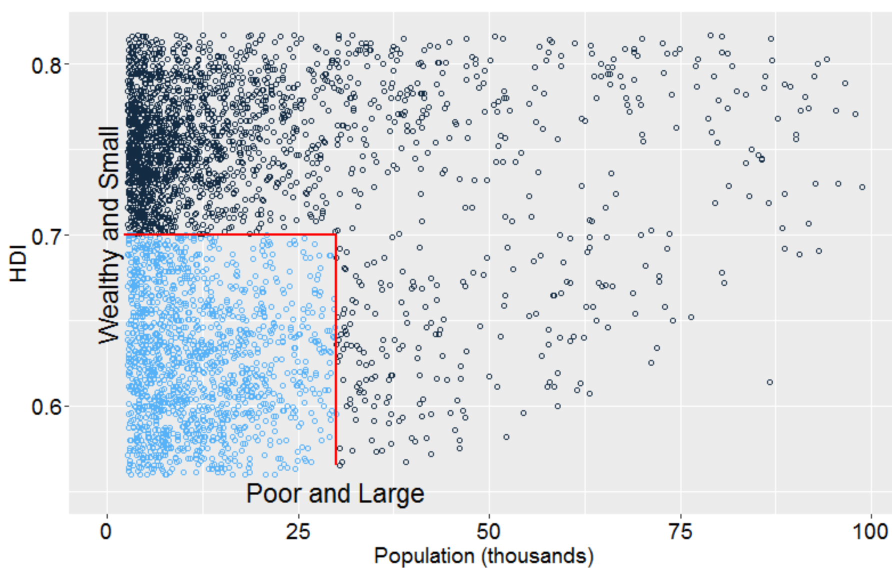
Municipalities in white are either outside of the bandwidth used for the RDD, located in the legal Amazon (high-left side) or have a share of poor population below 25% (most of the south). The map shows a total of 1,577 colored municipalities.

Figure 3.3: CCT Coverage vs. Number of Poor Families



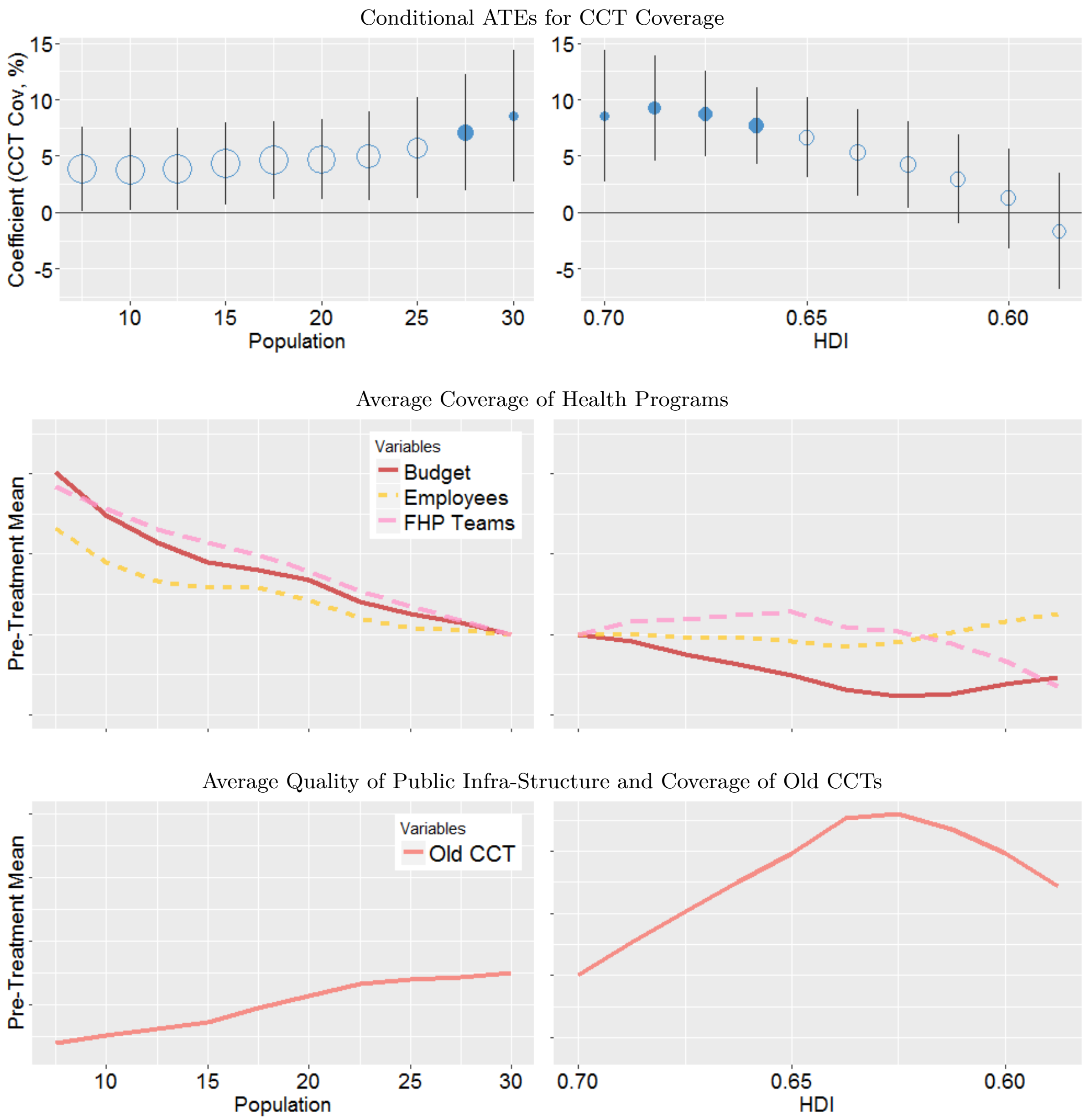
The y-axis shows in percentage points how much the CCT coverage was below the target. ** I adjust 2004 for 974,000 households receiving *Bolsa Família* and *Bolsa Escola*, as per the MDS. Even so, this coverage is still likely overstated (there is no adjustment for duplicity across old CCT benefits). CCT coverage includes the *Bolsa Família*, *Bolsa Escola*, *Bolsa Alimentação* and *Cartão Alimentação*.

Figure 3.4: Potential Sample and Treatment Frontier



The potential sample includes all municipalities within the central 95% percentile in population and HDI. The treatment frontier is the red line. Light blue dots represent municipalities eligible to treatment.

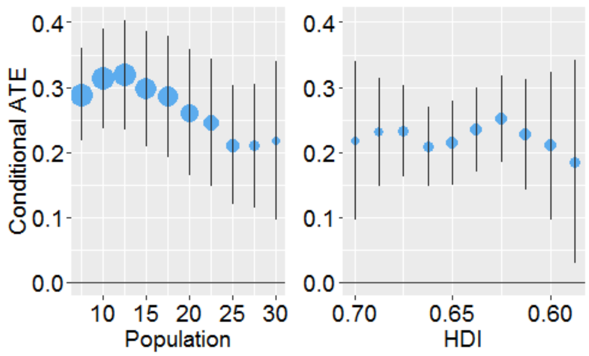
Figure 3.5: Heterogeneity of the CATE for CCT Coverage



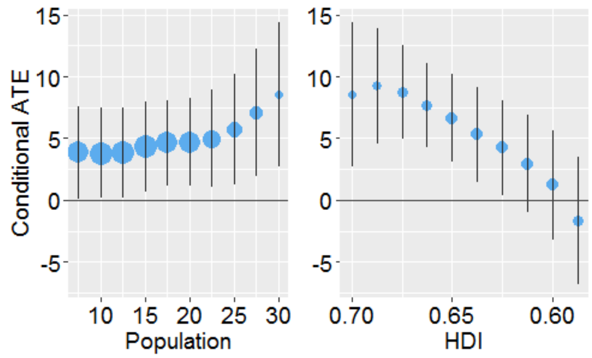
The y-axis shows in the first plot the conditional ATEs for CCT Coverage. The y-axis on the other plots shows the pre-treatment average of the variables (in common scale). For all charts, the left side has HDI fixed at 0.7 and population in 7,500-30,000. The right side has population fixed at 30,000 and HDI in 0.5875-0.7. The size of the dots represent the number of observations in each one of the 19 bins. I repeat the bin located at the origin in both sides.

Figure 3.6: Conditional ATE for Political Outcomes

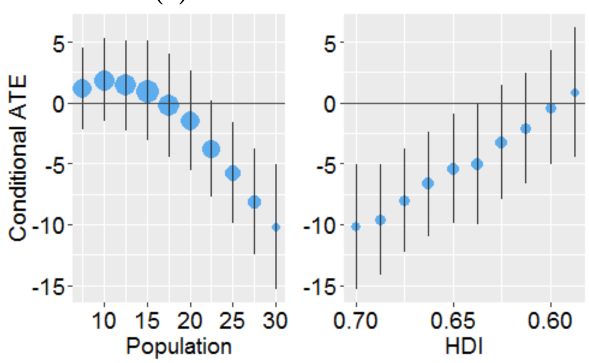
(a) Basic Health Funds



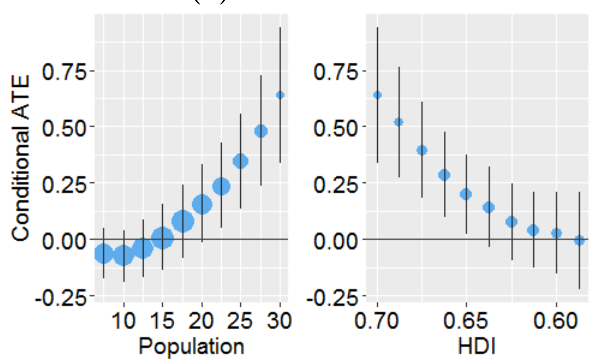
(b) CCT Coverage



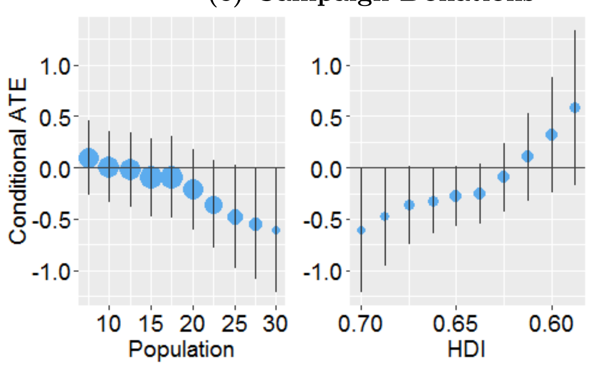
(c) Incumbent's Vote Share



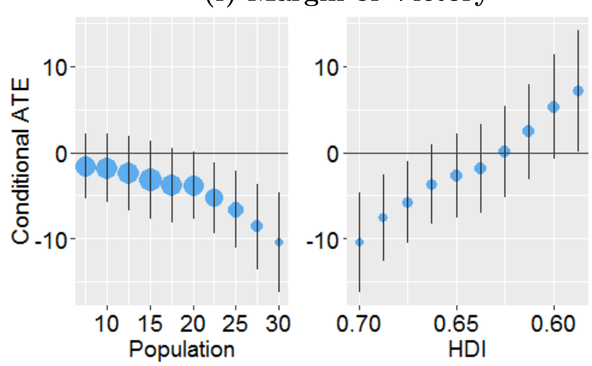
(d) Number of Candidates



(e) Campaign Donations

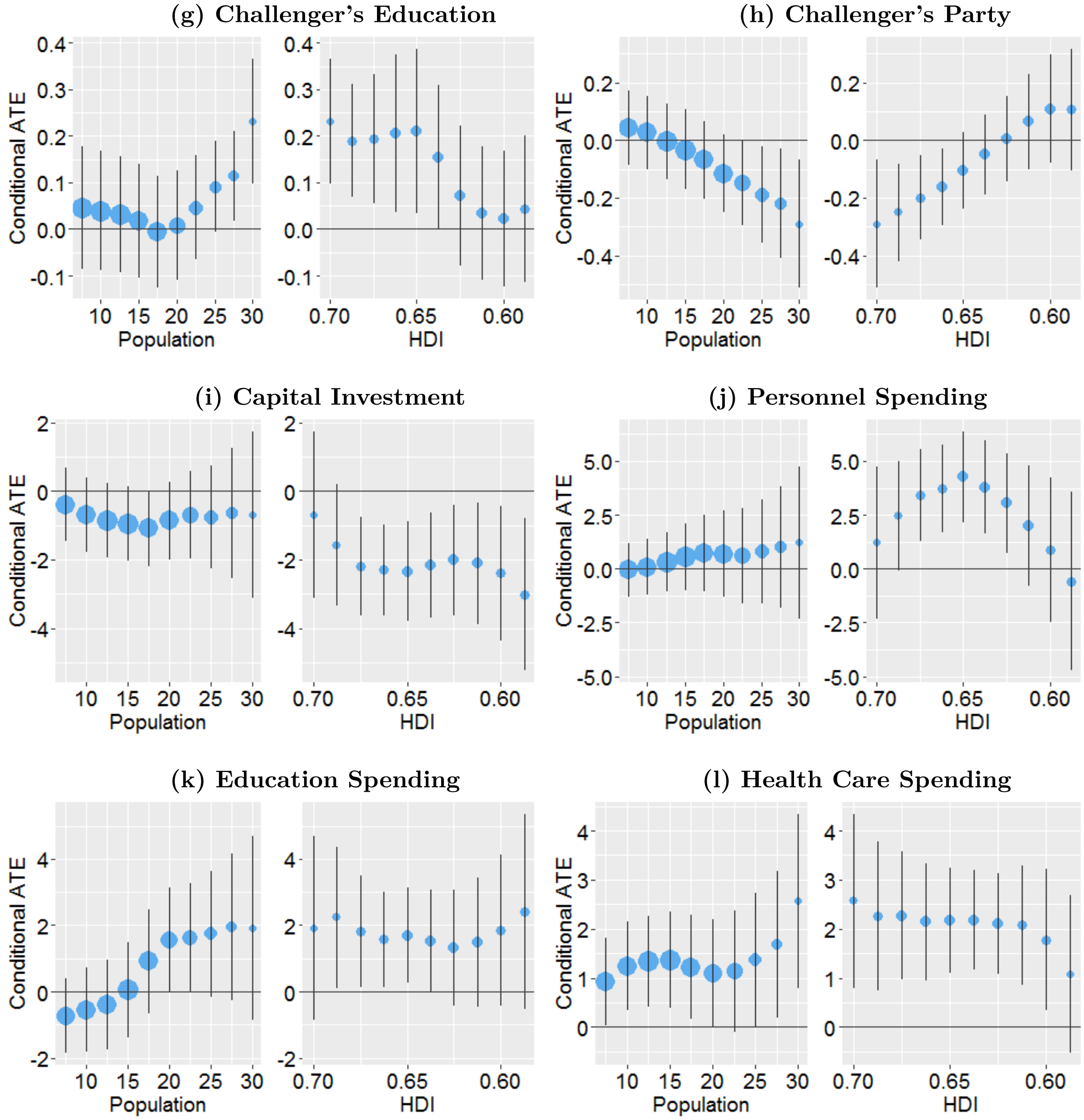


(f) Margin of Victory



The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI=0.7 and population in 7,500-30,000. The right side has population=30,000 and HDI in 0.7-0.5875. Regressions use the edge kernel, year and state effects, and controls. Bandwidth is set at 0.9 standard deviations for both scores. The size of the dots are the number of observations in each bin. I repeat the bin located at the origin in both sides. The bandwidth is the optimal for each variable.

Figure 3.7: Conditional ATE for Political Outcomes (Continued)



The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI=0.7 and population in 7,500-30,000. The right side has population=30,000 and HDI in 0.7-0.5875. Regressions use the edge kernel, year and state effects, and controls. Bandwidth is set at 0.9 standard deviations for both scores. The size of the dots are the number of observations in each bin. I repeat the bin located at the origin in both sides. The bandwidth is the optimal for each variable.

Table 3.1: Balance of Fixed and Pre-determined Variables

	PT Mean [90% CI]	(1)	Coefficient [90% CI] (2)	(3)	Opt. band (Pop,HDI) {Obs. / bin}
Bandwidth	*Optimal*	*Optimal*	*0.90*	*0.75*	
Latitude (degrees)	-40.59 [-41.18,-40.02]	-0.16 [-0.63,0.29]	-0.21 [-0.71,0.27]	-0.28 [-0.85,0.25]	(1.00,0.97) {584}
Longitude (degrees)	-13.16 [-13.96,-12.29]	-0.06 [-0.66,0.53]	-0.04 [-0.71,0.60]	-0.04 [-0.86,0.76]	(0.98,1.00) {596}
Schooling (% with high school)	9.43 [8.97,9.95]	0.65 [-0.09,1.46]	0.73 [-0.05,1.61]	0.66 [-0.24,1.64]	(1.00,0.97) {574}
Income Inequality (top 10% / bot. 40%)	21.42 [20.43,22.67]	2.31 [-0.06,4.81]	2.26 [-0.37,5.01]	2.50 [-0.61,5.81]	(1.00,1.00) {610}
Age Profile (share with 20-50)	39.60 [39.30,39.92]	-0.44 [-1.09,0.16]	-0.36 [-1.09,0.27]	-0.29 [-1.14,0.43]	(1.00,1.00) {607}
GDP per capita[a] (R$ '000)	2.83 [2.58,3.22]	0.07 [-0.15,0.30]	0.05 [-0.20,0.31]	0.02 [-0.27,0.35]	(0.99,0.99) {603}
Area[a] ('000 km2)	0.82 [0.67,0.99]	0.26 [-0.12,0.61]	0.28 [-0.14,0.66]	0.31 [-0.17,0.77]	(0.96,0.99) {574}
Urban pop. (% share)	63.01 [60.43,65.70]	1.13 [-3.56,5.80]	1.06 [-4.16,5.99]	-0.30 [-6.37,5.24]	(1.00,1.00) {609}
Gender (% share of male)	49.90 [49.71,50.09]	0.02 [-0.33,0.51]	0.03 [-0.37,0.61]	0.09 [-0.43,0.80]	(1.00,1.00) {607}
FHP teams (% coverage)	58.56 [52.75,64.02]	-0.19 [-11.65,11.01]	0.02 [-12.32,12.58]	-0.58 [-14.68,14.67]	(1.00,1.00) {611}
Poverty (% share)	39.87 [38.31,41.23]	0.75 [-2.06,2.99]	0.70 [-2.40,3.20]	0.78 [-2.86,3.87]	(1.00,1.00) {609}
CCT Coverage (% over target)	-14.34 [-19.11,-8.76]	1.20 [-6.97,9.75]	1.66 [-6.92,11.35]	3.16 [-6.55,14.65]	(0.99,1.00) {605}
Old CCT benefits (% of pop.)	18.57 [17.27,20.02]	-0.85 [-3.15,1.47]	-1.03 [-3.53,1.42]	-1.35 [-4.20,1.38]	(1.00,0.99) {606}
Obs. per bin			{476}	{314}	

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. The coefficients represent the average effect for the preferred frontier segment (6 bins). Regressions include year and state effects. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Table 3.2: Regression Discontinuity Results

	PT Mean [90% CI]	Coefficient [90% CI], avg. band. (Pop,HDI), {Obs. per bin} (1)	(2)	(3)	(4)
Kernel	*Edge*	*Edge*	*Epanec.*	*Gaussian*	*Uniform*
Main specification, optimal bandwidth					
Basic Health Funds[a]	1.76	0.22***	0.21***	0.20***	0.22***
(R$ million)	[1.64,1.86]	[0.14,0.29]	[0.13,0.29]	[0.10,0.30]	[0.11,0.34]
		(0.71,0.99)	(0.64,0.97)	(0.53,0.94)	(0.51,0.90)
		{329}	{269}	{180}	{166}
CCT Coverage	0.52	7.90***	7.65***	7.29***	6.92***
(% over target)	[-2.64,3.07]	[3.84,11.60]	[3.60,11.27]	[3.32,10.91]	[2.82,10.36]
		(1.00,1.00)	(1.00,0.98)	(1.00,0.94)	(1.00,0.92)
		{612}	{590}	{544}	{530}
No municipal controls or state effects, optimal bandwidth					
Basic Health Funds[a]		0.25***	0.24***	0.25***	0.27***
(R$ million)		[0.15,0.34]	[0.14,0.34]	[0.14,0.36]	[0.16,0.39]
CCT Coverage		5.92**	5.81**	5.36**	5.23**
(% over target)		[0.98,10.31]	[1.30,10.07]	[1.03,9.43]	[0.93,9.26]
Main specification, bandwidth = 0.90					
Basic Health Funds[a]		0.25***	0.27***	0.29***	0.30***
(R$ million)		[0.18,0.31]	[0.20,0.33]	[0.22,0.35]	[0.23,0.36]
CCT Coverage		8.44***	8.07***	7.53***	7.09***
(% over target)		[4.11,12.43]	[3.90,11.96]	[3.41,11.34]	[2.90,10.81]
Main specification, bandwidth = 0.75					
Basic Health Funds[a]		0.21***	0.23***	0.23***	0.24***
(R$ million)		[0.13,0.29]	[0.15,0.30]	[0.16,0.31]	[0.16,0.31]
CCT Coverage		8.93***	8.54***	8.25***	8.07***
(% over target)		[3.91,13.41]	[3.66,12.81]	[3.57,12.34]	[3.44,12.14]

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. The coefficients represent the ATE for the preferred frontier segment. The list of included municipal-level controls is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Table 3.3: Main Results: Political Outcomes

	Coefficient, [90% CI]				Avg Band.
	PT Mean	RDD	OLS	IV	(Pop. HDI)
	(1)	(2)	(3)	(4)	{Obs. / bin}
ELECTIONS					
Incumbent's Vote Share	50.90	-7.73***	0.03	-0.82**	(0.99,0.96)
(%)	[47.71,53.71]	[-12.42,-3.60]	[-0.08,0.12]	[-1.91,-0.28]	{431}
Number of candidates	2.31	0.39***	0.00	0.04***	(1.00,0.98)
(number)	[2.24,2.40]	[0.20,0.63]	[0.00,0.00]	[0.02,0.10]	{450}
Margin of Victory	16.53	-5.96**	0.07	-0.63**	(1.00,1.00)
(pct points)	[13.62,20.59]	[-11.69,-1.74]	[-0.04,0.17]	[-1.70,-0.09]	{467}
Campaign Donations[a]	11.29	-0.41*	0.00	-0.05*	(1.00,1.00)
(R$'000)	[11.03,11.46]	[-0.89,-0.06]	[-0.01,0.01]	[-0.14,0.00]	{448}
Challenger's Education	85.56	19.27**	-0.04	2.38*	(0.75,0.75)
(% with high school)	[74.87,92.72]	[4.56,33.57]	[-0.42,0.34]	[0.36,22.59]	{198}
Challenger's Party	51.58	-19.30*	-0.07	-1.99*	(1.00,1.00)
(1=clientelistic)	[43.63,60.24]	[-33.85,-3.42]	[-0.45,0.29]	[-5.26,-0.09]	{378}
BUDGET SHARES (by type)					
Capital Investment	10.56	-1.73*	-0.02	-0.15*	(1.00,1.00)
(% share)	[9.70,11.73]	[-3.34,-0.22]	[-0.05,0.00]	[-0.35,-0.02]	{352}
Personnel Spending	46.58	2.85**	0.01	0.25**	(1.00,1.00)
(% share)	[44.82,47.97]	[0.66,5.20]	[-0.03,0.05]	[0.07,0.60]	{352}
BUDGET SHARES (by function)					
Education	29.64	1.84**	0.01	0.23	(1.00,1.00)
(% share)	[28.33,30.90]	[0.34,3.91]	[-0.02,0.03]	[-0.02,1.04]	{350}
Health	22.22	2.17***	0.00	0.27**	(1.00,0.96)
(% share)	[21.18,23.38]	[1.05,3.73]	[-0.02,0.03]	[0.07,1.28]	{329}

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. The coefficients represent the ATE for the preferred frontier segment, under the edge kernel. All regressions include year and state effects. The list of included municipal-level controls is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Table 3.4: Robustness to Kernel Choice: Political Outcomes

	Coefficient, [90% CI]			
	(1)	(2)	(3)	(4)
Kernel	*Edge*	*Epanech.*	*Normal*	*Uniform*
ELECTIONS				
Incumbent's Vote Share	-7.73***	-7.39***	-7.08***	-6.68***
(%)	[-12.42,-3.60]	[-11.83,-3.50]	[-11.17,-3.19]	[-10.71,-2.82]
Number of candidates	0.39***	0.37***	0.34***	0.32***
(number)	[0.20,0.63]	[0.18,0.59]	[0.15,0.54]	[0.13,0.52]
Margin of Victory	-5.96**	-5.40**	-4.79**	-4.37**
(pct points)	[-11.69,-1.74]	[-10.41,-1.37]	[-9.34,-1.03]	[-8.64,-0.73]
Campaign Donations[a]	-0.41*	-0.38*	-0.46*	-0.46*
(R$'000)	[-0.89,-0.06]	[-0.85,-0.02]	[-0.93,-0.06]	[-0.94,-0.05]
Challenger's Education	19.27**	17.51**	13.96*	12.68*
(% with high school)	[4.56,33.57]	[3.68,31.20]	[1.11,25.75]	[0.14,23.88]
Challenger's Party	-19.30*	-16.61*	-13.80	-10.65
(1=clientelistic)	[-33.85,-3.42]	[-31.17,-0.57]	[-28.24,1.53]	[-25.70,4.26]
BUDGET SHARES (by type)				
Capital Investment	-1.73*	-1.65*	-1.68**	-1.62**
(% share)	[-3.34,-0.22]	[-3.15,-0.25]	[-3.10,-0.42]	[-2.98,-0.41]
Personnel Spending	2.85**	2.62**	2.56**	2.43**
(% share)	[0.66,5.20]	[0.53,4.85]	[0.57,4.74]	[0.50,4.64]
BUDGET SHARES (by function)				
Education	1.84**	1.64*	1.54*	1.40
(% share)	[0.34,3.91]	[0.19,3.59]	[0.12,3.37]	[-0.08,3.14]
Health	2.17***	2.09***	2.03***	2.00***
(% share)	[1.05,3.73]	[1.03,3.56]	[0.97,3.43]	[0.98,3.30]

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. The coefficients represent the ATE for the preferred frontier segment, under the optimal bandwidth. All regressions include year and state effects. The list of included municipal-level controls is described in the text.

Table 3.5: Other Robustness Tests: Political Outcomes

	Coefficient, [90% CI]			
	(1)	(2)	(3)	(4)
Bandwidth	*Optimal*	*0.90*	*0.75*	*Optimal*
Controls	*No*	*Yes*	*Yes*	*Yes*
Period	*2008-12*	*2008-12*	*2008-12*	*2000-04*
ELECTIONS				
Incumbent's Vote Share	-7.60***	-7.89***	-7.99**	1.08
(%)	[-11.95,-3.03]	[-12.95,-3.40]	[-14.03,-2.25]	[-2.96,5.71]
Number of candidates	0.35***	0.42***	0.46***	0.12
(number)	[0.14,0.59]	[0.22,0.68]	[0.23,0.77]	[-0.06,0.32]
Margin of Victory	-5.69*	-6.50**	-6.94**	-0.73
(pct points)	[-11.09,-0.91]	[-12.94,-1.77]	[-14.85,-1.28]	[-4.17,3.28]
Campaign Donations[a]	-0.40*	-0.47**	-0.55**	-0.58
(R$'000)	[-0.89,-0.04]	[-0.97,-0.08]	[-1.12,-0.13]	[-1.47,0.11]
Challenger's Education	17.65**	13.68*	19.31**	-0.65
(% with high school)	[4.33,30.17]	[1.79,25.01]	[4.60,33.60]	[-21.37,17.04]
Challenger's Party	-20.19*	-23.63**	-29.28**	-8.38
(1=clientelistic)	[-35.42,-3.49]	[-39.12,-6.49]	[-47.60,-10.10]	[-25.10,7.18]
BUDGET SHARES (by type)				
Capital Investment	-1.43	-1.62	-1.67	0.77
(% share)	[-3.01,0.02]	[-3.45,0.13]	[-4.05,0.70]	[-1.18,3.32]
Personnel Spending	2.73**	2.88*	3.08	0.20
(% share)	[0.60,5.19]	[0.34,5.54]	[-0.32,6.66]	[-1.65,2.21]
BUDGET SHARES (by function)				
Education	0.36	1.95*	2.25*	-1.20
(% share)	[-1.64,2.56]	[0.26,4.38]	[0.07,5.09]	[-2.56,0.09]
Health	2.61***	2.29***	2.44**	0.15
(% share)	[1.21,4.28]	[1.03,4.10]	[0.74,4.77]	[-1.76,1.81]

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. The coefficients represent the ATE for the preferred frontier segment, under the edge kernel. All regressions include year and state effects. The list of included municipal-level controls is described in the text.

Table 3.6: Main Results by Subsamples

Dependent Variable	Incumbent's Vote Share	Capital Investment	Education Spending	Health Care Spending
	Coefficient, [90% CI]			
by the Incumbent's Education				
High	-6.05**	-2.69**	3.47***	2.27**
	[-12.26,-1.39]	[-5.60,-0.60]	[1.72,6.53]	[0.68,4.90]
	{263}	{150}	{155}	{147}
Low	-11.13**	0.08	-1.08	1.49
	[-20.04,-3.21]	[-3.35,3.14]	[-3.83,1.63]	[-0.60,3.66]
	{205}	{129}	{122}	{114}
by the Density of Political Preferences Around the Neutral Level				
High	-10.79***	-3.63***	4.69***	2.52**
	[-21.70,-4.82]	[-6.11,-1.97]	[2.10,8.21]	[0.57,5.25]
	{215}	{176}	{174}	{164}
Low	-4.43	-0.89	0.07	1.25
	[-10.68,1.50]	[-3.09,1.54]	[-2.18,2.58]	[-0.65,3.21]
	{215}	{176}	{175}	{164}
by the Party's Clientelism Score				
High	-8.72***	-1.17	1.20	2.50**
	[-15.69,-3.17]	[-3.63,0.92]	[-0.69,4.14]	[0.82,5.40]
	{248}	{160}	{153}	{143}
Low	-2.51	-1.40	1.22	1.94
	[-11.07,4.86]	[-4.48,0.75]	[-0.72,4.74]	[-0.96,4.57]
	{183}	{100}	{104}	{099}
by Poverty in the Municipality				
High	-7.73***	-1.73*	1.84**	2.17***
	[-12.42,-3.60]	[-3.34,-0.22]	[0.34,3.91]	[1.05,3.73]
	{431}	{352}	{350}	{329}
Low	-4.43	-0.59	1.64	1.64
	[-11.44,4.58]	[-2.62,1.46]	[-1.07,4.85]	[-1.07,4.85]
	{355}	{353}	{345}	{345}

Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. Highly educated incumbents have some post-secondary education; less educated ones have less than than college. High density municipalities have above-median votes for "swing" parties in the 1998-2002 congress elections. High clientelism parties are defined by the DALP survey. High poverty sample has at least 25% of poor families (main specification). Low poverty sample has at most 45% of poor families. The coefficients represent the ATE for the preferred frontier segment (edge kernel). All regressions include year and state effects, and the municipal-level controls described in the text.

Chapter 4

Reelection Incentives and Fraud in Cash Transfers

The previous two Chapters show that, when cash transfers are effectively shielded from both local political manipulation and credit claiming, they reduce incumbency advantage in local elections. In this Chapter, I examine the scenario in which incumbent politicians are able to manipulate eligibility to CCT, and also benefit from it in the elections.

Corruption in the administration of public resources is a threat to the functioning of democracies. While corruption can take many forms, it is often driven by one of two motivations: politicians seeking personal rewards, or using the spoils of office to influence voters and gain reelection. The literature has well documented examples of rent extraction for personal benefit.[77] Against this rent-motivated corruption, the possibility of reelection can serve as a disciplining device that aligns the preferences of voters with the actions of politicians, therefore reducing corruption (Barro, 1973; Ferejohn, 1986; Duggan and Martinelli, 2016). On the other hand, when politicians illegally capture resources with the intent to direct them to voters, and influence

[77]See Ferraz and Finan (2011); Brollo and Troiano (2016) for their work in Brazil; Eggers and Hainmueller (2009) in Britain; and Querubin and Snyder (2013) in the USA.

elections,[78] the possibility of reelection creates incentives to fraud.[79]

This Chapter uses the administrative registry of the BF program (CadUnico) to measure income underreporting fraud. CadUnico contains the timing of enrollment, and the income reported by more than 24 million households in Brazil, for the purpose of program eligibility. I estimate that roughly 8 million households are currently receiving benefits by means of underreporting their income. Not all fraud, however, can be attributed to the households alone. The program enrollment is managed and overseen by municipal offices under the mayoral administrations, allowing for political manipulation of reported income.

I first demonstrate that the possibility of reelection for mayor creates incentives for fraud in the BF program. Using a regression discontinuity design (RDD), I identify the effects of reelection incentives on fraud by comparing municipalities were incumbents barely won the 2008 election (and were ineligible in 2012 due to a two-term limit), to ones were they barely lost (and new mayors were eligible in 2012). The evidence of fraud comes from the following: while the 2010 Census shows no significant differences in the income distribution and poverty levels of the two groups, a significantly higher share of the population reports very low income (in CadUnico) in the municipalities with reelection incentives.

This pattern is stronger near the election period, and it is uncorrelated to both contemporaneous measures of administrative and economic performance in the mu-

[78] See Camacho and Conover (2011) for an example in Colombia. The two motivations for rent extraction can co-exist, and often are mutually exclusive (the politician can chose to either use the rents for personal benefit or vote buying). This Chapter focuses on cases in which the institutional design leaves little room for extraction of personal benefits from the public sphere, but ample room for manipulation favoring voters. This is the case of the CCT program in Brazil.

[79] Politicians can also use public resources to influence voters without committing fraud (and without offering universal public goods). As an example, they can claim undeserved credit for the arrival of resources that they do not control. The case of foreign aid to the Philippines is well documented in Cruz and Schneider (2016). They can also use public resources to conduct tactical redistribution, by targeting selected groups of voters in order to maximize reelection probabilities (Dixit and Londregan, 1996). See Stokes et al. (2013) for a detailed account of clientelistic practices, which include the use of public resources for vote buying. This Chapter focuses on cases where the politician can only claim credit or manipulate a specific source of funds by illegal means (i.e. fraud or corruption).

nicipality, and other demographic information reported by poor households. It also lines up with the outcome of random government audits on reported income. This type of fraud generates expenses with BF that are nearly 7% higher in these municipalities, giving mayors ample ability to claim credit over the distribution of resources that are 5 times the average campaign spending in municipal elections.

Second, when the expectation of being audited increases in the municipality, mayors are less likely to promote fraud. Since 2004, the Brazilian government audits a random sample of municipalities every year, in order to examine the spending of federal resources, including the BF payments. Income underreporting fraud can trigger financial charges and legal action against both the household and members of the administration. The occurrence of audits in nearby municipalities should make the consequences of an audit more salient to mayors, and are used as a proxy for the mayor's audit risk perception. Given that audits are random, this Chapter can identify the change caused by audit risk in the effect of reelection incentives on the type of fraud mentioned above. Accordingly, the occurrence of at least one audit in a nearby municipality is enough to curtail all the politically-motivated fraud in election years.

Third, this Chapter shows that fraud helps incumbents to be reelected. Using matching over 40 observed characteristics of municipalities, parties and mayors, I compare the administrations that are top 25% in fraud to the remaining ones. For the sub-sample of all municipalities where the mayors actually ran for reelection in 2012, higher fraud increases the reelection probability by a significant 36%. This effect is higher where past elections were more competitive. Here again, there is strong evidence that fraud is uncorrelated with the administration performance in other areas, indicating that the increase in reelection probability is likely due to the mayor's ability to claim credit over illegally obtained BF transfers. All these results fit a probabilistic voting model tailored to include conditional cash transfers, and federal audits as an accountability device.

4.1 CCT Program Design, Credit Claiming and Manipulation of Eligibility

As described in Section 3.1 (Chapter 3), *Bolsa Família* (BF) is the largest CCT program in the world, and represents a significant share of all non-discretionary federal transfers to municipalities in Brazil. Given the country's history of abundant clientelism (Fried, 2012; Sugiyama and Hunter, 2013), and a decentralized structure of public spending (Ferraz and Finan, 2011; Brollo and Nannicini, 2012), it is not surprising that BF's political impact would raise the interest of both researchers[80] and the local press since its inception.[81]

CCT programs are known to increase the support for the federal government that implemented it. At the local level, although most CCT experiences severely limit the role of local politicians in the implementation, incumbents can still benefit from claiming credit for the program's arrival (Labonne, 2013; Rodriguez-Chamussy, 2015). In Brazil, however, undeserved credit claiming opportunities for local politicians are curtailed by a combination of program design and the political context.

First, different from most other CCTs around the world (e.g. Mexico, Philippines, Colombia), BF benefits are granted solely based on self-reported income (Handa and Davis, 2006). When designing the eligibility mechanism for CCT, there is a trade-off between inclusion and the quality of targeting. Simpler enrollment rules, such as self-reporting, are easy understood by beneficiaries and allow for more inclusion. They, however, come at a cost to targeting, given that ineligible house-

[80] See examples in De Janvry, Finan, and Sadoulet (2012); Sugiyama and Hunter (2013); Zucco (2013).

[81] The role of BF in shaping the Brazilian vote has been widely discussed by the press since the program's creation in 2003. These are examples of press articles (in Portuguese): http://goo.gl/6Ypa8u, http://goo.gl/e5KSSw, http://goo.gl/tr808O.

holds can easily access the benefit by underreporting income. While the Brazilian CCT was designed to maximize inclusion (Soares, Ribas, and Osório, 2010), the political implications of this enrollment technology are often overlooked. When individuals fully understand the eligibility rules and self-enroll, they are also less likely to attribute undeserved credit to a local politician.

Second, BF was designed to promote the federal government brand. The money is deposited directly into the beneficiary's account without the need for brokers or intermediaries, and a toll-free number gives them direct access to the federal management office if they need support. It is clear to beneficiaries who is eligible to the benefit, and who is paying for it.[82] Accordingly, surveys indicate that its beneficiaries perceive BF as being more immune to local political manipulation than other government programs (Rego and Pinzani, 2013; Sugiyama and Hunter, 2013).

Third, the Brazilian political environment has multiple parties (35), low ideological identification (Ames and Smith, 2010), and local elections are candidate-driven. Local coalitions do not respect party alliances at the federal level. In fact, it is common to observe coalitions including parties that are on opposing sides at the central government. In this context, local politicians have limited ability to claim credit for a successful policy implemented by their party at a higher level. Municipal elections in Brazil happen every four years, and there is a two-term limit for mayors.

While undermining credit claiming efforts, the BF management structure favors manipulation of program eligibility by local politicians. The program is managed jointly by the central and local administrations. Municipal governments are mainly responsible for household enrollment and data collection, which is done through the *Cadastro Unico* (CadUnico).[83] The federal government sets program targets,

[82] The BF card is carries the brand of the federal government on the front, and the toll-free number on the back.

[83] CadUnico is an integrated registry that contains a vast array of demographic information reported by households at the time of their enrollment/update. It can be accessed and updated by the local offices, and the data is used by the federal office to grant or deny BF benefits. CadUnico is also the platform for accessing other (smaller) poverty alleviation policies run by the federal

funding, eligibility criteria, and approves and denies benefits. The municipal offices do not control either the timing or the approval of benefits.[84] All households with monthly per capita (pc) income below half the minimum age (R$311 in 2012) are eligible to enroll in CadUnico. However, only families with pc income below R$140 are eligible to BF benefits, which are mainly of two types. The full benefit of R$70 is granted to all households with pc income below R$70, and it is not conditional on anything other than the self-reported income. The variable benefit (R$32-R$38) targets children (below 18) and pregnant women, and it is available to all households with pc income below R$140, conditional on school attendance and health check-ups.

Although local BF offices are responsible for the accuracy of the information entered into CadUnico, the only existing source of accountability for BF targeting are the audits run periodically by the federal government. The Office of the Comptroller General (CGU) audits the spending of earmarked federal resources on 120-180 random municipalities every year, which gives each municipality a probability of around 3% of being audited in any given year. These audits include the resources transferred to mayors to manage the CadUnico enrollment, and also interviews with a subset of the program beneficiaries to verify their reported income.[85] Income underreporting is considered a fraud, and households or public servants found responsible for this act can be prosecuted in a civil court.

To the extent that they can influence the enrollment/update process of CadUnico, incumbent mayors can use different strategies to benefit from the program.[86] This

government.

[84]For a more extensive analysis, see Lindert et al. (2007).

[85]In addition to audits, the federal government has recently started to cross the CadUnico data against data from other government databases such as RAIS (a database containing wage information for all employees in for formal market in Brazil), and CNIS (the database of retirement benefits). MDS also had a pilot program to use the remaining demographic information included (e.g. housing conditions, reported expenses) in CadUnico to detect inconsistencies in reported income, but this is still not the formal procedure for eligibility.

[86]There several reports in the local press (in Portuguese) describing attempts of local politicians to use the program in their own benefit. These include enrolling friends and family in the program, threatening households to cancel their benefits, or enrolling beneficiaries that are not eligible in exchange for votes. See some examples here: http://goo.gl/3RsfaW, http://goo.gl/QRTvLp,

paper focuses on the income underreporting fraud motivated by the reelection incentives of the mayor. Self-reported income as the only eligibility criteria generates a massive amount of income underreporting by households. Figure 4.1 shows the difference between the distributions of reported income in CadUnico and the 2010 Census.[87] The share of the population reporting income below R$70 (eligible to the full BF benefits) is much higher in CadUnico. While part of this underreporting can be attributed to the beneficiaries themselves,[88] this paper shows strong evidence that reelection incentives can lead mayors to promote income reporting fraud, and benefit electorally from it.

4.2 Theory: Reelection Incentives and Reverse Accountability

When politicians face the trade-off between providing public goods and extracting rent for personal benefit, accountability mechanisms such as reelection incentives had been shown to be effective in reducing corruption, as they realign the incentives of voters and politicians.[89] In the case of BF, the incentives of mayors and voters are aligned to increase program fraud. Income underreporting sanctioned by the local politician increases both the transfers to voters and the probability of reelection for incumbents, as it allows them to claim full credit for the extra benefit under a

http://goo.gl/D2LJTg, http://goo.gl/9GtLOy.

[87]The Census is conducted by a different government institute, and it is not used in any way for eligibility purposes in the BF program. The MDS, however, uses the Census data to update its estimates of eligible families on each municipality.

[88]Additionally, part of these gap can be attributed to the timing of both surveys, and to income volatility. In fact, the government official target of BF beneficiaries is not solely calculated based on reported income in the Census, but also includes an income volatility component. Nevertheless, there is clear evidence of income underreporting even after this adjustment. For example, for 2010, 21% of households reported income below R$140 in the Census survey, while the BF target is 29% of the population. However, the effective share of the households reporting income below R$140 in CadUnico is 41%.

[89]See Ferraz and Finan (2011) for an example in Brazil, and Duggan and Martinelli (2016) for a recent literature review on the subject.

very low cost (the BF resources do not come out of the local budget). I call this realignment reverse accountability.

This mechanism can be illustrated by this version of the probabilistic voting model adapted for CCT. This is a much simpler model than the one presented in the second Chapter. It follows the same general framework and, while it includes the possibility of claiming credit for CCTs, it does not include vote buying.[90] Let p be the share of poor households in the population n, one that is eligible to at least some benefits. Let ρ be the share of the poor (p) that are very poor, i.e. eligible to the full BF benefit (b). The very poor group cannot defraud the program, as they receive the full benefit without underreporting income. The share $(1-\rho)$ could only receive b if they underreport income. Define the share of households defrauding BF as f, i.e. the share of $np(1-\rho)$ that is actually underreporting income.[91]

In the spirit of the previous model, voters are prospective. They have utility coming from public goods received from the incumbent (g) and transfers, as long as their credit can be attributed to the local politician. I assume that g is exogenous, and constant over time and politician. Let h be the exogenous probability that the challenger will not allow and/or promote income underreporting if elected. Thus, when voters are allowed to underreport their income, they know that they continue to receive the extra benefit with probability $1-h$ if the challenger wins and probability 1 if they reelect the incumbent. Let α be the perceived probability that the municipality will be audited. The voter's utility coming from the incumbent is $g + fnp(1-\rho)(1-\alpha)b - \sigma$, where σ is a random variable that represents the overall popularity of the challenger, uniformly distributed in $[\frac{-1}{\psi}, \frac{1}{\psi}]$, so that ψ is the density of the distribution. The cost to the incumbent of promoting fraud is

[90] It is a model in the spirit of Persson and Tabellini (2000), and more recently Camacho and Conover (2011).

[91] For simplicity, I assume that the share $(1-p)$ of the population is not willing to commit fraud to enter the BF program in exchange of political support, so it will not be targeted by the municipality. This assumption has no relevant influence on the mechanism in place, and follows the literature on clientelism and vote buying to the extent that these practices thrive in the context of poverty.

increasing in f (assume cf^2), where c is exogenous.[92]

Following the Brazilian case, mayors can be reelected only once. In a two-period model, where winning incumbents have no reelection incentives in period 2, their only objective is the rents of office. Let r be the exogenous, maximum share of g that can be extracted as rent for personal gain. Because promoting BF fraud has a cost, but it only benefits the mayor through reelection probabilities, in period 2 the optimal fraud (f_2) for reelected mayors is zero, and rents are maximized at rg. Solving back to period 1, the incumbent's objective is to maximize his reelection probability π, which is shown below, subject to the cost of promoting fraud.

$$\pi = Prob[(rg + f_1 np(1-\rho)(1-\alpha)hb - rg) > \sigma] \tag{4.1}$$

Because the challenger cannot manipulate the eligibility, he only offers rg to voters.[93] Under the model assumptions, first-term incumbents maximize $\pi rg - \frac{c}{2} f_1^2$, solving for f_1 as shown below.

$$f_1 = \frac{\psi np(1-\rho)(1-\alpha)bhrg}{2c} \tag{4.2}$$

Several predictions of this model can be tested in the data. First, it is easy to see that, while second-term mayors have no incentives to promote income reporting fraud if $c > 0$, f_1 is positive for first-term mayors. Also, fraud should be higher in municipalities where the poor population is higher ($\frac{\partial f_1}{\partial p} > 0$), simply because fraud is more likely to have a meaningful impact on the election in these locations. Second, when the perceived probability of being audited is higher ($\frac{\partial f_1}{\partial \alpha} < 0$), mayors have less incentives to promote fraud, in keeping with the government's accountability mechanism. Third, higher fraud enhances the chances of reelection ($\frac{\partial \pi}{\partial f_1} > 0$) by

[92] In practice, c should be fairly low.

[93] In this model, because the rent extraction in period 1 does not serve as a signal of period 2 extraction (voters know that lame-duck mayors will always extract maximum rent), rent extraction is always maximized in period 1 as well.

aligning the preferences of poor voters and politicians.[94] Fourth, more competitive elections should increase both fraud ($\frac{\partial f_1}{\partial \psi} > 0$), and the effect of fraud on the re-election probability ($\frac{\partial^2 \pi}{\partial f_1 \partial \psi} > 0$). Although this paper's identification strategy does not allow for a formal test of these last two predictions, there is strong evidence indicating that their are in fact in line with the empirical results.

4.3 Data and Construction of Variables

The main data source in this paper is the administrative database with the full information of the CadUnico registry.[95] This database was extracted in Dec 2012, and it contains all the demographic information provided by the households during their most recent update process (households are required to update their information every two years, under the risk of losing their benefits). It also has data on which households are effectively receiving each type of BF benefit, and the date in which the information was last updated by the household, allowing me to observe how the income reporting changes as elections approach. Given that the identifying variation in this study is at the municipality level (i.e. the reelection incentives), all the household data is aggregated at the this level.

Within the group of households that are eligible to enroll in the CadUnico (pc income $\leq$ R\$311),[96] I define poor households as the ones with pc income below R\$140, and very poor households as the ones with per pc income below R\$70. The main variable used to measure fraud is the share of CadUnico-eligible households

[94]Given that the CCT transfers are funded by the federal government, and municipalities do not have binding coverage targets, the fact that some voters might be receiving the benefit does not affect the utility of other voters in the same municipality, i.e., they are not less or more likely to receive the benefit.

[95]This database is not available for public access, it was provided by the Ministry of Social Development (MDS) upon request.

[96]Approximately 8% of the entries in CadUnico had pc income above R\$311. These households are excluded from the sample. The higher income could be a consequence of a data entry error, or the fact that households that have pc income above the limit, but still have total monthly income of up to three minimum salaries, can also enroll (although they could never access BF benefits).

that are very poor, which are the ones eligible to the full benefit. I will also report supporting results for the following two variables: (1) the share of CadUnico-eligible households that are poor ; and (2) the share of BF-eligible households that are very poor. I built the same variables using data from the sample of the Census 2010, which was taken only two years before the 2012 election.

The Ministry of Social Development (MDS) provides monthly data on municipal CCT coverage, and the total value of transfers related to BF, from 2004 to 2015. This data, although only reported at the aggregated level, complements the CadUnico information for the period after the 2012 election (CadUnico data is not available after 2012). It also provides the IGD index, which measures the performance of the local administration in managing the BF program. The index is composed by four different scores, measuring the performance in checking health and school conditionalities, registry update, and coverage of the CadUnico.

Data on historical precipitation comes from the Climate Research Unit (CRU) of the University of East Anglia, and it is available for a 0.5 degrees grid in latitude and longitude, since 1970.[97] The remaining geographical (e.g. area, latitude, longitude) and demographic (e.g. population) data comes from the Brazilian Institute of Geography and Statistics (IBGE). Data on gender, income distribution, literacy and urban population comes from the 2000 IBGE Census.

As in Chapter 3, election data comes from the Superior Electoral Authority (TSE) and includes, in addition to election results, personal information of candidates such as gender, education, occupation, age, assets and campaign donations. Mayors with reelection incentives are the ones elected in 2008 for their first-term (first consecutive term). Mayors reelected in 2008 for their second consecutive term are the lame-duck mayors. The sample excludes municipalities were the lame-duck mayor stepped down before the end of the term. The reason is that the replacement

[97] I use the average monthly rainfall in 1970-2008, and the coefficient of variation of the rainfall, calculated as the standard deviation divided by the mean in this period.

mayor would actually have reelection incentives, without being elected in 2008.[98] I classify the main Brazilian political parties according to their level of clientelism (high or low), and ideology (left or right), based on the survey from the Democratic Accountability and Linkages Project (DALP).[99]

I use the number of recent CGU audits in neighbor municipalities[100] to proxy the expected probability that a certain municipality will be audited. The idea that audits in nearby locations make the audit risk more salient to politicians has recently been used in Lichand, Lopes, and Medeiros (2016). The main sample in this study includes only municipalities in the upper tercile in poverty, which are the ones with at least 67% of the households eligible to enroll in CadUnico (as of 2007).[101] This is in line with the model prediction that fraud should be lower in municipalities with low poverty rates, and also with the estimation strategy in Chapter 3. Table C.5 in the appendix shows the results for the low poverty sample of roughly the same size (bottom tercile in poverty).

[98]These municipalities are identified using the candidates in the 2012 election that reported their occupation as mayors, in municipalities were reelection incentives were absent. The results are robust to this exclusion, and the excluded observations represent only 1.6% of the sample

[99]The DALP (Democratic Accountability and Linkages Project) is a survey from 2008 where political experts from several countries respond to questions about the political behavior of local parties. The project is supported and made available by the Political Science Department at Duke University. I use the scores from the four questions related to the intensity with which parties use clientelistic exchanges to gather votes. The parties with an average score above 3 (out of 4 in an increasing scale) are identified as clientelistic parties, with value = 1. All other parties are identified with value = 0. All small parties that were not evaluated by DALP are identified as non-clientelistic. For ideology, I all major parties with a score lower than 5 in 0-10 scale were branded as left. I also included in the left group the small parties with a open socialist ideology.

[100]I do not include past audits in the municipality of interest, as those might have a confounding effect. There might be direct effects of recent audits on the amount of BF fraud in a given municipality.

[101]I calculate the poverty rate using the estimate of CadUnico eligible households made by the MDS, based on the 2006 PNAD survey, and the population count of 2007. This was the official target of BF beneficiaries until 2011, when it was re-calculated based on data from the 2010 Census. I still use the pre-2010 variable, as it pre-dates the 2008 election.

4.4 Empirical Strategy

In this Section I briefly describe the three different empirical strategies employed in this study.

4.4.1 Reelection Incentives and Fraud

The first part of this Chapter estimates the effect of reelection incentives on income underreporting fraud by comparing mayors elected for a first term in office, to mayors reelected for a second and last term. This strategy carries two main potential sources of bias (Ferraz and Finan, 2011; De Janvry, Finan, and Sadoulet, 2012), as illustrated in equation 4.3 below:

$$F_m = \beta_0 + \beta_1 \delta_m + \beta_2 experience_m + \beta_3 ability_m + \mu_m \tag{4.3}$$

F_m is the level of fraud in municipality m, and δ_m indicates a first-term mayor with reelection incentives ($\delta_m = 1$). The coefficient of interest is β_1. First, the experience bias reflects the fact that second-term mayors have, by design, more consecutive years of experience on the job than first-term mayors. For example, if political experience increases the willingness or ability of a mayor to promote fraud, then $\beta_1$will be biased upwards. The ability bias comes from a selection problem. Second-term mayors are the only group that was reelected for office after revealing their political ability. The estimate of β_1 will again be biased if both the politician's reelection chances, and the level of fraud, are correlated with either the politician ability or any characteristic of the municipality.

I employ a regression discontinuity design (RDD) to eliminate the potential ability bias coming from unobserved municipal characteristics that might determine reelection probabilities. The spirit of the RDD is to compare municipalities where incumbent mayors barely won the 2008 election, and therefore had no reelection incentives in 2012 (control group), to municipalities where they barely lost, and were replaced by first-term mayors that could be reelected in 2012 (treated group). This strategy provides a quasi-random assignment of first- and second-term mayors where elections were very close (Lee, 2008; Eggers et al., 2015), i.e. the margin of victory (MV_m) was nearly zero. I define the margin of victory as the difference in percentage points between the winner and the runner-up. The application of the RDD to close elections has been widely used in Brazil.[102] Under the assumptions of the RDD (Lee and Lemieux, 2010), the local average treatment effect (LATE) is identified for $MV_m = 0$. I estimate this effect non-parametrically using a local linear regression, as shown in the equation 4.4 below.

$$F_m = \beta_0 + \beta_1 \delta_m + \beta_2 MV_m + \beta_3 \delta_m MV_m + \gamma_s + \eta_p + \theta_m + \mu_m \tag{4.4}$$

The treatment effect is again denoted by β_1. I also include state effects (γ_s), party effects (η_p),[103] and a vector of municipal controls (θ_m).[104] This is usual in RDDs to reduce the sample variability (Lee and Lemieux, 2010). The local linear regression is weighted by the edge kernel, and estimated for a sample limited by a bandwidth around $MV_m = 0$. The main specification uses a bandwidth of 16pp (percentage points), which is roughly the average optimal bandwidth calculated for the three variables that measure fraud. For robustness, results will be shown for lower and higher bandwidths (13pp and 19pp), and also for an optimal bandwidth calculated by the widely used method proposed by Calonico, Cattaneo, and Titiunik (2014).

The RDD is only valid if covariates that are either fixed or pre-determined (de-

[102] For examples, see Boas and Hidalgo (2011); Ferraz and Finan (2011); Brollo and Nannicini (2012); Brollo and Troiano (2016).

[103] Brazil has a total of 26 states. There are 23 political parties represented in the sample.

[104] Unless otherwise noted, the following variables are included as controls: latitude; longitude; their interaction; the shares of male population and urban households (2000); the mayor's gender and education level (=1 if attended College); log of population; the share of votes PT had in the municipality in the 2006 presidential election; and the IGD index for the municipality for 2006-08.

termined before treatment) are balanced around at the discontinuity, so the control and treatment groups are not significantly different. Figure 4.3 (and C.2 in the appendix) show that the sample is balanced, under all bandwidths, for 26 out of 27 variables that include characteristics of the municipality, elected party, and mayor. The exception is the age of the elected mayor, which is significantly lower in the group of first-term mayors, most likely following the empirical design. In the appendix to this Chapter (Table C.5), I also show that the results estimated for a sub-sample of mayors between 40 and 55 years, where the variable age is balanced, are still significant and very close to the ones reported for the main sample.

In addition to the 27 pre-determined variables above, Figure 4.3 also shows that the variables measuring reported income levels in the 2010 census are also balanced. It is key for the results in this Chapter that municipalities with and without reelection incentives have the same share of poor households when measured by the income reported in the census survey. Evidence of income underreporting fraud comes from these municipalities having different shares of poor households when measured by income reported in the CadUnico survey (the one used for BF eligibility purposes).

To address the issues of both experience and ability biases, I follow Ferraz and Finan (2011). First, I show in Table 4.5 that the results remain robust if estimated only for a sample of first-term mayors that have previous political experience in the same municipality, i.e. they were mayors in 1996-2000 or 2000-04, or members of the local council in 2000-2004 or 2004-2008. The same Table also shows that the results remain robust when estimated only for a sample of first-term mayors with comparable political ability to second-term mayors, i.e., first-term mayors that actually ran for reelection in 2012, and won.

If the income underreporting is actually caused by an electoral strategy, as opposed to being a spillover of the mayor's administrative practices and performance, one should expect two things. First, the effects of reelection incentives on fraud should be uncorrelated with other performance measures in the mayoral administration, especially the ones related to the management of the BF program. I test this hypothesis on Table 4.3 in the appendix.

Second, the effects should be more significant near the election. Accordingly, the results estimated in this paper correspond to the fraud measured only for households that updated or enrolled in CadUnico in 2012, the year of the municipal election. These correspond to 45% of the households in the registry. Because households are only required to update their information every two years, and many do not comply, a still relevant part of the sample had their last income update on, or before 2011. This allows me to observe if these effects also occur in updates pre-election (Figure 4.5).

4.4.2 Expected Probability of Audits and Fraud

I use the number of CGU audits in the municipality's neighborhood in the first two years of the mayoral tenure (2009-10), which precede the pre-election period (2011-12), to test the model prediction that a higher expected probability of being audited reduces the mayor's incentives to promote fraud. I define a neighbor municipality as the one in the same micro-region (as defined by IBGE). The effect of audit risk on fraud is estimated by interacting the variable that measures reelection incentives (δ_m) with τ_m ($\tau_m = 1$ if there was any fraud in a neighbor municipality), as shown below:

$$F_m = \beta_0 + \beta_1\delta_m + \beta_2 MV_m + \beta_3\delta_m MV_m + \beta_4\tau_m + \beta_5\delta_m\tau_m + \beta_6\tau_m MV_m + \beta_7\delta_m\tau_m MV_m + \gamma_s + \eta_p + \theta_m + \mu_m \quad (4.5)$$

The original treatment effect (reelection incentives on fraud) is again denoted by β_1, but the impact of audit risk on this effect is given by β_4. In addition to the main results, I also show results including past audits in the same municipality, and restrict the neighborhood to radiuses of 110km and 55km, within the micro-region. As a placebo test, I estimate τ_m using audits in 2012-13 (the election was in 2012), when no significant effects should be observed.

4.4.3 Program Fraud and Reelection Probability

The identification of the causal effect of fraud on elections is not as straightforward. There are many variables that could potentially influence both the election results and the level of income underreporting in a given municipality. Here, I use a non-parametric matching procedure. Matching has been widely used in political science applications, and it relies on the same identification assumption as the OLS regression (selection-on-observables), i.e., that there are no variables other than the model covariates that affect both the outcome (election results) and the explanatory variable (fraud). It has, nevertheless, two major advantages: a less restrictive functional form assumption, and it guarantees that the control and treatment groups are balanced on the selected covariates (Sekhon, 2009).

For this purpose, I define the explanatory variable non-linearly, as an indicator of the top 25% administrations in fraud.[105] The matching procedure used here is the entropy balancing method proposed by Hainmueller (2012). This method assigns weights to the observations in the control group (i.e. municipalities that are in the bottom 75% in fraud) to ensure that treatment and control are balanced on all the selected covariates, and the groups are comparable. In order to minimize the potential omitted variables bias arising from the selection-on-observables assumption, I condition the effects on 40 different covariates. These include characteristics of

[105] Or the top 25% administration with a higher share of incomes reported below R$70 in the election year (2012) in CadUnico.

the municipality such as GDP, poverty rate and regional location; of the mayor, such as gender, education, occupation and assets; and of parties, such as coalition status and ideology. Table C.6 in the appendix shows that, as expected, several municipal characteristics are imbalanced before matching. The number of households reporting low income is expected to be correlated with some observed variables such as GDP, and poverty rate. However, all covariates are perfectly balanced after matching.

4.5 Result and Interpretation

In this Section I present the results, their interpretation, and their link to the theoretical framework.

4.5.1 Do Reelection Incentives Increase Fraud?

The results show that mayors with reelection incentives promote more fraud in BF program than lame-duck mayors. Table 4.2 shows the local average treatment effects obtained with the RDD, for the sample including all households that updated/enrolled on CadUnico in 2012 (election year), and for all municipalities with a share of poor families above 67%.[106] Under the main empirical specification (bandwidth of 16pp including all controls), the share of CadUnico-eligible households reporting income below R$70 is 5.4pp higher in municipalities were mayors have reelection incentives (7% higher). Similar results are found for the share of CadUnico-eligible households reporting income below R$140 (2.1pp higher), and BF-eligible households reporting income below R$70 (3.9pp higher). These results are robust to the exclusion of controls, state and party effects, and to different bandwidths (Table 4.2).

[106]In line with the model prediction, the effects described here are not observed for a sample of ”wealthier” municipalities, as shown in the appendix (Table 4.5).

In addition to the robust evidence of income underreporting fraud, the results from this section provide two interesting insights on the mechanism in place. First, from the larger universe of CadUnico eligible households, much higher effects are observed for income reported below R$70 (eligible to the full BF benefit), as opposed to R$140 (eligible to some BF benefits). This implies that the core of the fraud is to get households that are already poor (and likely eligible to some BF benefit) to receive the full BF benefit.[107] This practice is in line with the distribution of reported income shown in Figure 4.1, and well illustrated by the model. That is, this type of fraud has a lower search cost to the local offices (many of these households come to update their CadUnico profile regularly), it is harder to be detected by audits, and this beneficiaries are less costly to monitor (no extra health or school conditions attached).

Second, Figure 4.2 shows the distribution of municipalities across the discontinuity, according to their income underreporting level. Although the effects are not identified for points apart from the discontinuity ($MV_m = 0$), a visual examination shows that, for municipalities with less competitive elections on the right (with reelection incentives), the effects of higher fraud disappear. This is line with the theoretical prediction ($\frac{\partial f_1}{\partial \psi} > 0$), i.e., mayors elected by a large vote margin in 2008 are expected to rely less on fraud to be reelected in 2012.

4.5.2 The Magnitude of the Losses Triggered by Fraud

In Dec 2012, municipalities with reelection incentives had a total BF spending that was 7% higher than municipalities governed by lame-duck mayors (Table 4.2). To put this figure in perspective, it represents R$0.3mn in annual expenses for a municipality at the discontinuity, which is roughly five times the average cost of an election campaign. Additionally, it represents roughly three times what the

[107] which includes the monthly payment of R$70 with no conditions attached

municipal administration receives form the federal government to spend in the management of CadUnico and update/enrollment efforts. While 7% might not seem a major leakage for a program of the scale of BF, it represents a major source of resources that can be controlled by mayors looking for reelection.

Additionally, these effects are slightly conservative. It might take several months for the actual distribution of benefits to reflect changes in the income reported in recent enrollment/updates.[108] Table 4.2 illustrates this temporary disparity between income reporting and benefits. At the end of 2012, the share of households receiving the full BF benefit is not significantly higher for the entire set of beneficiaries in municipalities with reelection incentives. However, in 2014, the share of households receiving the full BF benefit is already significantly higher (by 0.6pp) for the entire set of beneficiaries treated municipalities, where a higher BF spending is also observed (a 12% higher spending).

4.5.3 Other Robustness, Placebo and Specification Tests

Figure 4.3 (full results in Table C.3 in the appendix) shows that, in line with the expectations, fraud is only observed in near the election (in 2011, and stronger in 2012), being absent from the two first years of the mayoral tenure. Table 4.3 shows that reelection incentives had no impact on both the performance of mayors in managing the BF program, measured by the components of the IGD index, and on contemporaneous economic growth, measured by the GDP growth between 2011-12 and 2010-12 in the municipality. These results show that it is unlikely that income underreporting is a spillover from the practices and performance of the administration, as opposed to fraud.

Table 4.5 shows the robustness of the main results to the potential experience and ability biases. The effects on election year-fraud remain significant for both the

[108] The local program administration has no control over the timing, and the actual approval and denial of benefits, see the appendix for a more detailed description of this process.

sub-samples of mayors with similar experience and ability to second-term mayors. In fact, the magnitude of the coefficients is much higher in the high-ability sub-sample, indicating that the main sample might be giving conservative estimates of the actual effects of reelection incentives of fraud. In other words, less political experience and ability might be in fact slightly reducing fraud. The appendix (Table C.5) also shows the results for the sample with balanced age, which produces results that are similar to the main specification.

Households also report other characteristics in CadUnico such as their monthly food and electricity expenses, size of their house, and the existence of a water connection. These variables are not actively used to define program eligibility.[109] It is possible that income underreporting is systematic enough that households also misrepresent the values of all variables consistently. It is more likely, however, that more attention is paid to the only variable that matters for eligibility, i.e., income. Accordingly, there are no significant differences in the effects of reelection incentives on the reporting of these variables in 2012, showing additional evidence that the income underreporting effects observed in the same year are indeed a consequence of fraud.

Finally, using information from the CadUnico, I trace the past income (pre-2012) of all households updating in 2012,[110] and aggregate the municipal variables only for the households that has their last update in the first two years of the mayoral tenure (2009-10). These results can be found in the appendix, Table C.4. There was no significant difference in past income across the groups of municipalities (with and without reelection incentives in 2012). This is evidence that the income reporting effects found in this study are indeed generated only in 2011-2012 (near the election year).

[109] The government uses the additional information in CadUnico to calculate an index of household development (IDF). There was a pilot project in place to use this index for detection of fraud in income reporting, but to the best of my knowledge, this was never implemented before 2012.

[110] This information is not available for households first enrolling in the registry in 2012.

4.5.4 External Accountability: Does Audit Risk Reduce Fraud?

The hypothesis arising from the theoretical model is that a higher risk of audits will reduce fraud in the election year. Table 4.6 shows the coefficient of interest, estimated using equation 4.5. The occurrence of recent audits in a municipality's neighborhood reduces the positive effect of reelection incentives on fraud. In fact, the magnitude of the results indicate that recent audits wipe out all these effects, given that an audit leads 6.8pp less households to report income below R$70, when compared to the effect observed for municipalities without increased audit risk. This result is robust to bandwidth and model specification, but it loses power and magnitude when the neighborhood is contracted to include a narrower radius. Table 4.6 also shows a placebo test. The effects disappear when estimated using either audits that happened after elections.

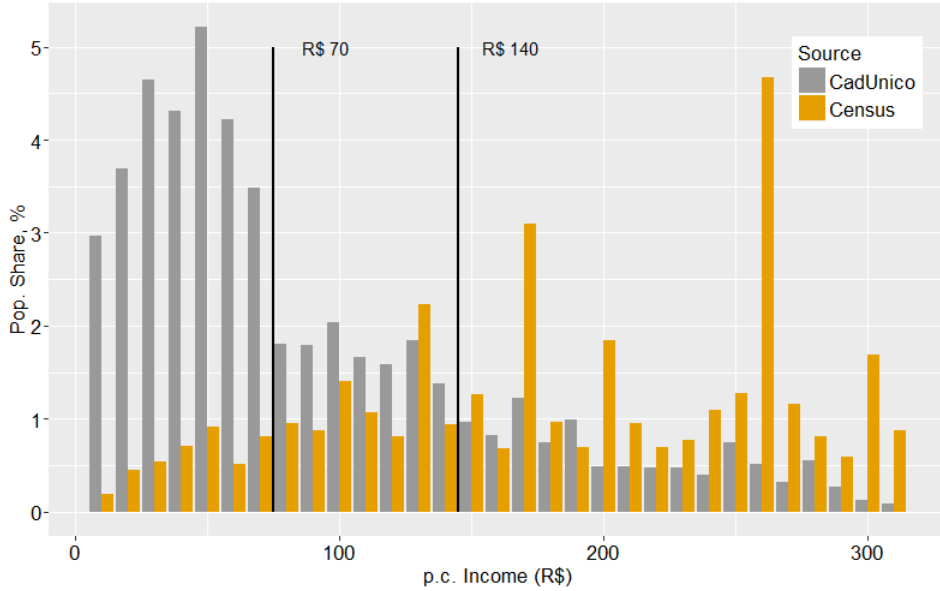
This is evidence that audits are an effective and relatively costly instrument to reduce program fraud, in the presence of reelection incentives. For perspective, the country-wide BF spending was R$20bi in 2012, which puts the estimated impact of fraud at more than R$1bi per year. At the same time, the annual budget of the CGU is currently less than R$100mn.

4.5.5 Does Fraud Help the Incumbent Politician?

In municipalities on the top 25% in fraud, incumbent mayors have a 36% higher probability of reelection (18pp higher), for a 6% higher share of votes (3pp higher). These results are observed for a sub-sample of municipalities where the past election (2008) had a margin of victory of 10pp or less (Table 4.7). They are also in line with both these predictions of the theoretical model: fraud should help incumbents to get reelected, and this effect should be higher where elections are more competitive. While these results are still significant but with a lower magnitude for the municipalities where the 2008 margin of victory was less than 15pp, they fade when the entire sample is used (margin of victory up to 100pp).

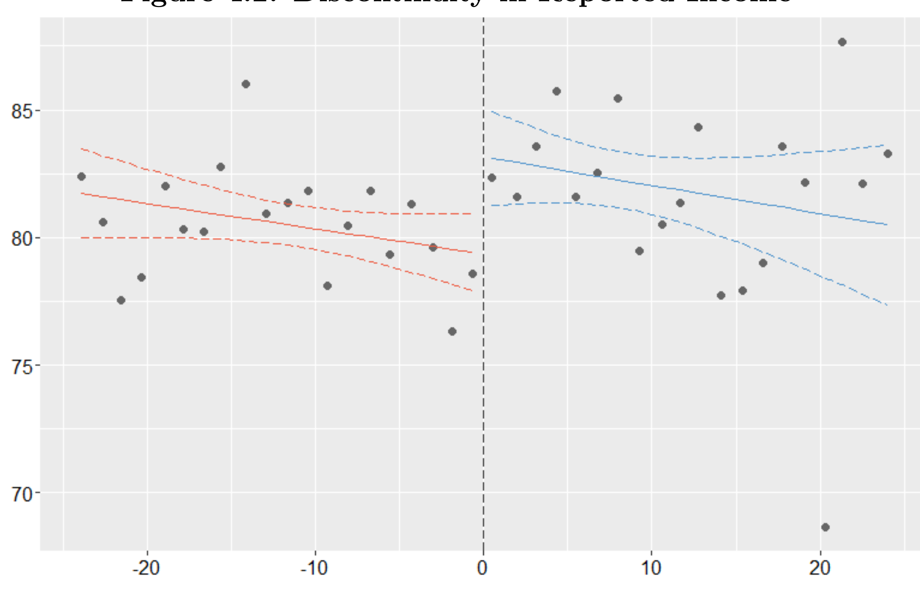
Table 4.7 also shows that being in the group with top 25% election year-fraud is not correlated with contemporaneous performance indicators of the mayors, measured by the components of the IGD index and GDP growth, as before. The electoral rewards of fraud are not correlated with the mayor's performance in other areas (BF management and economic growth). Finally, Figure 4.4 shows the results of an additional placebo test for the matching specification. We estimate the electoral rewards of being a top 25% municipality in low reported income, based on updates made by on CadUnico on the pre-election period (2009-2011). If the results found in this section come from some omitted variable bias that causes both a higher incumbency advantage and lower reported income, then the coefficients observed for the years other than 2012 should also be significant and strong. The fact that no effects are observed for these years is evidence that the electoral rewards estimated under this specification are coming from income reporting fraud in 2012 only.

Figure 4.1: Reported Monthly Income in Brazil: CadUnico vs. Census



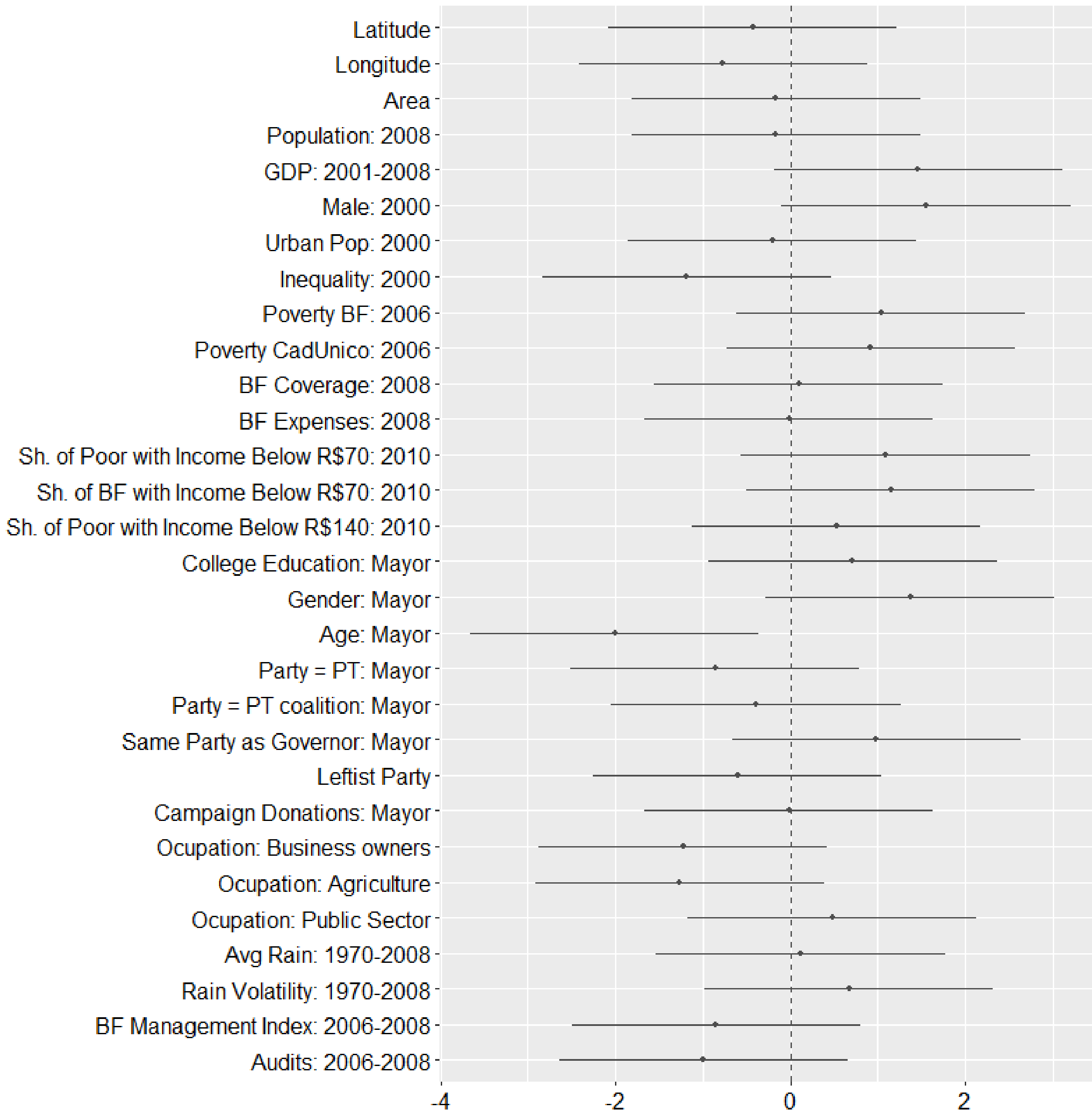
Data from the sample of the 2010 Census, and CadUnico updated as of Dec 2012. It excludes the households that reported zero income (10% of the CadUnico sample, and 11% of the Census sample).

Figure 4.2: Discontinuity in Reported Income



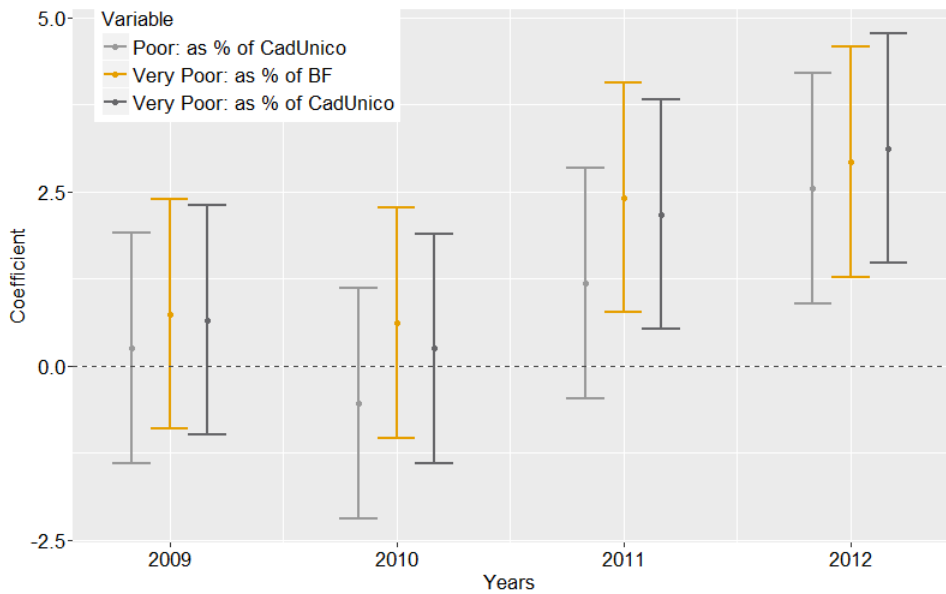
The right side shows the municipalities with reelection incentives. The solid lines are the local fit of a second degree polynomial, and the dashed lines are the 90% confidence intervals. The sample includes municipalities in the two upper terciles in poverty.

Figure 4.3: Balance of Fixed or Pre-Determined Variables



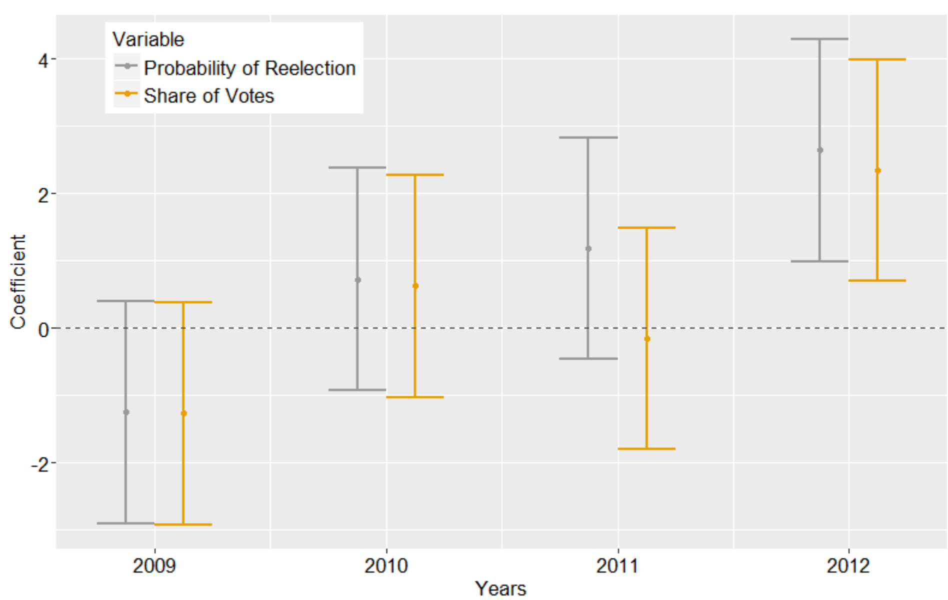
The chart shows the normalized confidence intervals (C.I.) for a confidence level of 90%. The C.I. are heteroskedasticity robust.

Figure 4.4: The Impact of Reelection Incentives on Fraud, by Update Year



The regression is estimated using the full specification, under a bandwidth of 16pp. The bars show the 90% confidence interval.

Figure 4.5: The Impact of Fraud on Electoral Performance, by Update Year



The regression is estimated using a matching algorithm, where the explanatory variable is a dummy indicating if the municipality is top 25% in fraud. The bars show the 90% confidence interval.

Table 4.1: Main Results: Fraud (Income Underreporting)

	PT Mean	Coefficient, [S.E.]				Band.
Depended Variable	(1)	(2)	(3)	(4)	(5)	{Obs.}
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*Optimal*
Excludes Municipal Covariates, State dummies, and Region-Party dummies						
Very Poor, ri≤R$70	81.227	4.848**	5.157**	5.196***	5.024**	15.0
(% of CadUnico Eligible)	[0.458]	[2.180]	[1.997]	[1.841]	[2.049]	{589}
Very Poor, ri≤R$70	88.464	3.576**	3.796**	3.828***	3.690**	14.8
(% of BF Eligible)	[0.341]	[1.641]	[1.502]	[1.383]	[1.553]	{585}
Poor, ri≤R$140	91.435	1.642	1.813*	1.848**	1.867**	17.0
(% of CadUnico Eligible)	[0.218]	[1.067]	[0.974]	[0.894]	[0.947]	{644}
Includes State dummies and Region-Party dummies						
Very Poor, ri≤R$70		6.001***	5.944***	5.768***	5.894***	15.0
(% of CadUnico Eligible)		[2.095]	[1.902]	[1.753]	[1.956]	{589}
Very Poor, ri≤R$70		4.199***	4.150***	4.023***	4.128***	14.8
(% of BF Eligible)		[1.615]	[1.453]	[1.330]	[1.511]	{585}
Poor, ri≤R$140		2.366**	2.383***	2.330***	2.409***	17.0
(% of CadUnico Eligible)		[0.955]	[0.880]	[0.817]	[0.858]	{644}
Full Specification (includes Municipal Covariates, State dummies, and Region-Party dummies)						
Very Poor, ri≤R$70		5.500***	5.444***	5.263***	5.375***	15.0
(% of CadUnico Eligible)		[1.921]	[1.744]	[1.611]	[1.794]	{589}
Very Poor, ri≤R$70		3.977***	3.916***	3.769***	3.885***	14.8
(% of BF Eligible)		[1.478]	[1.336]	[1.229]	[1.388]	{585}
Poor, ri≤R$140		2.058**	2.084**	2.049***	2.104***	17.0
(% of CadUnico Eligible)		[0.893]	[0.819]	[0.760]	[0.800]	{644}
Observations		544	617	698		

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table 4.2: Distribution of Benefits

Depended Variable	PT Mean (1)	Coefficient, [S.E.] (2)	(3)	(4)	(5)	Band. {Obs.}
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*Optimal*
Sample Aggregating All Households in CadUnico (MDS Data at the Municipal Level)						
full Benefits, Dec 2012	97.465	0.509	0.563	0.574*	0.579	17.1
(% of total benefits)	[0.099]	[0.422]	[0.376]	[0.339]	[0.362]	{651}
BF pc Benefit, Dec 2012[a]	5.064	0.012	0.016	0.014	0.014	19.2
(R$ '000/yr)	[0.006]	[0.024]	[0.022]	[0.020]	[0.020]	{700}
BF tot Benefits, Dec 2012[a]	12.741	0.067*	0.073**	0.067**	0.066**	19.4
(R$ mn/yr)	[0.034]	[0.037]	[0.034]	[0.032]	[0.032]	{704}
full Benefits, Dec 2014	98.409	0.618	0.618*	0.574*	0.534*	22.5
(% of total benefits)	[0.089]	[0.384]	[0.338]	[0.304]	[0.280]	{748}
BF pc Benefit, Dec 2014[a]	5.245	0.069**	0.065**	0.058**	0.066**	14.4
(R$ '000/yr)	[0.007]	[0.030]	[0.027]	[0.025]	[0.029]	{577}
BF tot Benefits, Dec 2014[a]	12.959	0.122**	0.119***	0.107***	0.113***	17.8
(R$ mn/yr)	[0.034]	[0.049]	[0.045]	[0.041]	[0.043]	{669}
Sample Aggregating Households that Updated their Information in the Election Year (2012)						
full Benefits, Dec 2012	97.862	0.458	0.499	0.523*	0.524	17.6
(% of total benefits)	[0.088]	[0.386]	[0.341]	[0.307]	[0.322]	{663}
BF Benefits, Dec 2012	78.544	-1.541	-1.247	-1.004	-1.371	14.8
(% of total enrollment)	[0.290]	[1.317]	[1.222]	[1.130]	[1.258]	{585}
BF pc Benefit, Dec 2012[a]	5.086	0.027	0.031	0.028	0.030	17.0
(R$ '000/yr)	[0.006]	[0.025]	[0.023]	[0.022]	[0.023]	{648}
BF tot Benefits, Dec 2012[a]	12.112	0.150	0.159*	0.149*	0.130*	23.3
(R$ mn/yr)	[0.038]	[0.093]	[0.086]	[0.079]	[0.072]	{760}
Observations		544	617	698		

[a]Coefficient estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table 4.3: Other Household Characteristics Reported in CadUnico

	PT Mean	Coefficient, [S.E.]				Band.
	(1)	(2)	(3)	(4)	(5)	{Obs.}
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*Optimal*
Sample of Households that Updated their Information in the Election Year (2012)						
Food Expenses	48.555	-3.584	-3.205	-2.897	-3.218	15.5
(pc avg)	[0.588]	[2.494]	[2.277]	[2.097]	[2.312]	{599}
Electricity Expenses	6.717	-0.332	-0.264	-0.176	-0.168	19.3
(pc avg)	[0.097]	[0.384]	[0.352]	[0.327]	[0.326]	{700}
Rooms	1.344	0.034	0.029	0.025	0.030	15.5
(pc avg)	[0.010]	[0.032]	[0.029]	[0.027]	[0.030]	{599}
Floor	63.823	-5.317*	-4.663	-4.166	-4.865	15.0
(share of built floor)	[0.798]	[3.215]	[2.966]	[2.738]	[3.044]	{587}
Water		544	617	698		
(share of water connection)	[0.500]	[2.251]	[1.705]	[1.265]	[1.435]	{2164}
Observations		872	1626	2556		

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table 4.4: Measures of the Mayor's Performance in 2012

	PT Mean	Coefficient, [S.E.]				Band.
	(1)	(2)	(3)	(4)	(5)	{Obs.}
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*Optimal*
School Conditionalities	0.895	0.016	0.016	0.015	0.016	13.5
(index in 0-1, 2012)	[0.004]	[0.014]	[0.013]	[0.012]	[0.014]	{555}
Health Conditionalities	0.835	-0.010	-0.005	-0.002	-0.010	13.0
(index in 0-1, 2012)	[0.005]	[0.022]	[0.020]	[0.018]	[0.022]	{544}
CadUnico Updates	0.793	-0.006	-0.004	-0.005	-0.004	16.0
(index in 0-1, 2012)	[0.004]	[0.017]	[0.016]	[0.015]	[0.016]	{619}
CadUnico Coverage	0.992	0.002	0.002	0.001	0.002	17.4
(index in 0-1, 2012)	[0.002]	[0.007]	[0.007]	[0.006]	[0.007]	{656}
GDP growth	0.003	-0.009	-0.016	-0.018	-0.006	12.5
(%, 2012/2011)	[0.006]	[0.016]	[0.015]	[0.015]	[0.016]	{531}
GDP growth	0.050	0.020	0.006	0.001	0.012	14.5
(%, 2012/2010)	[0.007]	[0.026]	[0.025]	[0.025]	[0.026]	{579}
Observations		544	617	698		

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table 4.5: Robustness to the Experience and Ability Biases

	P.T. Mean	Coefficient, [S.E.]				Band.
	(1)	(2)	(3)	(4)	(5)	{Obs.}
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*Optimal*
EXPERIENCE: mayors with reelection incentives that have previous political experience						
Very Poor, ri≤R$70	80.896	5.395**	5.012**	4.582**	5.032**	16.3
(% of CadUnico Eligible)	[0.534]	[2.207]	[2.106]	[1.961]	[2.114]	{383}
Very Poor, ri≤R$70	88.216	3.528**	3.291**	3.060**	3.332**	15.9
(% of BF eligible)	[0.394]	[1.733]	[1.655]	[1.532]	[1.674]	{372}
Poor, ri≤R$140	91.301	2.313**	2.161**	1.939**	1.896**	16.9
(% of CadUnico Eligible)	[0.261]	[1.031]	[0.966]	[0.893]	[0.875]	{394}
Observations		403	462	528	454	
ABILITY: mayors with reelection incentives that ran and won in 2012						
Very Poor, ri≤R$70	82.371	9.242***	9.186***	8.695***	9.139***	14.0
(% of CadUnico Eligible)	[0.500]	[2.165]	[2.039]	[1.920]	[2.025]	{419}
Very Poor, ri≤R$70	89.360	6.951***	6.783***	6.393***	6.887***	13.4
(% of BF eligible)	[0.364]	[1.645]	[1.542]	[1.448]	[1.547]	{407}
Poor, ri≤R$140	91.844	3.282***	3.424***	3.240***	3.366***	14.9
(% of CadUnico Eligible)	[0.242]	[1.083]	[1.013]	[0.944]	[0.992]	{432}
Observations		397	460	531	419	

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table 4.6: Effect of Audits on Fraud, by Different Specifications

	PT Mean	Coefficient, [S.E.]				
	(1)	(2)	(3)	(4)	(5)	(6)
Depended Variable: (Share of Poor with Income below R$70)						
Bandwidth	*16.0*	*16.0*	*16.0*	*13.0*	*16.0*	*19.0*
Effect of Recent Audits (2009-10)						
Micro-region,	0.567	-4.641	-5.100	-7.336**	-6.779**	-6.170*
(1=any audit)	[0.020]	[3.911]	[3.471]	[3.661]	[3.406]	[3.153]
Micro-region +[a]	0.608	-5.193	-6.293*	-7.695**	-7.175**	-6.559**
(1=any audit)	[0.020]	[3.933]	[3.492]	[3.684]	[3.424]	[3.166]
Micro-region, 111km rad.	0.532	-3.276	-5.973*	-8.033**	-7.298**	-6.752**
(1=any audit)	[0.020]	[4.019]	[3.454]	[3.648]	[3.378]	[3.125]
Micro-region, 55km rad.	0.364	1.267	-2.147	-3.235	-2.981	-2.657
(1=any audit)	[0.019]	[4.260]	[3.512]	[3.685]	[3.415]	[3.190]
Placebo Test						
Micro-region	0.334	-2.091	0.574	-1.017	-0.152	0.084
(1=any audit, 2012-13)	[0.019]	[4.328]	[3.566]	[4.011]	[3.703]	[3.389]
Observations	617	617	617	544	617	698
State, Party-Region Dum.	N	Y	Y	Y	Y	Y
Controls	N	N	N	Y	Y	Y

[a]It also includes past audits in the municipality. Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the audit dummies (i.e the share of municipalities that experienced audits in their neighborhood) for a municipality at the discontinuity segment, before treatment.

Table 4.7: Impact of BF Fraud on the Probability of Reelection

	PT Mean		Coefficient, [S.E.]	
Depended Variable	(1)	(2)	(3)	(4)
Margin of Victory, 2008	*≤15pp*	*≤10pp*	*≤15pp*	*All*
Election Results				
Share of Votes, 2012	47.970	3.060**	2.115*	-0.315
(%)	[11.984]	[1.308]	[1.185]	[0.983]
Prob. of Reelection, 2012	48.354	17.611***	12.017**	1.726
(%)	[50.036]	[6.680]	[5.609]	[4.708]
Mayoral Performance Index				
Health Conditionalities, 2011-12	0.844	-0.002	-0.007	-0.016
(index in 0-1)	[0.130]	[0.020]	[0.014]	[0.013]
School Conditionalities, 2011-12	0.905	-0.007	-0.017	-0.013
(index in 0-1)	[0.072]	[0.013]	[0.013]	[0.009]
CadUnico Coverage, 2011-12	0.995	0.003	-0.001	-0.002
(index in 0-1)	[0.024]	[0.005]	[0.004]	[0.003]
CadUnico Update, 2011-12	0.801	-0.005	-0.003	-0.003
(index in 0-1)	[0.090]	[0.015]	[0.012]	[0.010]
Overall Index (Igd), 2011-12	0.886	-0.003	-0.007	-0.009*
(index in 0-1)	[0.048]	[0.007]	[0.006]	[0.005]
GDP Growth, 2012-11	-0.332	0.020	0.016	0.012
(% yoy)	[11.413]	[0.012]	[0.011]	[0.010]
GDP Growth, 2012-10	4.688	-0.001	-0.006	-0.006
(% yoy)	[14.516]	[0.018]	[0.015]	[0.012]
Observations	527	406	527	740

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The regression was weighted by weights generated in the matching procedure. The explanatory variable is a dummy indicating if the municipality is top 25% in fraud (% of CadUnico-eligible households reporting income below R$70).

Chapter 5

Conclusion

This thesis studies the impact of a CCT program in the local politics of Brazil, a developing nation plagued by clientelism, vote buying and corruption. Chapters 2 and 3 show that, when transfers are properly shielded from the influence and manipulation of local politicians, the CCT program reduces incumbency advantage, increases both electoral competition and the quality of candidates, and weakens the support for clientelistic parties. Cash transfers also contribute to the political enfranchisement of the poor by shifting spending into redistributive health and education services. The theory reconciles these empirical findings by showing that cash transfers reduce the ability of the incumbents to raise support with vote buying. These effects are observed over a nine-year period.

Chapter 4 examines the relationship between CCT and local politics from a different angle. When the program can be used by politicians to influence voters, the possibility of reelection aligns corruption with both the preferences of voters and the politician's objective, giving raise to a mechanism of reverse accountability. This Chapter exposes one way in which politicians illegally extract rents from the BF program, and use them to increase their reelection prospects. By promoting income underreporting fraud, mayors can make not-so-poor households eligible to receive federal funds. These households will, in turn, reward mayors with votes. This Chapter also shows that this reverse accountability calls for external disciplining devices to keep politicians and voters in check. The expectation of being audited is shown to drastically reduce the effects of reelection incentives on income reporting fraud in the BF program.

Both the approach of the first two Chapters, and the approach of Chapter 4, estimate the same thing: the effect of higher CCT coverage on local politics. The results obtained with the different approaches, however, point in the opposite direction. While Chapter 2 and 3 show that CCT coverage reduces incumbency advantage and increases electoral competition, Chapter 4 shows that incumbents can benefit from the program with the use of fraud. These results can be re-conciled after considering the nature and the persistence of each effect, in the context of Brazil. The negative effects shown in Chapter 4 are characteristic of the period when beneficiaries are first joining the program, and the manipulation of eligibility by politicians can create credit claiming opportunities. Similar results already exist in the literature (De Janvry, Finan, and Sadoulet, 2012; Labonne, 2013). However, given the program characteristics (long permanence, easy access to the central government office for complaints), and the local political environment (two-term limit for mayors), these effects are likely short-lived.

In Chapter 2 and 3, the effects are not only estimated for a longer period (two election cycles), but they also represent a lasting impact of the CCT program. The mechanism proposed requires that beneficiaries see the CCT income shock as permanent, in a way that would change their consumption decisions, and utility. Accordingly, these effects should last as long as the CCT benefit is stable, and should survive through different political cycles and incumbents (they do not concern the relationship between voters and one specific politician, i.e., the one that might have given them access to CCT). In other words, after a short-period in which credit can be claimed by politicians for getting the program benefits to some voters, the long-term effects of the CCT program that come through the impact of the permanent income increase on vote buying should be positive. They should work against incumbents that replace public good with clientelistic exchanges.

This set of results also has various policy implications. First, Chapter 3 suggests that policy spillovers matter. In addition to the political impacts of the CCT program, it shows that a small differential in funding for a health program generated a much larger impact on the CCT distribution. Second, CCT programs may increase the ability of national governments to shape local politics in their favor. This is even more significant when considering that the literature has shown that politicians are able to reap electoral rewards, at the national level, by implementing the program. Third, when properly shielded from local political interests, a CCT program can be treated as an exogenous income increase. In this context, these findings are useful to inform other policies that aim for a similar impact on the incomes of the poor.

These final results also have significant policy implications. In a program of the scale and scope of BF (and in many other CCT programs around the world), even a small source of fraud could represent billions in losses. This study estimates losses with fraud that could surpass R$1bn a year. Put in perspective, disciplining devices as audits are an very efficient strategy to reduce fraud, at a relatively low cost (the budget of the auditing unit CGU is less than R$100mn/year).

The overall findings in this thesis raise at least one important question that might be theme for future research: what are the social welfare implications of the political effects of a CCT program? There are a few potential developments here.

First, while Chapters 2 and 3 show that the CCT program, when effectively bypassing local politicians, is positive for the poor sectors of the population, the overall welfare effects remain uncertain. Anticipating the long-term consequences of less capital-intensive public goods are beyond the ability of the identification strategy employed here. Also, the magnitude of the effects has to be treated carefully when applied to a more general context. The level of the coefficients may well depend on institutional features that are specific to the Brazilian case, and the magnitude may still be affected by a residual direct impact of the FHP funding, even though the evidence on the exclusion restriction convincingly supports the direction of the results presented here.

Second, to the extent that politicians can defraud the CCT program, corruption might be offsetting part of the welfare improvements generated by the program. Although this thesis shows that fraud can counteract some of the incumbency advantage reduction effects shown in Chapter 3, a precise estimation of the overall impact of fraud on the entire range of political outcomes is beyond the scope of the empirical strategy employed here (a back of the envelop calculation, however, indicates that the long-term positive effects of the CCT are at least four times larger than the short-lived negative effects). At least from a pure economic standpoint, it is likely that fraud is inefficient from a social welfare perspective. Assuming that the marginal utility of consumption increases as income decreases, fraud is welfare reducing if very poor households cannot access the full BF benefit due to income underreporting by undeserving households. In 2011, the government estimated that nearly 800 thousand poor households in Brazil did not have the benefit, even though the average BF coverage across municipalities was above 100%.

Bibliography

Alston, Lee J. and Bernardo Mueller. 2006. "Pork for Policy: Executive and Legislative Exchange in Brazil." *Journal of Law, Economics, and Organization* 22 (1):87–114.

Alves, Denisard and Chris Timmins. 2003. "Social Exclusion and the Two-Tiered Healthcare System of Brazil." *In J.R. Behrman, A.G. Trujillo, and M. Szekely (eds.), Who is In and Who is Out: Social Exclusion in Latin America. Washington, DC: Inter-American Development Bank.* .

Ames, Barry and Amy Erica Smith. 2010. "Knowing Left from Right: Ideological Identification in Brazil, 2002-2006." *Journal of Politics in Latin America* 2 (3):3–38.

Anderson, Siwan, Patrick Francois, and Ashok Kotwal. 2015. "Clientelism in Indian Villages." *American Economic Review* 105 (6):1780–1816.

Baez, Javier E., Adriana Camacho, Emily Conover, and Roman A. Zarate. 2012. "Conditional Cash Transfers, Political Participation, and Voting Behavior." *Policy Research Working Papers* .

Barro, Robert J. 1973. "The control of politicians: An economic model." *Public Choice* 14 (1):19–42.

Boas, Taylor C. and F. Daniel Hidalgo. 2011. "Controlling the Airwaves: Incumbency Advantage and Community Radio in Brazil." *American Journal of Political Science* 55 (4):869–885.

Boix, Carles and Susan C. Stokes. 2009. "Political Clientelism." *The Oxford Handbook of Comparative Politics* .

Brollo, Fernanda and Tommaso Nannicini. 2012. "Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil." *American Political Science Review* 106:742–761.

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. 2013. "The Political Resource Curse." *American Economic Review* 103 (5):1759–96.

Brollo, Fernanda and Ugo Troiano. 2016. "What Happens When a Woman Wins an Election? Evidence from Close Races in Brazil." *Working Paper* .

Brusco, Valeria, Marcelo Nazareno, and Carol Stokes. 2004. "Vote Buying in Argentina." *Latin American Research Review* 39 (2):66–88.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6):2295–2326.

Camacho, Adriana and Emily Conover. 2011. "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy* 3 (2):41–65.

Clark, Damon and Paco Martorell. 2014. "The Signaling Value of a High School Diploma." *Journal of Political Economy* 122 (2):282–318.

Cruz, Cesi, Philip Keefer, and Julien Labonne. 2015. "Incumbent Advantage, Voter Information, and Vote Buying." *Working Paper* .

Cruz, Cesi and Christina J. Schneider. 2016. "Foreign Aid and Undeserved Credit Claiming." *Working Paper* .

De Janvry, Alain, Frederico Finan, and Elisabeth Sadoulet. 2012. "Local electoral incentives and decentralized program performance." *Review of Economics and Statistics* 94 (3):672–685.

De La O, Ana L. 2013. "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico." *American Journal of Political Science* 57 (1):1–14.

———. 2015. *Crafting Policies to End Poverty in Latin America.* Cambridge University Press.

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78 (6):1863–1903.

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

Duggan, John and César Martinelli. 2016. "The Political Economy of Dynamic Elections: Accountability, Commitment, and Responsiveness." *Working Paper* .

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7 (1):1–26.

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder. 2015. "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races." *American Journal of Political Science* 59 (1):259–274.

Eggers, Andrew C. and Jens Hainmueller. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103:513–533.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50 (1/3):5–25.

Ferraz, Claudio and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123 (2):703–745.

———. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101 (4):1274–1311.

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

Fiszbein, Ariel, Norbert Schady, Francisco H.G. Ferreira, Margaret Grosh, Niall Keleher, Pedro Olinto, and Emmanuel Skoufias. 2009. "Reducing Present and Future Poverty. World Bank Policy Research Report." *World Bank Policy Research Report* .

Fried, Brian J. 2012. "Distributive Politics and Conditional Cash Transfers: The Case of Brazil's Bolsa Familia." *World Development* 40 (5):1042–1053.

Fujiwara, Thomas. 2015. "Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil." *Econometrica* 83 (2):423–464.

Fujiwara, Thomas and Leonard Wantchekon. 2013. "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." *American Economic Journal: Applied Economics* 5 (4):241–55.

Gerber, Alan S., Daniel P. Kessler, and Marc Meredith. 2011. "The Persuasive Effects of Direct Mail: A Regression Discontinuity Based Approach." *The Journal of Politics* 73:140–155.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20 (1):25–46.

Handa, Sudhanshu and Benjamin Davis. 2006. "The Experience of Conditional Cash Transfers in Latin America and the Caribbean." *Development Policy Review* 24 (5):513–536.

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

Hidalgo, F. Daniel and Simeon Nichter. 2015. "Voter Buying: Shaping the Electorate through Clientelism." *American Journal of Political Science* .

IBOPE. 2005. "Compra de Votos nas Eleicoes 2004." *Ibope Opinião* .

Imbens, Guido W. and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3):933–959.

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142 (2):615 – 635.

IPEA. 2011. "SIPS - Sistema de Indicadores de Percepção Social." *IPEA* .

Jacob, Brian A and Lars Lefgren. 2004. "Remedial education and student achievement: A regression-discontinuity analysis." *Review of economics and statistics* 86 (1):226–244.

Keele, Luke J. and Rocío Titiunik. 2014. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* .

Labonne, Julien. 2013. "The local electoral impacts of conditional cash transfers: Evidence from a field experiment." *Journal of Development Economics* 104 (0):73 – 88.

Larreguy, Horacio, John Marshall, and Laura Trucco. 2015. “Breaking Clientelism or Rewarding Incumbents? Evidence from an Urban Titling Program in Mexico.” *Working Paper* .

Lee, David S. 2008. “Randomized experiments from non-random selection in U.S. House elections.” *Journal of Econometrics* 142 (2):675 – 697.

Lee, David S. and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48 (2):281–355.

Lichand, Guilherme, Marcos F. M. Lopes, and Marcelo C. Medeiros. 2016. “Is Corruption Good For Your Health?” *Working Paper* .

Lindert, Kathy, Anja Linder, Jason Hobbs, and Benedicte de la Briere. 2007. “The Nuts and Bolts of Brazil s Bolsa Familia Program: Implementing Conditional Cash Transfers in a Decentralized Context.” *World Bank Working Papers* .

Manacorda, Marco, Edward Miguel, and Andrea Vigorito. 2011. “Government Transfers and Political Support.” *American Economic Journal: Applied Economics* 3 (3):1–28.

Nichter, Simeon. 2008. “Vote Buying or Turnout Buying? Machine Politics and the Secret Ballot.” *American Political Science Review* 102:19–31.

Pande, Rohini and Benjamin Olken. 2012. “Corruption in Developing Countries.” *Annual Review of Economics* 4:479–505.

Papay, John P., John B. Willett, and Richard J. Murnane. 2011. “Extending the regression-discontinuity approach to multiple assignment variables.” *Journal of Econometrics* 161 (2):203 – 207.

Persson, Torsten and Guido E. Tabellini. 2000. “Political Economics: Explaining Economic Policy.” *Cambridge, Massachussets: MIT Press* .

Querubin, Pablo and James M. Snyder. 2013. "The Control of Politicians in Normal Times and Times of Crisis: Wealth Accumulation by U.S. Congressmen, 1850–1880." *Quarterly Journal of Political Science* 8 (4):409–450.

Reardon, Sean F. and Joseph P. Robinson. 2010. "Regression discontinuity designs with multiple rating-score variables." *Center for Education Policy Analysis Working Papers* .

Rego, Walquiria L. Rego and Alessandro Pinzani. 2013. "Vozes do Bolsa Familia." *Editora Unesp* .

Rodriguez-Chamussy, Lourdes. 2015. "Local Electoral Rewards from Centralized Social Programs: Are Mayors Getting the Credit?" *IDB Working Paper Series* (550).

Ruppert, D. and M. P. Wand. 1994. "Multivariate Locally Weighted Least Squares Regression." *The Annals of Statistics* 22 (3):1346–1370.

Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12 (1):487–508.

Soares, Fábio Veras, Rafael Perez Ribas, and Rafael Guerreiro Osório. 2010. "Evaluating the impact of Brazil's Bolsa Familia: Cash transfer programs in comparative perspective." *Latin American Research Review* 45 (2):173–190.

Stokes, Susan, Thad Dunning, Marcelo Nazareno, and Valeria Brusco. 2013. "Brokers, Voters and Clientelism." *New York: Cambridge University Press* .

Stokes, Susan C. 2005. "Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina." *American Political Science Review* null:315–325.

Sugiyama, Natasha Borges and Wendy Hunter. 2013. "Whither Clientelism? Good Governance and Brazil's Bolsa Familia Program." *Comparative Politics* 46 (1):43–62.

Vicente, Pedro C. and Leonard Wantchekon. 2009. "Clientelism and vote buying: lessons from field experiments in African elections." *Oxford Review of Economic Policy* 25 (2):292–305.

Wong, Vivian C., Peter M. Steiner, and Thomas D. Cook. 2013. "Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables A Comparative Study of Four Estimation Methods." *Journal of Educational and Behavioral Statistics* 38 (2):107–141.

Zajonc, Tristan. 2012. "Essays on Causal Inference for Public Policy." *PhD Dissertation. Harvard University* .

Zucco, Cesar. 2013. "When Payouts Pay Off: Conditional Cash-Transfers and Voting Behavior in Brazil: 2002-2010." *American Journal of Political Science* 47 (3).

Appendix A

Appendix to Chapter 2

A.1 Proofs of Propositions

Proof of proposition 1. The incumbent maximizes the total utility of voters given by the equation below:

$$\begin{aligned} TOT.U_I = \alpha^H \phi^H [\delta \log(w) + (1-\delta)\log(b - t - (1-s)v)] \\ +\alpha^L \phi^L [\delta \log(y) + (1-\delta)\log(t - sv)] \\ +\alpha^S \phi^S [\delta \log(y + v\eta\theta) + (1-\delta)\log(t - sv)] \end{aligned}$$

The first order conditions for the maximization problem are:

$$t : \alpha^H \phi^H \frac{(1-\delta)}{b - t - (1-s)v} = \alpha^L \phi^L \frac{(1-\delta)}{t - sv} + \alpha^S \phi^S \frac{(1-\delta)}{t - sv}$$

$$v : \alpha^H \phi^H \frac{(1-\delta)(1-s)}{b - t - (1-s)v} + (\alpha^L \phi^L + \alpha^S \phi^S) \frac{(1-\delta)s}{t - sv} = \alpha^S \phi^S \eta\theta \frac{\delta}{y + v\eta\theta}$$

The equations above can be solved for for t and v , with $\mu = \frac{\delta}{(1-\delta)}$:

$$t = \frac{(\alpha^S \phi^S \mu s + \phi - \alpha^H \phi^H) b\eta\theta + y(\phi(1-s) - \alpha^H \phi^H)}{(\alpha^S \phi^S \mu + \phi)\eta\theta}$$

$$v = \frac{\alpha^S \phi^S \mu b \eta \theta - \phi y}{(\alpha^S \phi^S \mu + \phi)\eta\theta}$$

Taking the value of v calculated above, this is the derivative with respect to income:

$$\frac{dv}{dy} = -\frac{\phi}{(\alpha^S \phi^S \mu + \phi)\eta\theta}$$

Proof of proposition 2. Using the results obtained above, the value of $\frac{dt}{dy}$ and $\frac{d(b-t)}{dy}$ are given by:

$$\frac{dt}{dy} = \frac{\phi(1-s) - \alpha^H \phi^H}{(\alpha^S \phi^S \mu + \phi)\eta\theta} \quad \text{and} \quad \frac{d(b-t)}{dy} = -\frac{\phi(1-s) - \alpha^H \phi^H}{(\alpha^S \phi^S \mu + \phi)\eta\theta}$$

The condition on the parameters that give $\frac{dt}{dy} > 0$ are given by the inequality below:

$$\frac{(\alpha^L + z\alpha^S)\phi^L}{\alpha^H \phi^H} > \frac{s}{(1-s)}$$

It remains to show that this result would not be obtained in the absence of vote buying. Take the utility below, given by the incumbent, in a model without v and without the swing group.

$$TOT.U_I = \alpha^H \phi^H [\delta \log(w) + (1-\delta)\log(b-t)] + \alpha^L \phi^L [\delta \log(y) + (1-\delta)\log(t)]$$

The maximization problem can be solved for the value of t, which does not depend on income, as follows:

$$t = \frac{\alpha^L \phi^L b}{\phi}$$

Proof of proposition 3. Equation 2.3 shows the total utility faced by the voter, from both the challenger and the incumbent. The incumbent's vote share is given by $\pi_I = TOT.U_I - TOT.U_C + \lambda$. Substituting into that equation the values of t and v found before, and rearranging, I have:

$$\pi_I = \alpha^L \phi^L \delta \log(\frac{y}{y + v\epsilon\theta}) + \alpha^S \phi^S \delta \log(\frac{y + v\eta\theta}{y + v\epsilon\theta}) - \lambda + \frac{1}{2}$$

Taking the derivative with respect to income, and rearranging,

$$\frac{d\pi_I}{dy} = \delta\theta[\alpha^L \phi^L \frac{y + v\epsilon\theta}{y}(\frac{\epsilon(v - y\frac{dv}{dy})}{(y + v\epsilon\theta)^2}) - \alpha^S \phi^S \frac{y + v\epsilon\theta}{y + v\eta\theta}(\frac{(v - y\frac{dv}{dy})(\eta - \epsilon)}{(y + v\epsilon\theta)^2})] \tag{A.1}$$

For ($\frac{d\pi_I}{dy} < 0$), the term in brackets needs to be negative, which results in the following condition:

$$z = \frac{\phi^S}{\phi^L} \geq \frac{y + v\eta\theta}{y}$$

The ratio of the density of the political preference parameters for the swing and poor groups has to be at least the ratio of the marginal utilities of consumption for the same groups. The condition can also be written as a function of the parameters, as:

$$z > \frac{\phi^L \alpha^S \mu(\frac{b}{y}\eta\theta + 1) - (\phi - \alpha^S \phi^S)}{(\mu + 1)\phi^L \alpha^S})$$

Proof of proposition 4. First, I show that vote buying is decreasing on the density of political preference. For the value of v calculated above, the derivative with respect to ϕ^L is:

$$\frac{dv}{d\phi^L} = \frac{-\alpha^L \alpha^S \phi^L z\mu(b\eta\theta + y)}{(\alpha^S \phi^S \mu + \phi)^2 \eta^2 \theta^2}$$

and the derivative of this equation with respect to income is:

$$\frac{dv}{d\phi^L dy} = \frac{-\alpha^L \alpha^S \phi^L z\mu}{(\alpha^S \phi^S \mu + \phi)^2 \eta^2 \theta^2}$$

Now, the derivative of the incumbent's vote share with respect to income, shown in equation A.1, simplifies to:

$$\frac{d\pi_I}{dy} = \delta\epsilon\theta \frac{1}{y + v\epsilon\theta}(v - y\frac{dv}{dy})[\alpha^L \phi^L z \frac{1}{y} - \alpha^S \phi^L z \frac{1}{y + v\eta\theta} \frac{(\eta - \epsilon)}{\epsilon}]$$

Rearranging the terms even further, and using the identity $\frac{(\eta-\epsilon)}{\epsilon} = \frac{\alpha^L}{\alpha^S}$, I have:

$$\frac{d\pi_I}{dy} = \delta\epsilon\theta\alpha^L A[\phi^L(\frac{1}{y} - zB)]$$

Where,

$$B = \frac{1}{(y + v\eta\theta)} = \frac{\alpha^S \phi^L z\mu + \phi}{\alpha^S \phi^L z\mu(b\eta\theta + y)} \quad \text{with} \quad \frac{dB}{d\phi^L} = \frac{(\alpha^L \phi^L - \alpha^H \phi^H)\alpha^S z\mu(b\eta\theta + y)}{(\alpha^S \phi^L z)^2 \mu^2 (b\eta\theta + y)^2}$$

Defining $\alpha = \frac{\eta-\epsilon}{\eta} = \frac{\alpha^L}{\alpha^L+\alpha^S}$, I also have:

$$(v - y\frac{dv}{dy}) = \frac{\alpha^S \phi^L z\mu b}{\alpha^S \phi^L z\mu + \phi} \quad \text{and} \quad (y + v\epsilon\theta) = \frac{\alpha^S \phi^L z\mu(b\epsilon\theta + y) + \phi\alpha y}{\alpha^S \phi^L z\mu + \phi}$$

And,

$$A = \frac{(v - y\frac{dv}{dy})}{(y + v\epsilon\theta)} = \frac{\alpha^S \phi^L z\mu b}{\alpha^S \phi^L z\mu(b\epsilon\theta + y) + \phi\alpha y} \quad \text{with} \quad \frac{dA}{d\phi^L} = \frac{b\alpha\mu y\alpha^S z\alpha^H \phi^H}{[\alpha\phi^L z\mu(b\epsilon\theta + y) + \phi\alpha y]^2} > 0$$

Now, the equation of interest is the following:

$$\frac{d\pi_I}{dyd\phi^L} = \delta\epsilon\theta\alpha^L \frac{dA}{d\phi^L}[\phi^L(\frac{1}{y} - zB)] + \delta\epsilon\theta\alpha^L A[(\frac{1}{y} - zB) - \phi^L z \frac{dB}{d\phi^L}]$$

Rearranging,

$$\frac{d\pi_I}{dyd\phi^L} = \delta\epsilon\theta\alpha^L[\frac{dA}{d\phi^L}\phi^L(\frac{1}{y} - zB) + A(\frac{1}{y} - zB) - A\phi^L z \frac{dB}{d\phi^L}]$$

The equation above has three terms inside the brackets. From the assumption supporting the result of Proposition 3, I have $(\frac{1}{y} - zB) < 0$. Thus, because $\frac{dA}{d\phi^L} < 0$, the first two terms inside the brackets are negative. If the sum of these two terms has an absolute value that is higher than $A\phi^L z \frac{dB}{d\phi^L}$, then $\frac{d\pi_I}{dyd\phi^L} < 0$. A stronger condition to achieve this result is to simply require that $\frac{dB}{d\phi^L} \geq 0$. This occurs if $\alpha^L\phi^L \geq \alpha^H\phi^H$, i.e., the poor group in the population is a relatively more attractive target for public good distribution than the wealthy group. Even if, $\alpha^L\phi^L < \alpha^H\phi^H$, the result holds if the value of $\frac{dB}{d\phi^L}$ is small enough.

Now for the prediction related to budget allocation, the derivative of $\frac{dt}{dy}$ with respect to ϕ^L gives the result, as shown below.

$$\frac{dt}{dyd\phi^L} = \frac{(\alpha^L + z\alpha^S)\alpha^H\phi^H}{(\alpha^S\phi^S\mu + \phi)^2\eta\theta} > 0$$

Proof of proposition 5. With A and B defined as before, I use the derivative of the incumbent's vote share with respect to income shown in equation A.1, which simplifies to:

$$\frac{d\pi_I}{dy} = \delta\epsilon\theta\alpha^L A[\phi^L(\frac{1}{y} - zB)]$$

This result depends on the value of $\frac{d\pi_I}{dyd\alpha^L}$, which is calculated maintaining the size of the swing group fixed (i.e. an increase in α^L causes a decrease of the same magnitude in α^H). Defining $C = [\phi^L\frac{1}{y} - \phi^S B] < 0$, the equation of interest is the following:

$$\frac{d\pi_I}{dyd\alpha^L} = -\frac{\alpha^L}{\alpha^L + \alpha^S}\delta\epsilon\theta AC + \delta\epsilon\theta AC + \delta\epsilon\theta\alpha^L\frac{dA}{d\alpha^L}C + \delta\epsilon\theta\alpha^L A\frac{dC}{d\alpha^L} \tag{A.2}$$

Taking the derivative of A with respect to α^L,

$$\frac{dA}{d\alpha^L} = \frac{\alpha^S\phi^S\mu b[\alpha^S\phi^S\mu(b\epsilon\theta - \phi y) - (\phi^L - \phi^H)\alpha^L y]}{(\alpha^L + \alpha^S)^2(\alpha^S\phi^S\mu(b\epsilon\theta + y) + \phi\alpha y)^2}$$

The condition that makes $\frac{dA}{d\alpha^L} > 0$ is $\alpha^S\phi^S\mu(b\epsilon\theta - \phi y) > (\phi^L - \phi^H)\alpha^L y$. Now taking the derivative of B with respect to α^L,

$$\frac{dB}{d\alpha^L} = \frac{(\phi^L - \phi^H)}{\alpha^S\phi^S\mu(b\eta\theta + y)}$$

The derivative is positive if $(\phi^L - \phi^H)$. Equation A.2 has four terms. The sum of the first and second term is negative. For $\frac{d\pi_I}{dyd\alpha^L} < 0$, it is sufficient to have one of the last two terms being negative and higher than the other. For simplicity, I will present the conditions to make both terms negative. If $\phi^L > \phi^H$, the last term is negative, so as long as the difference between these densities is small, the third term is also negative. If, however, $\phi^L < \phi^H$, automatically the third term is negative and the fourth is positive. As long as we have $-\frac{dA}{d\alpha^L}[\phi^L(\frac{1}{y} - zB)] > -z\phi^L\frac{dB}{d\alpha^L}$, the result also holds. Overall, the relative distance between the densities in the political preferences of the wealthy and poor groups should be small, independent on the direction it takes.

Now for the prediction related to budget allocation, I take the derivative of the equation below with respect to α^L:

$$\frac{dt}{dy} = \frac{\phi(1-s) - \alpha^H\phi^H}{(\alpha^S\phi^S\mu + \phi)\eta\theta}$$

which is positive as follows:

$$\frac{dt}{dyd\alpha^L} = \frac{(\phi^L - \phi^H)[(1-s)\alpha^S\phi^S\mu + \alpha^H\phi^H] + \phi^H(\alpha^S\phi^S\mu + \phi)}{(\alpha^S\phi^S\mu + \phi)^2\eta\theta} > 0$$

Finally, we show the that $\frac{dv}{dyd\alpha^L} < 0$ if $\phi^L > \phi^H$, as follows:

$$\frac{dv}{dyd\alpha^L} = \frac{-(\phi^L - \phi^H)(b\eta\theta + y)\alpha^S\phi^S\mu}{(\alpha^S\phi^S\mu + \phi)^2\eta\theta}$$

Proof of proposition 6. Now, if the politicians differ in type, we might have different allocations of t and v across candidates. For the incumbent, taking the first derivative of the voter's utility with respect to income, and substituting the optimal values, I have,

$$\frac{dTOT.U_I}{dy} = (1-\delta)\frac{\phi s + \alpha^H\phi^H}{y} + \alpha^L\phi^L\delta\frac{1}{y} + (1-\delta)\frac{(\alpha^L\phi^L + \alpha^S\phi^S)^2}{\phi b\eta\theta + y(\alpha^L\phi^L + \alpha^S\phi^S)} + \alpha^S\phi^S\delta\frac{1}{b\eta\theta + y}$$

Taking the derivative with respect to type,

$$\frac{dTOT.U_I}{dyd\theta} = b\eta[-(1-\delta)\frac{\phi(\alpha^L\phi^L + \alpha^S\phi^S)^2}{[\phi b\eta\theta + y(\alpha^L\phi^L + \alpha^S\phi^S)]^2} - \delta\frac{\alpha^S\phi^S}{(b\eta\theta + y)^2}] < 0$$

With $\frac{dU_I}{dyd\theta} < 0$, a clientelistic type would always have a higher utility fall coming from an income shock than a public good-type of politician. As for the allocation to pro-poor goods, it is easy to see that if $\frac{dt}{dy} > 0$, then $\frac{dt}{dyd\theta} < 0$, as it is given by the expression below:

$$\frac{dt}{dyd\theta} = -\frac{\phi(1-s) - \alpha^H\phi^H}{(\alpha^S\phi^S\mu + \phi)\eta\theta^2}$$

A.2 A Generic Model

In this section, I present the same model using a generic form for the utility function. This version shows the sensitivity of the main theoretical predictions to the assumption regarding the complementary of the consumption of public and private goods in the voter's utility. I continue to assume that the utility is concave, and all the properties for maximization apply. In this section, I denote the first and second derivatives of the utility using subscripts.[111] Also, define $\alpha^H\phi^H = H$, $\alpha^L\phi^L = L$ and $\alpha^S\phi^S = S$, for simplicity.

Proposition A1. *A CCT program will reduce vote buying ($\frac{dv}{dy} < 0$) if the consumption of private and public goods is complementary or neutral in the voter's utility ($U_{gc} \geq 0$), and $U_{gcc} \geq 0$. If they are substitute, or $U_{gcc} < 0$, the result holds for a small enough magnitude of U_{gc}.*

The first order conditions for maximization are now:

$$t : -HU_g^H + LU_g^L + SU_g^S = 0$$

$$v : -H(1-s)U_g^H - LsU_g^L - SsU_g^S + S\eta\theta U_c^S = 0$$

Taking the total derivative of the equations above, and rearranging,

$$[HU_{gg}^H + LU_{gg}^L + SU_{gg}^S]dt + [H(1-s)U_{gg}^H - sSU_{gg}^S - sLU_{gg}^L + S\eta\theta U_{gc}^S]dv + [LU_{gc}^L + SU_{gc}^S]dy = 0 \tag{A.3}$$

[111] For example, U_{gg}^L denotes the second derivative of the utility of group L with respect to the consumption of the public good.

$$[H(1-s)U_{gg}^{H} - sSU_{gg}^{S} - sLU_{gg}^{L} + S\eta\theta U_{gc}^{S}]dt + [H(1-s)^2U_{gg}^{H} + \gamma s^2U_{gg}^{L} + Ss^2U_{gg}^{S} - Ss\eta\theta U_{gc}^{S} + S\eta^2\theta^2U_{cc}^{S}]dv + [-LsU_{gc}^{L} - SsU_{gc}^{S} + S\eta\theta U_{cc}^{S}]dy = 0 \quad (A.4)$$

Now for simplicity, I rewrite the equations above as $adt + bdv + ddy = 0$ and $bdt + cdv + edy = 0$. The two equation system can now be solved as $\frac{dv}{dy} = \frac{bd-ae}{ac-b^2}$. By the second order conditions of the maximization problem, $ac - b^2 > 0$. Also, after substituting equations A.3 and A.4 I have:

$$bd = [H(1-s)U_{gg}^{H} - sSU_{gg}^{S} - sLU_{gg}^{L} + S\eta\theta U_{gc}^{S}] * [LU_{gc}^{L} + SU_{gc}^{S}]$$

$$-ae = [HU_{gg}^{H} + LU_{gg}^{L} + SU_{gg}^{S}] * [LsU_{gc}^{L} + SsU_{gc}^{S} - S\eta\theta U_{cc}^{S}]$$

These terms can be rearranged into the equations below:

$$TERM.1 = HSU_{gg}^{H}U_{gc}^{S} + (HU_{gg}^{H} + S\eta\theta U_{gc}^{S})LU_{gc}^{L}$$

$$TERM.2 = -(HU_{gg}^{H} + LU_{gg}^{L})S\eta\theta U_{cc}^{S} + S^2\eta\theta(-U_{gg}^{S}U_{cc}^{S} + (U_{gc}^{S})^2)$$

$$\frac{dv}{dy} = \frac{TERM.1 + TERM.2}{ac - b^2} \quad (A.5)$$

By the concavity of the utility function, $TERM.2 < 0$. In the sequence, I examine potential cases for the values of U_{gc}^{L} and U_{gc}^{S} from $TERM.1$. First, it is easy to see that, if $U_{gc} = 0$, then $TERM.1 < 0$ and $\frac{dv}{dy} < 0$.

Now, consider the case in which $U_{gc} > 0$. If $U_{gcc} \geq 0$, then $U_{gc}^{S} > U_{gc}^{L}$. Replacing the value of U_{gc}^{L} by U_{gc}^{S} in the equation above, we have the following condition:

$$U^S_{gc} < \frac{HU^H_{gg}SU^S_{gc} - HU^H_{gg}S\eta\theta U^S_{cc} + S\eta\theta(L+S)(-U^S_{gg}U^S_{cc} + (U^S_{gc})^2)}{-HLU^H_{gg}}$$

The last term in the nominator was set at its highest value when I imposed $U^L_{gc} = U^S_{gc}$, but it remains negative. Thus, the entire term on the right side of the inequality above is negative, which means that any positive values of U^L_{gc} and U^S_{gc} would satisfy the conditions for $\frac{dv}{dy} < 0$. If in fact $U_{gcc} < 0$, and $U^S_{gc} < U^L_{gc}$, it is easy to see from equation A.5 that $\frac{dv}{dy} < 0$ for any values of U^L_{gc}, if $-HU^H_{gg} > S\eta\theta U^S_{gc}$. This is still a stronger condition than the one required for the result, as if it does not hold, we can still have $\frac{dv}{dy} < 0$ if:

$$U^L_{gc} < \frac{HU^H_{gg}SU^S_{gc} - (HU^H_{gg} + LU^L_{gg})S\eta\theta U^S_{cc} + S^2\eta\theta(-U^S_{gg}U^S_{cc} + (U^S_{gc})^2)}{-L(HU^H_{gg} + S\eta\theta U^S_{gc})}$$

Now, consider the case in which $U_{gc} < 0$. The following condition must hold for the result:

$$HSU^H_{gg}U^S_{gc} + (HU^H_{gg} + S\eta\theta U^S_{gc})LU^L_{gc} < (HU^H_{gg} + LU^L_{gg})S\eta\theta U^S_{cc} - S^2\eta\theta(-U^S_{gg}U^S_{cc} + (U^S_{gc})^2)$$

Proposition A2. *If there is no vote buying, a CCT program will shift the budget allocation towards the pro-poor public good ($\frac{dt}{dy} > 0$) only if the consumption of private and public goods is complementary in the voter's utility ($U_{gc} > 0$). With vote buying, the result also holds if public and private consumption are substitute or neutral ($U_{gc} \leq 0$), for small enough values of U_{gc} and s.*

If public and private consumption are complements, CCT will increase the marginal utility of the public good for the poor, relatively more than for the wealthy, because y increases and w does not. In this case, $\frac{dt}{dy} > 0$ even without vote buying, as shown below. However, if they are substitutes or neutral, the result can only be achieved with vote buying. It follows from Proposition A1 that, under certain conditions, the incumbent will reallocate resources from v back into t and $b-t$. If $U_{gc}=0$, than the income shock will have no direct effect on the marginal utility of the public good for the poor, so the amount of t and $b-t$ reallocated will depend on three things. First, the relative marginal utilities U_g^H , U_g^L and U_g^S. Second, the relative attractiveness of each group as an electoral target. Third, it will depend on s.

First, I show that without clientelism the result can only be achieved if $U_{gc}^L>0$. The first order condition is:

$$t : -HU_g^H + LU_g^L = 0$$

Taking the total derivative of the equation and rearranging, $\frac{dt}{dy}$ is positive if and only if $U_{gc}>0$, by the second order conditions for maximization of a concave function.

$$\frac{dt}{dy} = \frac{U_{gc}^L}{-HU_{gg}^H - LU_{gg}^L}$$

Now, using the notation from Proposition A1, the system shown in equations A.3 and A.4 can be solved as $\frac{dt}{dy}=\frac{be-cd}{ac-b^2}$. Again, by the second order conditions of the maximization problem, $ac-b^2>0$. After substituting equations A.3 and A.4, I have:

$$be = [H(1-s)U_{gg}^H - sSU_{gg}^S - sLU_{gg}^L + S\eta\theta U_{gc}^S] * [-LsU_{gc}^L - SsU_{gc}^S + S\eta\theta U_{cc}^S]$$

$$-cd = [H(1-s)^2U_{gg}^H + Ls^2U_{gg}^L + Ss^2U_{gg}^S - 2S\eta\theta sU_{gc}^S + S\eta^2\theta^2U_{cc}^S] * [-LU_{gc}^L - SU_{gc}^S]$$

The signal of the sum of these two terms defines the signal of $\frac{dt}{dy}$. I will split these two terms in the following four terms:

$$T1 = -s[HU_{gg}^{H} + SU_{gg}^{S} + LU_{gg}^{L}] * S\eta\theta U_{cc}^{S}$$

$$T2 = -SL\eta^2\theta^2 U_{cc}^{S} U_{gc}^{L}$$

$$T3 = -H(1-s)U_{gg}^{H} * [LU_{gc}^{L} + SU_{gc}^{S}]$$

$$T4 = S\eta\theta s U_{gc}^{S} * [LU_{gc}^{L} + SU_{gc}^{S}] + SH\eta\theta U_{gg}^{H} U_{cc}^{S}$$

So $\frac{dt}{dy} > 0$ if $T1 + T2 + T3 + T4 > 0$. First, $T4$ is always positive and $T1$ is always negative. Also, it is easy to see that, if $U_{gc} \geq 0$, then $T2 + T3 > 0$, by the concavity of the utility function. It remains the following condition on s to achieve the result.

$$s < \frac{TERM.1 + TERM.2}{[HU_{gg}^{H} + SU_{gg}^{S} + LU_{gg}^{L}] * S\eta\theta U_{cc}^{S}}$$

where

$$TERM.1 = -SL\eta^2\theta^2 U_{cc}^{S} U_{gc}^{L} - H(1-s)U_{gg}^{H} * [LU_{gc}^{L} + SU_{gc}^{S}]$$

$$TERM.2 = S\eta\theta s U_{gc}^{S} * [LU_{gc}^{L} + SU_{gc}^{S}] + HS\eta\theta U_{gg}^{H} U_{cc}^{S}$$

From the equation above, if $U_{gc} < 0$, then $T2+T3 < 0$. Because the first term in the nominator is positive for a concave function, small enough values of U_{gc} will give the result. The condition on s remains, as long as $T4 > -(T1 + T2 + T3)$. Finally, for $U_{gc} = 0$, we have that $T2 = T3 = 0$, so the result holds if $-H(1-s)U_{gg}^{H} > -s(SU_{gg}^{S} + LU_{gg}^{L})$.

Proposition A3. *If the density of the political preference of the swing group is high enough around zero, then there are parameter values for which CCT reduce the vote share of the incumbent.*

First, it is easy to see that the result cannot be achieved without vote buying, as there is no incumbency advantage. Equation 2.1 gives the vote share of the incumbent. The total derivative of this equation with respect to income is zero without vote buying.

Now with vote buying, taking the total differential of equation 2.3 for the incumbent's portion, I have:

$$\begin{aligned}\frac{dTOT.U_I}{dy} &= -HU_g^{IH}[\frac{dt}{dy} + (1-s)\frac{dv}{dy}] + LU_g^{IL}[\frac{dt}{dy} + s\frac{dv}{dy}] \\ &+SU_g^{IS}[\frac{dt}{dy} + s\frac{dv}{dy}] + LU_c^{IL} + SU_c^{IS} + S\eta\theta U_c^{IS}\frac{dv}{dy}\end{aligned}$$

For the challenger's portion, both poor and swing groups will receive the same amount of public and private transfers. Therefore, $U^{CL} = U^{CS}$, and the total differential is:

$$\begin{aligned}\frac{dTOT.U_C}{dy} &= -HU_g^{CH}[\frac{dt}{dy} + (1-s)\frac{dv}{dy}] + (L+S)U_g^{CL}[\frac{dt}{dy} + s\frac{dv}{dy}] \\ &+(L+S)U_c^{CL} + (L+S)\epsilon\theta U_c^{CL}\frac{dv}{dy}\end{aligned}$$

First, we know that, with equal allocations, we have $HU_g^{CH} = HU_g^{IH}$. Now, the change in the vote share of the incumbent after CCT is defined by:

$$\frac{d\pi_I}{dy} = \frac{dt}{dy}(LU_g^{IL} + SU_g^{IS} - (L+S)U_g^{CL}) + LU_c^{IL} + SU_c^{IS} - (L+S)U_c^{CL}$$
$$+\frac{dv}{dy}(S\eta\theta U_c^{IS} - (L+S)\epsilon\theta U_c^{CL} + LsU_g^{IL} + SsU_g^{IS} - (\gamma+S)sU_g^{CL})$$

Rearranging, and using the identity $\phi^S = z\phi^L$, we find the equation for z that supports $\frac{d\pi_I}{dy} < 0$.

$$z > \frac{\alpha^L[\frac{dt}{dy}(U_g^{IL} - U_g^{CL}) + (U_c^{IL} - U_c^{CL}) + \frac{dv}{dy}(-\epsilon\theta U_c^{CL} + sU_g^{IS} - sU_g^{CL})]}{-\alpha^S[\frac{dt}{dy}(U_g^{IS} - U_g^{CL}) + (U_c^{IS} - U_c^{CL}) + \frac{dv}{dy}(\eta\theta U_c^{IS} - \epsilon\theta U_c^{CL} + sU_g^{IS} - sU_g^{CL})]}$$

A.3 Incorporating Rent Seeking

The new allocations for v and t in the simpler version of the rent seeking model are seen below. Under these assumptions, there is always a positive amount of rent being extracted, but there is no change in the model predictions for $\frac{dv}{dy}$, $\frac{dt}{dy}$ or $\frac{d\pi_I}{dy}$.

$$v = \frac{\alpha^S\phi^S\eta\theta\mu(b-\bar{r}) - \phi y}{(\alpha^S\phi^S\mu + \phi)\eta\theta}$$

$$t = \frac{(\alpha^S\phi^S\mu s + \phi - \alpha^H\phi^H)(b-\bar{r})\eta\theta + y(\phi(1-s) - \alpha^H\phi^H)}{(\alpha^S\phi^S\mu + \phi)\eta\theta} + \frac{s\bar{r}}{\alpha^S\phi^S\mu\eta\theta}$$

For the more refined model, the first order conditions are:

$$r : \gamma = \frac{R\beta\psi}{\phi}[\alpha^H\phi^H\frac{(1-\delta)(1-s)}{b-t-(1-s)(v+r)} + (\alpha^L\phi^L + \alpha^S\phi^S)\frac{(1-\delta)s}{t-s(v+r)}]$$

$$t : \alpha^H \phi^H \frac{(1-\delta)}{b-t-(1-s)(v+r)} = \alpha^L \phi^L \frac{(1-\delta)}{t-s(v+r)} + \alpha^S \phi^S \frac{(1-\delta)}{t-s(v+r)}$$

$$v : \alpha^H \phi^H \frac{(1-\delta)(1-s)}{b-t-(1-s)(v+r)} + (\alpha^L \phi^L + \alpha^S \phi^S) \frac{(1-\delta)s}{t-s(v+r)} = \alpha^S \phi^S \eta\theta \frac{\delta}{y+v\eta\theta}$$

And the values of $\frac{dv}{dy}$, $\frac{dt}{dy}$ and $\frac{dr}{dy}$ are:

$$\frac{dt}{dy} = \frac{(\phi(1-s) - \alpha^H \phi^H)\delta R \beta \psi \phi}{\eta\theta(\delta R \beta \psi(\alpha^S \phi^S \mu + \phi) + \gamma\phi\mu)}$$

$$\frac{dv}{dy} = \frac{-\phi(\delta R \beta \psi + \gamma\mu)}{\eta\theta(\delta R \beta \psi(\alpha^S \phi^S \mu + \phi) + \gamma\phi\mu)}$$

$$\frac{dr}{dy} = \frac{\phi\gamma\mu}{\eta\theta(\delta R \beta \psi(\alpha^S \phi^S \mu + \phi) + \gamma\phi\mu)}$$

For the reelection probability, the intuition is that a higher maximum rent ($\bar{r}$) reduces the incumbent's vote share because the expected utility of voters from the incumbent in period 2 is lower, while it is not for the challenger. This can been seen in the equation below:

$$\begin{aligned} TOT.U_I = {} & \alpha^H \phi^H (1-\delta) \log(b-t-(1-s)(v+0.5(\bar{r}+r))) + \alpha^L \phi^L \delta \log(y) \\ & +(\alpha^L \phi^L + \alpha^S \phi^S)(1-\delta) \log(t-s(v+0.5(\bar{r}+r))) + \alpha^S \phi^S \delta \log(y+v\eta\theta) \end{aligned}$$

However, the income shock would also have a lower effect on the fall in vote shares ($\frac{d\pi_I}{dy}$), as the incumbency advantage created by vote buying is already less effective with the existence of rents, as described above. Below I show the new equation for the vote shares.[112] The last two terms are the same as before, and the sum of their derivatives with respect to income should be negative to sustain $\frac{d\pi_I}{dy} < 0$.

$$\begin{aligned}\pi_I = \alpha^H\phi^H(1-\delta)\log(\frac{b-t-(1-s)(v+0.5(\bar{r}+r))}{b-t-(1-s)(v+r)}) \\ +(\alpha^S\phi^S+\alpha^L\phi^L)\delta\log(\frac{t-s(v+0.5(\bar{r}+r))}{t-s(v+r)}) \\ \frac{(1-s)}{2}\frac{dr}{dy}(b-t-(1-s)(v+0.5(\bar{r}+r)))]\end{aligned}$$

The first two terms differ from the previous model, and their derivative with respect to income is positive, as seen below. This means that the existence of rents imposes stronger conditions on the minimum necessary strength of the vote buying networks to achieve $\frac{d\pi_I}{dy} < 0$.

$$\begin{aligned}\frac{d\pi_I(1st)}{dy} = cons1*[(-\frac{dt}{dy}-(1-s)\frac{dv}{dy}-\frac{(1-s)}{2}\frac{dr}{dy})(\frac{(1-s)}{2}(\bar{r}-r)) \\ +\frac{(1-s)}{2}\frac{dr}{dy}(b-t-(1-s)(v+0.5(\bar{r}+r)))]\end{aligned}$$

$$\text{with}\quad cons1 = \alpha^H\phi^H(1-\delta)\frac{1}{[b-t-(1-s)(v+r)][b-t-(1-s)(v+0.5(\bar{r}+r)]}$$

$$\frac{d\pi_I(2nd)}{dy} = cons2*[(\frac{dt}{dy}-s\frac{dv}{dy}-\frac{s}{2}\frac{dr}{dy})\frac{s}{2}(\bar{r}-r)+\frac{s}{2}\frac{dr}{dy}(t-s(v+0.5(\bar{r}+r)))]$$

[112] To simplify the analysis, assume that both politicians have an equal true type R.

$$\text{with} \quad cons2 = (\alpha^S \phi^S + \alpha^L \phi^L)\delta \frac{1}{[t - s(v + r)][t - s(v + 0.5(\bar{r} + r))]}$$

Appendix B

Appendix to Chapter 3

B.1 Effects on Hiring

The empirical evidence in this study indicates that the budget shift to redistributive services of health and education happens in tandem with an increase in the budget share spent with personnel. I briefly examine the available hiring data[113] to provide insight on how this budget reallocation can affect public services in the long term. The estimation results are show in Table B.7.

In Brazil, public servants are hired through a competitive, meritocratic process, which includes a written exam. Employees hired through this process have job security guaranteed by the legislation. Mayors are allowed to bypass the process in two ways. First, the creation of ”political” positions, usually reserved for high level executive jobs in the administration. These are often allocated to allies, or used as political exchange. Second, the hiring of temporary employees for a specific project and a limited tenure.

Table B.7 shows the effects, at the discontinuity, on both the total employment and the share of the permanent work force hired in political jobs. These effects are

[113] Hiring data comes from the IBGE annual survey of municipalities conducted in 2004, 2008 and 2012. Not all municipalities reported information and the 1999 survey has no data on interns or temporary employees. For this reason, I only use data from 2004-12. Where the 2008 and 2012 surveys report both separate categories for interns and temporary employees, the 2004 survey only reports the interns category. The patterns in the data indicate that this category also includes temporary employees, which I will assume that it does. I only include in the sample the municipalities that had a non-zero change in the calculated variables, as a zero change it is more likely to be a reporting error. Nevertheless, the results are not sensitive to this exclusion. Finally, for consistence the sample is limited to the municipalities for which the budget allocation data is also available.

shown for both the level and rate of change in these variables. The only noteworthy effect is a 8.2pp higher change in the share of temporary hiring in 2008-12. In the data, temporary hiring is generally positively correlated with budget changes.[114] This indicates that it might be a procedure used by mayors to conduct budget changes quickly, without committing to long term labor costs.

The education level of the labor force provides another evidence that this type of hiring is consonant with the budget reallocation narrative in this study. The share of temporary workers with less than a high school degree is 17pp lower for treated municipalities. This is in line with the expectation that a shift to health and education would increase the demand for high-skilled jobs (e.g. doctors, health agents, teachers), while the fall in capital investment should reduce the demand for low-skilled jobs (construction workers).

While these results are consistent with the budget shift, their persistence over the long term might have implications for public good distribution that are still unknown, and beyond the scope of this study. If temporary jobs outlive the budget shift and are not replaced by permanent ones[115], there are at least two potential issues arising. First, there is a trade-off between quality and effort in public service. Where the lack of a formal selection process for temporary jobs might negatively impact the quality of public service, the absence of job security might work as a mechanism to extract a higher effort. Second, the persistence of this type of hiring could be evidence that mayors are using a different avenue to conduct clientelistic exchanges with the wealthier population. Because there is no significant effect observed in 2008-12 in the political hires, the evidence for this argument seems limited now, but it might worth examining in the future.

[114] I calculate a coefficient of variation for the budget allocation, as the sum of the squared changes in budget shares across tenures, for 2004-12. This coefficient is positively correlated (0.1) to the absolute change in the share of temporary employees.

[115] In 2007-2011, 77% of the municipalities opened at least one formal hiring process. However, there is no significant difference between treated and non-treated locations.

B.2 Bandwidth Selection

Bandwidth selection is a significant component of RDDs, as the bandwidth is usually the tool used to control the trade-off between bias and efficiency in the estimation. Although there are plenty of approaches to select the bandwidth in one-score RD designs,[116] there is still sparse literature discussing similar methodologies for the MRDD. In fact, most of the literature applying some form of MRDD does not discuss the issue at all.[117] I will use a plug-in algorithm for bandwidth selection building on the work of Zajonc (2012). In this section I briefly discuss the main practical challenges in implementing the procedure in the multivariate context. I describe the algorithm on that paper and how it tackles those problems, and I propose alterations. All technical notation in this section is taken from that paper to facilitate reference.

The method follows the plug-in algorithm developed for the single-score case in Imbens and Kalyanaraman (2012). In a nutshell, plug-in methods aim to find an optimal bandwidth by the minimization of an expression for the mean squared error (MSE) at the cut-off, in three main steps. First, a theoretical expression for the minimum MSE is calculated, as a function of bandwidth and other parameters from the data. The MSE expression is a combination of terms for the bias and the variance in the estimation. Second, these parameters are estimated using the data, with the exception of the bandwidth. Third, they are plugged back into the original MSE expression to derive the optimal bandwidth.

The main difficulties for the two-score case are described here. First, the plug-in method does not have a closed form solution for more than one bandwidth. Zajonc

[116] See Imbens and Lemieux (2008); Imbens and Kalyanaraman (2012); **?** for various discussions on different methodologies of bandwidth selection for the single-score RDD.

[117] The discussion is absent is most cases either because the multiple variables are collapsed to one and the one-dimensional approaches are used, or the methodology is parametric. Dell (2010) shows the results for different bandwidths in longitude and latitude, but there is no formal approach to defining their levels.

(2012) uses the same bandwidth for both scores for a feasible solution. Second, bandwidths cut the kernel differently along the frontier, so the shape of the treated and non-treated subsets is endogenous to the bandwidth selection near the origin.[118] Here, the solution was to calculate the bandwidth only for points far from the origin. Finally, unreliable bandwidth values can arise due to the assumptions involved in the calculation of the MSE parameters. (Imbens and Kalyanaraman, 2012) correct this problem with a regularization term. Zajonc (2012) calculates the optimal bandwidth for various points and uses the minimum value for under-smoothing.

I propose a different approach for tackling these three problems. First, I allow the bandwidths to be different for each score variable. Thus, instead of using an expression for the single optimal bandwidth, I will minimize the MSE expression numerically for different pairs of bandwidths. This provides efficiency gains in nearly all cases, as the two-dimensional bandwidth outperforms the unique one in terms of MSE. Furthermore, I use an elliptical bandwidth instead of a rectangular one, to ensure that the points within a lower distance from the cut-off are used.[119]

Second, if the sample is not balanced along the frontier, the optimal bandwidth calculated away from the origin will not be optimal, or relevant, for the estimation at the origin. This is a key problem in this study, as the area of the frontier used for the main specification is near the origin. Thus, I expand the algorithm to include the calculation of optimal bandwidths at this point (Pop=30 and HDI=0.7), where the kernel will cut the treatment frontier in a well defined manner. This change requires the estimation of cross derivatives from a second-degree polynomial on the score variables, which was not required before.

Third, for regularization I will run a constrained minimization of the MSE ex-

[118] The origin is defined as the point where the two segments of the two-dimensional treatment frontier connect. In the case of this study, it is where population = 30 and HDI = 0.7

[119] The algorithm will produce the values for the sides of a rectangular bandwidth. I will use an ellipse centered at the same cut-off with an area that equals the area of the rectangle produce by the selection algorithm. It will have radiuses that are slightly higher than the sides of the rectangular bandwidth, but it will exclude the distant points in the corners of the rectangle.

pression, using a cap for the bandwidth. The nature of the problem of bandwidth selection is to find the optimal value, in light of the trade-off between bias and variance. However, whereas variance is salient in the regression results, bias is not. Thus, proposing a cap effectively limits the amount of bias that the researcher is willing to accept, at a cost of higher variance. It remains the issue of setting an appropriate cap. For simplicity, I will run the original algorithm for the first stage (CCT coverage) at points away from the origin, and select one minimum unique bandwidth for the scores, using that as a cap for the new algorithm.

For reference, all the other parameters, including the pilot bandwidths, are kept as proposed by the original algorithm. Finally, whenever the description of an equation is not detailed, it is because it fully replicates a step shown in Zajonc (2012). The algorithm is described below. Items 1-4 are taken from that paper, with a small adjustment to item 4. Steps 5-7 were modified as described above. The score variables are also normalized by their standard deviation to be in a common scale.

1. Using the entire sample, calculate the standard deviations for population σ_p and HDI σ_h.
2. Select a pilot bandwidth for each variable using Scott's rule (e.g.$\hat{h} = \sigma_p n^{-\frac{1}{6}}$) and limit the sample to those bandwidths.
3. Calculate the conditional variance $\hat{v}(p, h)$ and density $\hat{f}(p, h)$.
4. Apply again a rule-of-thumb bandwidth to create a subsample for the estimation of the second derivatives, which are calculated using a second degree local polynomial regression on both sides of the discontinuity. Here, although I keep the rule-of-thumb bandwidth originally proposed, I add a cap at of 1.65 in order to avoid using the extreme points in the estimation of the second derivatives, i.e. 5% of the sample on each side. Estimate the second derivatives

for both sides.

5. Plug-in the parameters calculated above and the hessian matrix M^j, where $j = (0,1)$ represents the treatment status, in the following formula for the MSE:

$$MSE = (Bias_1 + Bias_2 + Bias_3)^2 + 2 * Variance \tag{B.1}$$

where,

$$\begin{aligned}
Bias_1 &= (M^0_{12} - M^1_{12})h_p h_h \sigma_p \sigma_h C_2 \\
Bias_2 &= (M^0_{11} - M^1_{11})h_p^2 \sigma_p^2 C_3 \\
Bias_2 &= (M^0_{22} - M^1_{22})h_h^2 \sigma_h^2 C_4 \\
Variance &= \frac{v(p,h)}{n h_p h_h \sigma_p \sigma_h f(p,h)} C_1
\end{aligned}$$

The constants (C_1, C_2, C_3, C_4) are specific to the kernel and the region of the frontier used for the MSE estimation. The horizontal frontier is defined as the region where population varies between 0 and 30 (thousand) and HDI=0.7. The vertical frontier has population = 30 and HDI in the range 0.5-0.7. The values of (C_1, C_2, C_3, C_4) are shown in the table below.

	C1	C2	C3	C4
	(Away with Pop=30,Origin, Away with HDI =0.7)			
Edge	(3.20,11.91,3.20)	(0.00,-0.06,0.00)	(0.08,-0.05,-0.05)	(-0.05,-0.05,0.08)
Epanechnikov	(2.70,9.80,2.70)	(0.00,-0.07,0.00)	(0.10,-0.06,-0.06)	(-0.06,-0.06,0.10)
Normal	(2.06,7.28,2.06)	(0.00,-0.11,0.00)	(0.15,-0.08,-0.08)	(-0.08,-0.08,0.15)
Uniform	(2.00,7.00,2.00)	(0.00,-0.13,0.00)	(-0.17,-0.08,-0.08)	(-0.08,-0.08,-0.17)

The term $Bias_1$ goes to zero when the equation is estimated away from the origin. The expressions above are an expansion of the components of bias and variance used in the theoretical MSE expression defined in Ruppert and Wand (1994). They are reproduced below with the notation from this paper.

Conditional bias:

$\mathbb{E}[\hat{m}(p,h) - m(p,h) \mid (P,H)] = \frac{e_1' N_{p,h}^{-1}}{2} \int_{D_{p,h,H}} w' k(u) u' H^{\frac{1}{2}} M(p,h) H^{\frac{1}{2}} u du + Op(tr(H))$

Conditional variance:

$\mathbb{V}[\hat{m}(p,h) \mid (P,H)] = [n^{-1}|H|^{-\frac{1}{2}} e_1' N_{p,h}^{-1} T_{p,h} N_{p,h}^{-1} e_1 / f(x)] * v(p,h) * (1 + o_p(1))$,

where

$$N_{p,h} = \int_{D_{p,h,H}} w' w K(u) du$$
$$T_{p,h} = \int_{D_{p,h,H}} w' w K(u)^2 du$$
$$w = [1 \ u']$$

As for notation, $H^{\frac{1}{2}}$ is a bandwidth matrix assumed to be diagonal as $diag([\, h_p \sigma_p \ h_i \sigma_i \,])$,

M^0 and M^1 are the hessian matrices for the second degree polynomial estimated using pilot bandwidths for the non-treated and treated subsamples, respectively. $K(u)$ is the kernel,

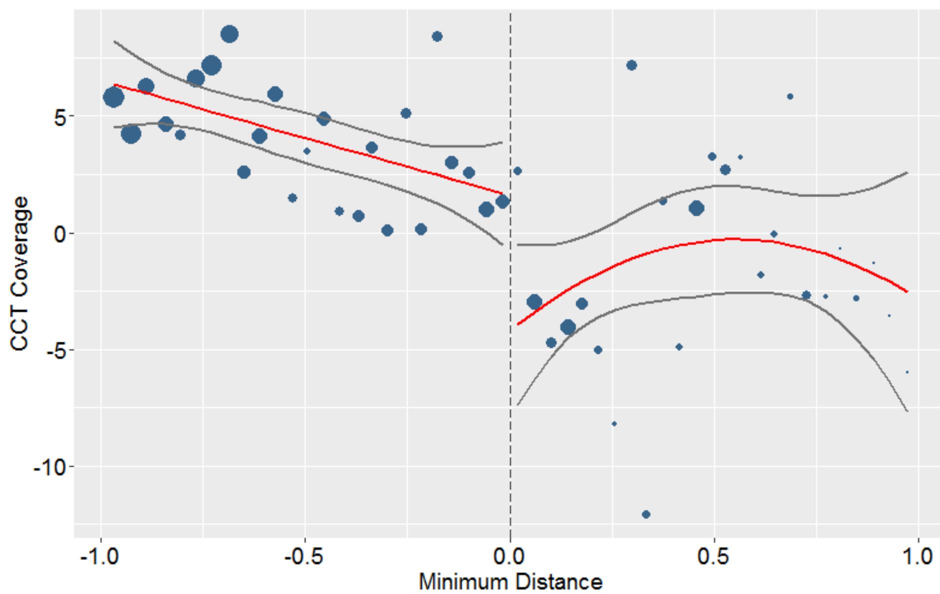
$u = [u1, u2]'$,e is defined as a vector of the same length as w, with 1 as the first element and 0 in all other elements. D^1_{x,H_1} and D^0_{x,H_0} are the sets of treatment and control points, respectively; within a bandwidth from x and within the support of the kernel K.

6. Find the pair (h_p, h_h) that minimizes the MSE expression, constraining the maximum bandwidth to a cap. I will use the cap of 1.0 for the entire sample, which is the minimum bandwidth found for the instrument using the original algorithm and one unique bandwidth for the two variables under the edge and epanechnikov kernels.

7. The steps 1-6 are repeated for 5 points away from the origin on both the vertical and horizontal frontiers. Where HDI = 0.7, I use Pop=(5-15) in 2.5 intervals and where Pop=30, I use HDI=(0.575-0.625) in 0.0125 intervals. Pick the minimum of (h_p, h_h) for each frontier and use as a starting point (first bin on each side). Calculate the (h_p, h_h) for the origin and linearly interpolate for all the k points used to estimate the CATE between the extremes and the origin. For example, for incumbency advantadge under the edge kernel, the minimum bandwidth in the horizontal dimension is $(h_p, h_h) = (0.68, 0.87)$, the bandwidth at the origin was $(h_p, h_h) = (1.00, 1.00)$ and the minimum bandwidth in the vertical dimension was $(h_p, h_h) = (1.00, 0.79)$.

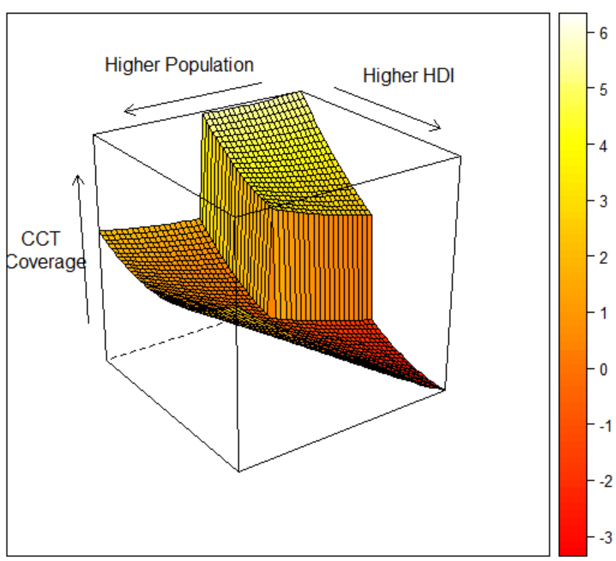
8. Steps 1-7 are repeated for each kernel.

Figure B.1: CCT Coverage: One-dimension RDD



The vertical line represents the treatment frontier (Pop = 30,000 and HDI = 0.7). Blue dots are the average CCT coverage for each one of the 25 bins on each side of the discontinuity. The x-axis shows the minimum distance of each observation to the treatment frontier. The red lines are fitted by local linear regression on the unbined data. The grey lines are the 90% confidence level.

Figure B.2: CCT Coverage: Three Dimensional Parametric Fit



The surface is fitted with a quadratic polynomial on normalized population and HDI. The sample includes all observations within 2 distance units from treatment.

Figure B.3: Conditional ATE for the log of Total Doctor's Visits

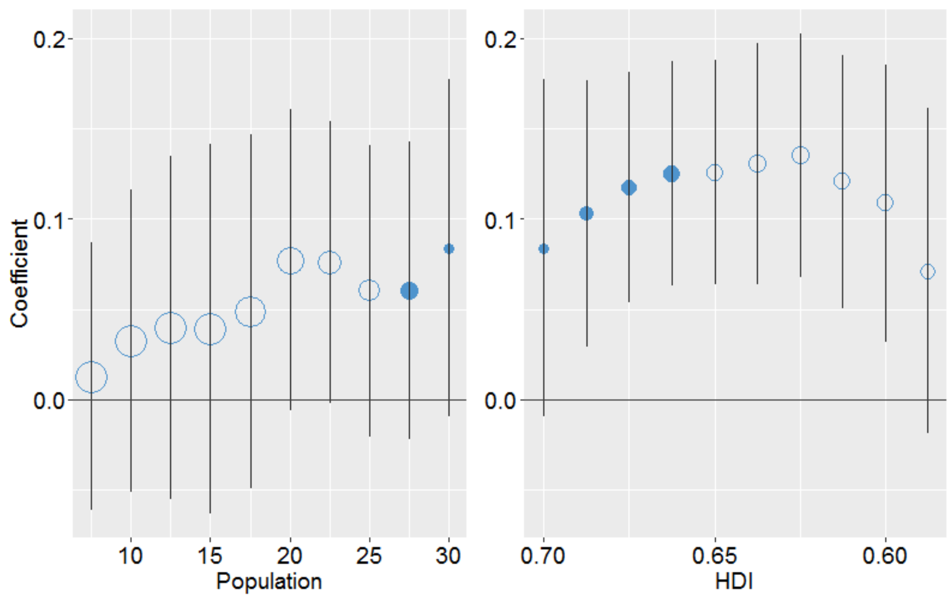
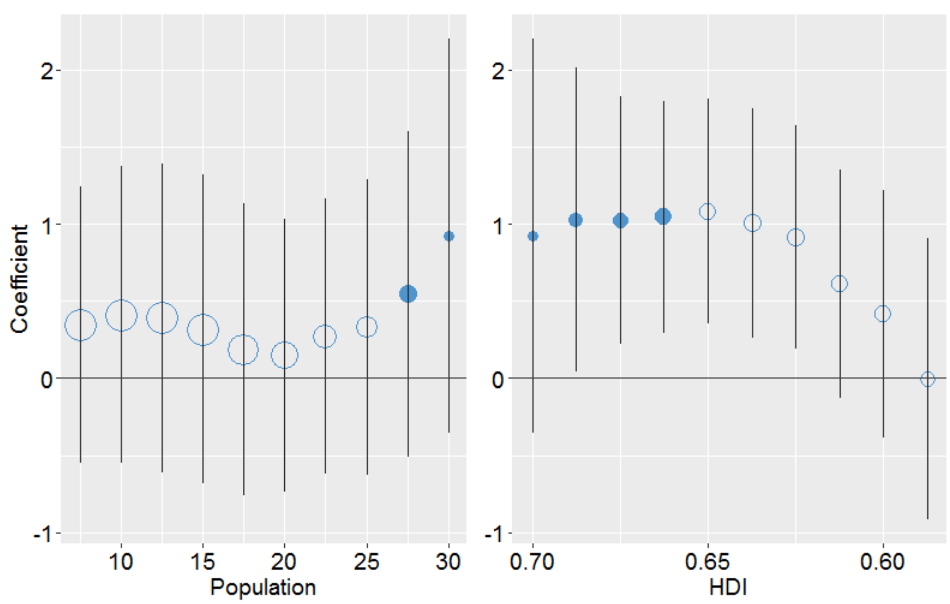


Figure B.4: Conditional ATE for the Total Doctor's Visits per Family



The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI=0.7 and population in 7,500-30,000. The right side has population=30,000 and HDI in 0.7-0.5875. Regressions use the edge kernel, year and state effects, and controls. Bandwidth is set at 0.9 standard deviations for both scores. The size of the dots are the number of observations in each bin. I repeat the bin located at the origin in both sides.

Table B.1: Optimal Bandwidths

Kernel	Edge	Epanechnikov	Normal	Uniform
Basic Health Funds	(0.71,0.99)	(0.64,0.97)	(0.53,0.94)	(0.51,0.90)
CCT Coverage	(1.00,1.00)	(1.00,0.98)	(1.00,0.94)	(1.00,0.92)
Incumbent's Vote Share	(0.99,0.96)	(0.99,0.93)	(0.99,0.90)	(0.99,0.89)
Number of candidates	(1.00,0.98)	(0.99,0.96)	(0.99,0.93)	(0.99,0.92)
Margin of Victory	(1.00,1.00)	(1.00,1.00)	(1.00,0.99)	(1.00,0.99)
Campaign Donations (log)	(1.00,1.00)	(0.99,1.00)	(0.86,0.96)	(0.82,0.95)
Challenger's Education	(0.75,0.75)	(0.75,0.75)	(0.75,0.75)	(0.75,0.75)
Challenger's Party	(1.00,1.00)	(1.00,1.00)	(1.00,0.99)	(1.00,0.98)
Turnout	(1.00,1.00)	(0.99,1.00)	(0.86,0.98)	(0.81,0.97)
Total Budget (log)	(0.94,1.00)	(0.85,0.99)	(0.71,0.96)	(0.68,0.95)
Capital Investment	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)
Personnel Spending	(1.00,1.00)	(1.00,1.00)	(1.00,0.96)	(0.99,0.95)
Other Spending	(1.00,1.00)	(1.00,0.99)	(0.99,0.95)	(0.99,0.94)
Education	(1.00,1.00)	(0.99,1.00)	(0.98,1.00)	(0.97,1.00)
Health	(1.00,0.96)	(1.00,0.94)	(0.99,0.92)	(0.99,0.91)
Administration	(1.00,1.00)	(1.00,1.00)	(0.98,0.90)	(0.98,0.84)
Urbanization	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)
Social Security	(0.98,1.00)	(0.96,1.00)	(0.93,0.96)	(0.93,0.95)
Transportation	(0.99,0.98)	(0.99,0.97)	(0.99,0.93)	(0.99,0.93)

Average optimal bandwidths calculated for the preferred frontier segment. They are expressed as (Pop.,HDI).

Table B.2: Main Results: Other Political Outcomes

	Coefficient, [90% CI], {Obs per bin}				
	P.T. Mean	RDD	OLS	IV	RDD
	(1)	(2)	(3)	(4)	(5)
Period	*2008-12*	*2008-12*	*2008-12*	*2008-12*	*2000-04*
ELECTIONS					
Turnout	84.55	0.44	0.03	0.03	-0.06
(%)	[83.60,85.41]	[-1.13,2.03]	[-0.01,0.06]	[-0.17,0.27]	[-1.52,1.37]
		{468}	{468}	{468}	{494}
BUDGET SHARES (by type)					
Total Budget[a]	149.14	-0.04	0.00	0.00	0.08
(R$ million)	[139.80,161.01]	[-0.13,0.03]	[0.00,0.00]	[-0.01,0.00]	[-0.02,0.20]
		{325}	{325}	{325}	{438}
Other Spending	42.75	-1.15	0.02	-0.10	-0.81
(% share)	[41.59,43.91]	[-3.05,0.82]	[-0.02,0.05]	[-0.34,0.05]	[-3.08,1.35]
		{351}	{351}	{351}	{477}
BUDGET SHARES (by function)					
Administration	14.36	-1.15	0.01	-0.13	0.30
(% share)	[13.42,15.40]	[-3.16,0.96]	[-0.03,0.04]	[-0.67,0.25]	[-1.94,2.83]
		{353}	{353}	{353}	{454}
Urbanization	9.75	-0.12	-0.01	-0.03	2.05*
(% share)	[9.10,10.40]	[-1.76,1.24]	[-0.03,0.02]	[-0.50,0.19]	[0.13,4.48]
		{353}	{353}	{353}	{454}
Social Security	6.25	-0.06	-0.01	0.00	-0.45
(% share)	[5.59,7.10]	[-1.50,1.12]	[-0.02,0.01]	[-0.23,0.18]	[-1.63,0.60]
		{347}	{347}	{347}	{447}
Transportation	2.71	-0.79**	-0.01	-0.10	-1.09**
(% share)	[2.28,3.21]	[-1.34,-0.19]	[-0.02,0.00]	[-0.29,0.05]	[-2.03,-0.19]
		{344}	{344}	{344}	{441}

[a]Estimated in log(Variable). [b]Coefficient for the period 2000-04 contains data from 2004 only. Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. The coefficients represent the ATE for the preferred frontier segment, under the edge kernel. All regressions include year and state effects. The list of included municipal-level controls is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Table B.3: Political Outcomes at Different FPM Population Thresholds

	Pop. nod = 23,772		Pop. nod = 37,356	
	Coeff., [90% CI]	{Obs. per bin}	Coeff., [90% CI]	{Obs. per bin}
Basic Health Funds[a] (R$ million)	-0.07* [-0.14,0.00]	{1053}	0.02 [-0.10,0.14]	{354}
CCT Coverage (% over target)	-3.00 [-7.00,0.77]	{1053}	-4.81 [-10.08,1.61]	{354}
Incumbent's Vote Share (%)	2.08 [-1.90,5.99]	{804}	0.36 [-7.94,6.51]	{271}
Number of candidates (number)	-0.16 [-0.34,0.01]	{804}	0.07 [-0.19,0.33]	{271}
Margin of Victory (pct points)	4.18* [0.44,8.29]	{804}	-0.51 [-12.94,8.81]	{271}
Campaign Donations[a] (R$'000)	0.44* [0.03,0.91]	{761}	-0.26 [-0.88,0.38]	{260}
Challenger's Education (% with high school)	-0.01 [-0.12,0.10]	{637}	10.56 [-7.30,31.92]	{219}
Challenger's Party (% clientelistic)	0.10 [-0.06,0.25]	{637}	-5.70 [-37.88,28.66]	{219}
Capital Investment (% share)	0.10 [-1.18,1.28]	{591}	0.97 [-1.97,4.27]	{207}
Personnel Spending (% share)	0.48 [-1.72,2.65]	{591}	1.05 [-3.37,5.46]	{207}
Education (% share)	1.22 [-0.37,2.72]	{591}	-1.72 [-4.90,0.80]	{203}
Health (% share)	-0.91 [-2.18,0.24]	{591}	1.64 [-2.08,6.62]	{203}

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. In order to isolate the effect of the FPM, treatment is attributed to municipalities only on the basis of population, so in this case municipalities with population below the cuttoff and HDI above 0.7 are considered as treated. Standard errors are clustered by municipality. The coefficients represent the ATE for the frontier with IDH between 0.65-0.70, and population as indicated in the table header. Bandwidths are set as one standard deviations of population and HDI. Regressions include year and state effects, and the municipal-level controls listed in the text.

Table B.4: Change in the Political Impact of the FHP post-2004

	Does it have FHP? 1=Yes		FHP in 2008-12		
	(1A)	(1B)	(2A)	(2B)	{Obs.}
ELECTIONS					
Incumbent's Vote Share	2.90***	2.76***	-2.44*	-2.55**	{7675}
(%)	[0.53]	[0.51]	[1.29]	[1.25]	
Number of candidates	-0.05**	-0.06**	0.05	0.06	{7675}
(number)	[0.02]	[0.02]	[0.05]	[0.05]	
Margin of Victory	0.67	0.85	-0.12	-0.78	{7675}
(pct points)	[0.55]	[0.53]	[1.31]	[1.24]	
Challenger's Education	1.01	-0.02	8.75*	5.07	{5877}
(% with high school)	[1.94]	[1.88]	[4.61]	[4.43]	
Challenger's Party	-0.50	-0.29	-2.78	-3.58	{5879}
(% clientelistic)	[2.01]	[1.90]	[4.82]	[4.59]	
BUDGET SHARES (by type)					
Capital Investment	0.06	-0.05	-0.77**	-0.78**	{7606}
(% share)	[0.19]	[0.19]	[0.38]	[0.38]	
Personnel Spending	-0.71**	-0.58**	2.89***	2.62***	{7606}
(% share)	[0.25]	[0.25]	[0.45]	[0.47]	
Education	-0.23	-0.22	-0.30	-0.41	{7382}
(% share)	[0.17]	[0.17]	[0.39]	[0.41]	
Health	1.22***	1.27***	1.02**	0.85**	{7382}
(% share)	[0.16]	[0.17]	[0.31]	[0.30]	

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. The first two columns estimate the effect on political outcomes of having the FHP present in the municipality in 2000. The following two columns (2A and 2B) estimate this effect is 2008-12. Columns (1A) and (2A) include dummies for micro-regions (10 municipalities on average) and columns (1B) and (2B) dummies for macro-regions (40 municipalities on average). The sample includes only smaller municipalities (below 60 thousand) and is not limited by poverty levels is order to allow enough variation in the FHP presence is 2008-12. All regressions include a dummy for 2004 and control for latitude, longitude, their interaction and 2000 HDI of the municipality. The regression for the vote shares of the incumbent also includes the past vote share of the candidate as a control.

Table B.5: Robustness of the Exclusion Restriction

	Coefficient, [90% CI], {Obs. per bin}			
	Segment B (Strong IV)		Segment A (Weak IV)	
Controls	*Yes*	*No*	*Yes*	*No*
CCT Coverage (% over target)	7.90*** [3.84,11.60]	6.97** [2.15,11.38]	4.14** [0.76,7.54]	3.85* [0.17,7.39]
Incumbent's Vote Share (%)	-7.73*** [-12.42,-3.60]	-7.60*** [-11.95,-3.03]	0.78 [-2.30,4.27]	1.02 [-1.99,4.58]
Number of candidates (number)	0.39*** [0.20,0.63]	0.35*** [0.14,0.59]	0.00 [-0.12,0.12]	0.00 [-0.12,0.12]
Margin of Victory (pct points)	-5.96** [-11.69,-1.74]	-5.69* [-11.09,-0.91]	-2.66 [-6.68,1.23]	-2.58 [-6.39,1.29]
Campaign Donations[a] (R$'000)	-0.41* [-0.89,-0.06]	-0.40* [-0.89,-0.04]	-0.04 [-0.36,0.29]	-0.05 [-0.38,0.29]
Challenger's Education (% with high school)	19.27** [4.56,33.57]	17.65** [4.33,30.17]	2.51 [-9.19,14.28]	1.74 [-10.00,13.73]
Challenger's Party (% clientelistic)	-19.30* [-33.85,-3.42]	-20.19* [-35.42,-3.49]	-1.83 [-14.91,9.73]	-2.37 [-15.57,9.43]
Capital Investment (% share)	-1.73* [-3.34,-0.22]	-1.43 [-3.01,0.02]	-0.79 [-1.90,0.16]	-0.60 [-1.68,0.41]
Personnel Spending (% share)	2.85** [0.66,5.20]	2.73** [0.60,5.19]	0.36 [-1.00,1.74]	-0.03 [-1.44,1.40]
Education (% share)	1.84** [0.34,3.91]	0.36 [-1.64,2.56]	0.03 [-1.12,1.32]	-0.41 [-1.69,0.91]
Health (% share)	2.17*** [1.05,3.73]	2.61*** [1.21,4.28]	1.20** [0.34,2.10]	1.24** [0.27,2.26]

[a]Coefficient estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Segment A: 7,500≤Pop≤17,500 , HDI = 0.7. Segment B: 0.65≤HDI≤0.70, Pop = 30,000. Standard errors are clustered by municipality. The average effect is estimated using the edge kernel for 6 bins. Regressions include year and state effects. The municipal controls are listed in the text.

Table B.6: Health Outcomes

	PT Mean [90% CI]	Coefficient, [90% CI], {Obs. per bin} 2008-12	2000-04
Number of Visits[a]	86.54	0.11***	0.04
('000 per year, 4-year avg.)	[83.39,89.50]	[0.04,0.17]	[-0.10,0.20]
		{476}	{542}
Children below 2y[a]	0.38	0.10**	0.07
('000 in any given month)	[0.36,0.39]	[0.02,0.17]	[-0.06,0.21]
		{476}	{542}
Number of Babies Born[a]	0.11	0.10*	0.10
('000 per year, 4-year avg.)	[0.10,0.11]	[0.01,0.19]	[-0.02,0.23]
		{476}	{542}
Visits per Family	11.41	1.06**	-0.54
(per year)	[11.07,11.75]	[0.22,1.87]	[-1.74,0.49]
		{476}	{542}
Mortality Rate	10.79	1.30	1.65
(children less than 11m)	[9.59,12.18]	[-0.59,3.20]	[-1.66,4.97]
		{470}	{529}
Pre-Natals	66.27	0.49	3.20
(% of pregnancies)	[63.20,69.26]	[-5.01,6.19]	[-3.12,10.17]
		{423}	{408}
Children $<$ 2y Vaccinated	95.71	0.03	1.46
(% of children)	[94.93,96.32]	[-0.87,1.01]	[-1.04,3.75]
		{476}	{542}

[a]Estimated in log (Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are clustered by municipality. All regressions include year and state effects and are estimated using the edge kernel and a bandwidth of 0.90, set as before. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment, in 2008-12.

Table B.7: Hiring Outcomes

	PT Mean [90% CI]	Coefficient [90% CI]	{Ob. per bin}
Total Employment[a] ('000)	1.19 [1.12,1.27]	-0.04 [-0.14,0.05]	{475}
Share of Political Employment[a] (%)	9.31 [8.25,10.57]	-0.05 [-0.28,0.17]	{476}
Share of Temporary Employment[a] (%)	19.26 [15.23,23.19]	0.17 [-0.15,0.56]	{431}
Total Employment (chg.) (pp)	28.06 [22.63,38.47]	0.66 [-10.29,9.38]	{474}
Share of Political Employment (chg.) (pp)	16.68 [2.79,64.59]	-5.65 [-24.24,2.27]	{471}
Share of Temporary Employment (chg.) (pp)	2.11 [-1.37,4.96]	8.19*** [4.09,13.63]	{464}
Share of Less Educated Employees (% of total employment)	28.06 [25.44,30.80]	-4.27 [-9.55,0.92]	{421}
Share of Less Educated Employees (% of temporary employment)	30.81 [25.28,42.88]	-17.20*** [-34.89,-7.30]	{421}
Formal hiring process in 2007-11 (1=Yes)	0.76 [0.66,0.83]	-0.06 [-0.23,0.11]	{476}

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. The coefficients represent the ATE for the preferred frontier segment. Standard errors are clustered by municipality. All regressions include year and state effects and are estimated using the edge kernel and a bandwidth of 0.9 standard deviations of population and HDI. The list of municipal-level controls are in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Appendix C

Appendix to Chapter 4

C.1 Income Reporting, Cancellation and Distribution of BF Benefits

There are two main reasons why the actual distribution of benefits in 2012 will not immediately reflect the fraud in that same year. First, 10% of all CadUnico households entered the system for the first time in 2012. This is 22% of all updates in the election year. Local offices can manipulate income reporting, but not the timing or approval of the actual benefit, which in many cases takes several months. Second, BF has a permanence rule since 2009, stating that households that increased their reported income above the threshold are still eligible to keep their former benefit for a grace period that could last up to two years (i.e. until their next scheduled update).

Accordingly, Table 3.1 below shows that 26% of the CadUnico households should be receiving BF benefits (yellow cells), but are not (mostly due to recent enrollment/updates), and 11% of households are receiving benefits while they should not, mostly due to the permanence rule.

Table C.1: Bolsa Família Targeting

% of Households	Reported p.c. Monthly Income, R$ [0,70]	(70,140]	(140,311]	Total
Full Benefits	43.3	7.6	3.5	54.3
Variable Benefit Only	0.3	4.7	0.9	6.0
No Benefit	14.4	11.0	14.3	39.7
Total	58.0	23.4	18.6	100.0

The cells contain the percentage of households in each category. The yellow cells on the left side represent households that are receiving less benefits than they should. The blue cells (right side) are the households that are receiving more benefits than they should.

Table C.2: Balance of Fixed and Pre-determined Variables

	PT Mean	Coefficient, [S.E.]				Opt. band.
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*{Obs.}*
Latitude	-7.962	-0.565	-0.385	-0.298	-0.307	18.69
(degrees)	[0.188]	[1.002]	[0.904]	[0.824]	[0.831]	{688}
Longitude	-41.986	-1.157	-0.821	-0.581	-1.471	11.45
(degrees)	[0.247]	[1.156]	[1.067]	[0.988]	[1.222]	{498}
Area[a]	0.701	0.325	0.283	0.245	0.326	12.97
(km2)	[0.001]	[0.245]	[0.225]	[0.207]	[0.245]	{544}
Population, 2008[a]	13.230	-0.066	-0.023	-0.019	-0.020	18.34
(thousand)	[1.033]	[0.154]	[0.142]	[0.131]	[0.133]	{679}
GDP, avg 2001-08[a]	4.175	0.116	0.096	0.078	0.105	14.81
(pc R$ '000)	[1.016]	[0.072]	[0.066]	[0.061]	[0.068]	{585}
Male, Census 2000	50.854	0.402	0.357	0.320	0.418*	12.47
(share of pop)	[0.055]	[0.248]	[0.230]	[0.214]	[0.252]	{531}
Urban, Census 2000	45.228	-0.226	-0.644	-1.099	-0.727	16.38
(share of pop)	[0.706]	[3.434]	[3.125]	[2.862]	[3.087]	{629}
Inequality, Census 2000	0.572	-0.016	-0.015	-0.014	-0.016	14.8
(Gini)	[0.003]	[0.014]	[0.012]	[0.011]	[0.013]	{585}
Poverty, MDS 2006	55.282	1.236	1.267	1.414	1.393	18.62
(% of BF eligible)	[0.293]	[1.326]	[1.226]	[1.131]	[1.142]	{684}
Poverty, MDS 2006	76.183	0.897	0.936	1.073	1.020	17.72
(% of CadUnico eligible)	[0.245]	[1.103]	[1.018]	[0.939]	[0.971]	{667}
Observations		544	617	698		

[a]Coefficient estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Table C.2: Balance of Fixed and Pre-determined Variables (continued)

	P.T. Mean	Coefficient, [S.E.]				Opt. band.
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*{Obs.}*
BF Coverage, 12/2008	94.771	0.065	0.284	0.424	0.533	22.19
(% of BF Target 06)	[0.680]	[3.275]	[3.016]	[2.781]	[2.612]	{742}
BF Expenses, 12/2008[a]	154.879	-0.050	-0.002	0.009	0.009	19.33
(R$ '000/month)	[0.001]	[0.157]	[0.144]	[0.133]	[0.132]	{704}
ri $\leq$ R$70, Census 2010	30.250	1.522	1.688	1.806	1.557	14.83
(% of CadUnico Eligible)	[0.346]	[1.701]	[1.551]	[1.421]	[1.604]	{586}
ri $\leq$ R$70, Census 2010	67.831	1.418	1.554	1.600	1.434	14.19
(% of BF Eligible)	[0.309]	[1.491]	[1.357]	[1.244]	[1.431]	{573}
ri $\leq$ R$140, Census 2010	63.447	0.444	0.673	0.868	0.624	15.63
(% of CadUnico Eligible)	[0.285]	[1.392]	[1.272]	[1.168]	[1.285]	{602}
Mayor's Education	0.410	0.066	0.060	0.050	0.064	14.07
(share with college)	[0.020]	[0.093]	[0.085]	[0.078]	[0.090]	{570}
Mayor's Gender	0.154	0.100	0.093	0.089	0.092	16.17
(share of female)	[0.015]	[0.075]	[0.068]	[0.062]	[0.067]	{624}
Mayor's Age	47.801	-3.885**	-3.487**	-3.381**	-3.533**	15.22
(years)	[0.391]	[1.907]	[1.736]	[1.588]	[1.776]	{593}
Mayor's Party	0.076	-0.051	-0.038	-0.029	-0.052	12.71
(share of PT)	[0.011]	[0.048]	[0.045]	[0.041]	[0.048]	{538}
Mayor's Party	0.390	-0.047	-0.032	-0.017	-0.036	15.27
(share of PT coalition)	[0.020]	[0.089]	[0.082]	[0.076]	[0.084]	{595}
Observations		544	617	698		

[a]Coefficient estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Table C.2: Balance of Fixed and Pre-determined Variables (continued)

	P.T. Mean	Coefficient, [S.E.]				Opt. band.
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*{Obs.}*
Mayor's Party	0.148	0.070	0.063	0.067	0.068	13.51
(share of state govt party)	[0.014]	[0.070]	[0.064]	[0.059]	[0.069]	{557}
Mayor's Party	0.284	-0.068	-0.046	-0.040	-0.068	13.15
(share of leftist)	[0.018]	[0.082]	[0.076]	[0.070]	[0.082]	{548}
Campaign Donations	53.712	-0.009	-0.003	0.001	0.004	22.37
(R$ '000 in 2008)	[0.001]	[0.245]	[0.227]	[0.211]	[0.200]	{737}
Occupation: Business	0.206	-0.117	-0.081	-0.067	-0.109	13.65
(share of business owners)	[0.016]	[0.072]	[0.066]	[0.061]	[0.071]	{559}
Occupation: Agriculture	0.166	-0.078	-0.074	-0.072	-0.074	17.71
(share in agriculture)	[0.015]	[0.062]	[0.058]	[0.055]	[0.056]	{667}
Occupation: Public Sector	0.094	0.051	0.026	0.013	0.047	13.41
(share in agriculture)	[0.012]	[0.062]	[0.055]	[0.050]	[0.060]	{555}
Avg Rain	86.596	2.472	0.849	0.102	1.057	14.97
(mm, 1970-2008)	[1.613]	[8.005]	[7.244]	[6.602]	[7.477]	{588}
Rain Volatility	1.023	0.022	0.029	0.031	0.029	21.14
(se/avg, 1970-2008)	[0.009]	[0.047]	[0.043]	[0.039]	[0.037]	{726}
BF Management Index	0.769	-0.017	-0.012	-0.009	-0.015	14.15
(IGD, 2006-2008)	[0.003]	[0.016]	[0.014]	[0.013]	[0.015]	{573}
Audits in Microregion	1.388	-0.229	-0.186	-0.207	-0.222	20.38
(number in 2006-2008)	[0.048]	[0.202]	[0.187]	[0.174]	[0.169]	{714}
Observations		544	617	698		

[a]Coefficient estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment.

Table C.3: Income Underreporting by Year of Last Update

	Coefficient, [S.E.]			
Update Year	2009	2010	2011	2012
Very Poor, ri≤R$70	1.989	0.573	3.669**	5.444***
(% of CadUnico Eligible)	[3.006]	[2.293]	[1.687]	[1.744]
Very Poor, ri≤R$70	1.457	0.995	2.884**	3.916***
(% of BF eligible)	[1.953]	[1.614]	[1.192]	[1.336]
Poor, ri≤R$140	0.633	-0.861	1.226	2.084**
(% of CadUnico Eligible)	[2.492]	[1.611]	[1.035]	[0.819]
Observations	617	617	617	617

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table C.4: Distribution of Benefits for Households Updating in 2012

	PT Mean	Coefficient, [S.E.]				Band.
	(1)	(2)	(3)	(4)	(5)	{Obs.}
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*Optimal*
Income Reporting (Income reported in the previous updated that took place in 2009-2010)						
Very Poor, ri≤R$70	85.061	0.905	1.277	1.536	1.126	15.2
(% of CadUnico Eligible)	[0.400]	[1.590]	[1.475]	[1.367]	[1.505]	{590}
Very Poor, ri≤R$70	88.889	1.060	1.186	1.276	1.207	16.2
(% of BF eligible)	[0.309]	[1.244]	[1.146]	[1.056]	[1.139]	{625}
Poor, ri≤R$140	95.439	-0.213	0.083	0.298	-0.237	12.7
(% of CadUnico Eligible)	[0.186]	[0.735]	[0.696]	[0.649]	[0.741]	{538}
Change in Income (from previous report before 2012)						
Income went above R$70	6.313	-3.201***	-2.980***	-2.831***	-3.123***	13.5
(% of CadUnico)	[0.267]	[1.042]	[0.938]	[0.861]	[1.020]	{558}
Income went below R$70	5.063	-0.420	-0.691	-0.827*	-0.722	16.4
(% of CadUnico)	[0.173]	[0.590]	[0.538]	[0.495]	[0.532]	{631}
Observations		544	617	698		

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table C.5: Heterogeneity of Results: by Mayor's Age and Poverty Rate

	P.T. Mean	Coefficient, [S.E.]				Band.
	(1)	(2)	(3)	(4)	(5)	{Obs.}
Bandwidth	*16.0*	*13.0*	*16.0*	*19.0*	*Optimal*	*Optimal*
Sample with municipalities where the elected mayors had between 40 and 55 years of age						
Very Poor, ri≤R$70	81.803	6.385***	6.250***	6.016***	6.236***	16.3
(% of CadUnico Eligible)	[0.573]	[2.165]	[2.039]	[1.920]	[2.025]	{383}
Very Poor, ri≤R$70	88.919	3.739**	3.837**	3.802***	3.834**	15.9
(% of BF eligible)	[0.422]	[1.645]	[1.542]	[1.448]	[1.547]	{372}
Poor, ri≤R$140	91.624	3.187***	2.943***	2.729***	2.891***	16.9
(% of CadUnico Eligible)	[0.273]	[1.083]	[1.013]	[0.944]	[0.992]	{394}
Mayor's Age	47.343	-1.029	-0.888	-0.855	-0.921	14.6
(years)	[0.229]	[1.150]	[1.043]	[0.947]	[1.088]	{352}
Observations		332	375	422	383	
Sample with municipalities in the bottom tercile in poverty						
Very Poor, ri≤R$70	42.571	-4.610	-4.283	-3.668	-4.743	16.3
(% of CadUnico Eligible)	[0.729]	[3.585]	[3.217]	[2.973]	[3.719]	{383}
Very Poor, ri≤R$70	59.728	-2.852	-2.748	-2.172	-2.820	15.9
(% of BF eligible)	[0.668]	[3.353]	[3.015]	[2.784]	[3.473]	{372}
Poor, ri≤R$140	69.340	-3.206	-2.688	-2.379	-3.225	16.9
(% of CadUnico Eligible)	[0.521]	[2.369]	[2.136]	[1.988]	[2.376]	{394}
Observations		452	533	585	423	

Significant at: 99% ***, 95% **, 90% *. Standard errors are heteroskedasticity robust. The list of included municipal-level covariates is described in the text. Pre-treatment means correspond to the predicted values of the variables for a municipality at the discontinuity segment, before treatment. Optimal bandwidths are calculated using the CCT algorithm.

Table C.6: Balance Before and After Matching

Variables	Unweighted Mean Treat.=0	Unweighted Mean Treat.=1	Difference	Weighted Mean Treat.=0
Latitude	-8.52	-6.90	1.62***	-6.90
Longitude	-41.97	-42.79	-0.820	-42.79
Population, 2008[a]	2.53	2.60	0.080	2.60
Region, Northeast	0.80	0.87	0.07*	0.87
Region, Southeast	0.04	0.01	-0.03***	0.01
Region, South	0.02	0.00	-0.02***	0.00
Region, Midwest	0.00	0.01	0.010	0.01
Rain Volatility, 1970-2008	1.01	0.99	-0.020	0.99
Avg Rain, 1970-2008	100.85	107.90	7.040	107.90
pc GDP, 2001-08[a]	1.45	1.33	-0.11***	1.33
Pop Literacy, 2000	63.19	60.04	-3.15***	60.04
Pop Gender, 2000	50.71	50.86	0.150	50.86
Urban Pop, 2000	46.06	41.95	-4.11**	41.95
GINI, 2000	0.57	0.57	0.000	0.57
Transfers from federal govt, 2008[a]	9.12	9.17	0.050	9.17
Rate of Very Poor, 2010	67.54	70.33	2.79***	70.33
Poverty, 2006	54.70	58.29	3.59***	58.29
BF Coverage, Dec 08	93.94	94.98	1.040	94.98
CadUnico Coverage, Dec 08	103.37	101.53	-1.840	101.53
Total BF Spending, Dec 08[a]	11.87	12.02	0.15**	12.02

[a]log(Variable). Significant at: 99% ***, 95% **, 90% *.

Table C.5: Balance Before and After Matching (Continued)

Variables	Unweighted Mean Treat.=0	Unweighted Mean Treat.=1	Difference	Weighted Mean Treat.=0
IGD Index, 2006-08	0.76	0.75	-0.010	0.75
Audits in City, 2009-10	0.08	0.08	-0.010	0.08
Audits in Micro-region, 2009-10	0.86	0.86	0.000	0.86
PT's Presidential Votes, 2006	67.27	69.64	2.37*	69.64
Margin of Victory, 2008	6.60	6.25	-0.350	6.25
Gender	0.13	0.17	0.040	0.17
Education, College	0.42	0.38	-0.040	0.38
Age	45.08	46.73	1.66*	46.73
Campaign Donations, 2008[a]	10.88	10.89	0.010	10.89
Assets, 2008[a]	5.03	4.99	-0.050	4.99
Mayor's Votes, 2008	50.31	49.91	-0.400	49.91
Occupation, Public Sector	0.12	0.13	0.010	0.13
Occupation, Business	0.24	0.16	-0.08**	0.16
Occupation, Agriculture	0.13	0.19	0.060	0.19
Political Experience (Mayor)	0.13	0.11	-0.030	0.11
Political Experience (Legislator)	0.13	0.20	0.07*	0.20
PT	0.11	0.08	-0.020	0.08
PSDB	0.13	0.10	-0.030	0.10
PT's Local Coalition	0.43	0.36	-0.070	0.36
Leftist Party	0.31	0.28	-0.030	0.28

[a]log(Variable). Significant at: 99% ***, 95% **, 90% *.