CASH TRANSFERS, CLIENTELISM, AND POLITICAL ENFRANCHISEMENT: EVIDENCE FROM BRAZIL*

Anderson Frey†

Abstract

This paper uses Brazil's *Bolsa Família* to show that poverty alleviation policies that are shielded from the influence of political intermediaries can reduce incumbency advantage, increase both electoral competition and candidate quality, reduce support for clientelistic parties, and lead incumbents to increase redistributive spending. The paper exploits a nonparametric multivariate regression discontinuity design and employs a novel identification strategy for the variation in program coverage. The theory proposes that cash transfers, by reducing the vulnerability of poor voters, make incumbents lose their ability to replace public good distribution with vote buying, pushing them to channel effort away from clientelism.

JEL Classification: C14,C36,D72,P16,O12,O54

Political leaders in many developing democracies are able to reap electoral rewards by replacing public good distribution with private transfers targeted to groups or individual voters. This practice often takes the form of clientelism and vote buying.[1] Clientelism thrives in poor environments,[2] it is associated with low political competition, perpetuation of political machines, and inefficient public good provision (see reviews on Boix and Stokes (2009) and Hicken (2011)). Shielding redistributive policies from the influence of clientelistic networks is one of the main challenges for effective policy decentralization in the developing world (Bardhan and Mookherjee, 2005).

*I thank Francesco Trebbi, Siwan Anderson, Patrick François and Thomas Lemieux for the extensive discussions and suggestions. I also thank the participants at various seminars at VSE, Vanderbilt, Rochester, York, Ottawa and Harris (Chicago) for the comments and suggestions. All errors are my own.

†Department of Political Science, University of Rochester. Harkness Hall, 320B. Rochester, NY, 14627. email: anderson.frey@rochester.edu.

Despite the established relationship between clientelism and poverty, the causal channels linking redistributive policies to a potential reduction in clientelism remain vastly unexplored.[3] Understanding this mechanism, however, is key for effective policy design. This paper is concerned with the impact of income increases generated by such policies, more specifically cash transfers, in local public good distribution and leadership selection. Will voters become less likely to engage in clientelism as their income increases? Will voters change the way in which they select politicians? Can poverty alleviation be a tool for institutional change?

The paper examines these questions using the largest Conditional cash transfers (CCT) program in the world, Brazil's *Bolsa Família* (BF). CCT programs are one of the main poverty alleviation policies currently adopted by developing nations. For the most part, these programs are designed to provide a stable income for the extremely poor, and to prevent capture by political intermediaries.[4] To the extent that their implementation strategy is successful in a clientelistic context, CCT programs can prompt significant changes in local political structures.

Brazilian politics are characterized by clientelism and pork barrel spending across all levels (Alston and Mueller, 2006; Fried, 2012). Vote buying is part of the culture, and more than 450 mayors were impeached due to the practice between 2000 and 2009. Using a survey in Northeastern municipalities, Sugiyama and Hunter (2013) show that 66% of the survey respondents were aware of *quid pro quo* offers for votes, and 28% reported having received recent offers.[5]

Using the Brazilian CCT experience, this paper shows evidence that transfers shielded from the influence of local politicians improve the electoral process and contribute to the political enfranchisement of the poor. Higher coverage reduces the incumbency advantage for mayors, increases both electoral competition and the quality of candidates, weakens support for parties historically associated with clientelism, and prompts an increase in health care and education spending.[6] These effects have been observed for up to nine years after the program creation.

In order to show how these results are reconciled in a clientelistic environment, I propose a probabilistic voting model where the incumbent chooses to allocate effort to either implementation of pro-poor public goods or clientelism. Clientelistic exchanges become possible when a share of the already poor population becomes extremely vulnerable due to negative income shocks. Uncertainty about the willingness of challengers to engage in such practices is what sustains incumbency advantage. In this context, the CCT program provides a permanent income increase that reduces the probability of the poor becoming extremely vulnerable, therefore reducing incumbency advantage. As vote buying becomes less attractive, incumbents try to recover the lost votes by funneling effort back into public goods.

BF has been aggressively marketed as a federal policy, thereby limiting the ability of local politicians to claim credit for its implementation. There are, however, two main challenges to identifying the effects of this program on politics: the implementation was not randomized, and the program is jointly run by the central and municipal governments. This gives mayors some leeway to manipulate the selection of beneficiaries. Accordingly, this paper estimates the political effects of BF using a novel identification strategy. It relies on the cross-municipal variation in CCT coverage that was generated by a small discontinuity in funding for the Family Health Program (FHP), a household-based health care program run by municipalities. Since August 2004, municipalities with population below 30,000, and human development index (HDI) below 0.7, are eligible to receive 50% additional FHP funding. This paper instruments local CCT coverage with this funding differential, using a multivariate regression discontinuity design (MRDD).

The identification strategy is only effective if the FHP has a strong impact on the CCT coverage (first stage), and no meaningful direct effects on political outcomes (exclusion restriction). With respect to the first stage, the FHP funding affects municipal CCT coverage through two institutional features of the BF program. First, access to information was key to the enrollment of potential beneficiaries in the early years. Where local administrations had few resources to promote enrollment, the FHP teams had sufficient capillarity and penetration among the poor to disseminate the information. A survey with 10,000 poor households in Brazil shows that more than 10% of BF beneficiaries came to know about the program through their family doctor. Second, the coverage differentials created between 2004 and 2006 persisted over the long term (2012+), due to the limits on program enrollment and to the registry update rules.

There is also strong evidence that the political effects of the FHP operate almost exclusively through the channel of CCT coverage (exclusion restriction). The first piece of evidence emerges from the use of a MRDD for the estimation. MRDDs differ from single-variable RDD cases on the basis of their implementation methods and their interpretive framework. For example, treatment effects are identified along a frontier of points on a two-variable space, rather than on a single point. By taking advantage of the first stage heterogeneity in effects along the treatment frontier, this paper shows that the effects on various political outcomes are only observed in the frontier segments where we also observe FHP effects on CCT coverage. For the most part, in subsamples where the FHP does not affect CCT coverage, the political outcomes also remain unaffected.

The second piece of evidence for the exclusion restriction comes from examining changes in the effects of the FHP on political outcomes over time. Using an alternative dataset with the arrival date and coverage of the FHP in each municipality, the paper shows that these direct effects, if they exist, change significantly after 2004 (BF started in 2003).The direction of these changes also matches the sign of coefficients estimated at the discontinuity (with the MRDD). Finally, the magnitude and direction of the estimated coefficients with the MRDD indicate that the results cannot be mechanically driven by the 50% FHP funding increase alone. The increases in provision of health services are 9 times larger then what is expected in the absence of CCT coverage.

This paper finds that the 50% additional FHP funding generates a 7.9 percentage points (pp) higher CCT coverage for municipalities having at least 25% of poor families. A 10pp increase in CCT coverage triggers a 8.3pp vote loss for the incumbent, which is mostly driven by the effect on less educated politicians or candidates that belong to clientelistic parties. Higher CCT coverage also prompts a 6.4pp fall in the margin of victory, and 0.4 more candidates in the mayoral race. Finally, the share of budget labeled pro-poor (health and education) increases by 4.8pp (nearly 10%). These results are robust to the choice of kernel and bandwidth in the nonparametric estimation, and are generated based on a data-driven plug-in algorithm for bandwidth selection in the multivariate case detailed in the appendix.

This paper also tests an alternative mechanism to clientelism, which based on party ideology. In Brazil, the line that separates programmatic and clientelistic parties are similar to the one that defines left an right-wing parties. One could argue that higher CCT coverage should generate a shift in electoral support toward leftist parties, especially since the leftist President Lula (PT, 2003-2010) is widely seen by the poor population as the main responsible for the spread of CCT benefits (Zucco, 2013). In turn, incumbents from left-wing parties would also focus more effort on pro-poor public good distribution. However, the results show that politicians (both incumbents an challengers) from leftist parties are no more likely to lose votes, or to shift effort towards pro-poor public good, than their right-wing counterparts. Heterogeneous effects are only observed for parties when they are categorized by historical association with clientelism.[7]

This paper contributes mainly to three strands of the literature in both economics and political science. First, it informs the extensive research on the relationship between political incentives and different forms of redistribution. One constant in this literature is the connection between vulnerability and the predominance of clientelistic forms of redistribution (see the seminal work of Dixit and Londregan (1996) and reviews in Hicken (2011); Boix and Stokes (2009)). The key finding in this paper is that persistent income increases can be an effective tool to reduce clientelism and to foster more programmatic forms of redistribution. Closer to this paper is the research focusing on various aspects of the mechanics of clientelism around the world. Finan and Schechter (2012) focuses on 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, and Brusco, Nazareno, and Stokes (2004) discuss the role of discount rate in vote buying in Argentina. 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.

Second, the literature on political effects of CCT is generally concerned with national politics, which is the level at which the program is implemented. A number of papers point out that electoral support for the incumbent increases following CCT implementation in Uruguay (Manacorda, Miguel, and Vigorito, 2011), Colombia (Baez et al., 2012), Mexico (De La O, 2013), and Brazil (Zucco, 2013). Labonne (2013) and Rodriguez-Chamussy (2015) examine the impact of CCT on incumbency advantage using randomized program roll-outs in the Philippines and Mexico, respectively.[8] Unlike the results presented here, 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 selected the beneficiaries and therefore could claim credit for the policy. The authors show that higher program performance was associated with higher incumbency advantage. To my knowledge, this is the first paper to provide a comprehensive picture of the way in which cash transfers affect a variety of political outcomes, in a context where the ability of mayors to claim credit for the program is limited.

Finally, by extending the nonparametric method proposed proposed by Zajonc (2012), and using a formal approach to bandwidth selection, this paper departs from the strategies commonly used for the estimation of multivariate regression discontinuity designs (MRDDs). These strategies include collapsing the multiple variables to one dimension and applying a single RDD (Jacob and Lefgren, 2004; Dell, 2010; Clark and Martorell, 2014), or estimating a parametric function over the two-variable space (Reardon and Robinson, 2010; Wong, Steiner, and Cook, 2013). To my knowledge, this is the first paper that uses the average treatment effect heterogeneity to inform an IV regression.

The remainder of the paper is structured as follows. Section 2 hast a brief overview of the institutional background; Section 3 proposes a theoretical model; Section 4 describes the identification strategy and explains how the theoretical predictions are tested; Section 5 describes the data and construction of variables; Section 6 presents the empirical strategy for the MRDD; Section 7 shows the results and their interpretation; and Section 8 concludes.

1 Institutional Background

1.1 Bolsa Família (Brazilian CCT program)

Bolsa Família (BF) is the largest CCT program in the world, covering 13 million households (Dec 2012). It was created in 2003 with the unification of other smaller CCT programs previously run by different government ministries.[9] Contrary to other CCT implementations (e.g. Mexico and Philippines), BF was rolled out simultaneously across all of Brazil. In a nutshell, households with per capita income below a certain threshold are eligible to receive a monthly government grant, which varies according to the number of children in the family. 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. Permanence in the program is subject to compliance with conditionalities, particularly school attendance health check-ups. The BF operations are run jointly by the central and local governments.[10] The former determines major guidelines,[11] controls the approval and cancellation of benefits, and pays beneficiaries directly through a debit-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 families.

1.2 Cash Transfers and Local Politics

Brazilian politics are characterized by low of ideological identification (Ames and Smith, 2010), a highly fragmented party system with candidate-driven elections, and local political coalitions that span the entire ideological spectrum. Mayors have a two-term limit and elections are held every four years under majority rule in one round.[12] Voting is mandatory and the average turnout in the 2008 and 2012 elections was 83%. Accordingly, this paper will focus on the reelection of candidates and not parties.

The implementation of most government policies in Brazil is decentralized. Programs are often jointly financed by federal, state, and municipal administrations. The majority of revenues for small municipalities come in the form of transfers from higher levels of government; and clientelism is 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. In this climate, 71% of eligible incumbents ran for reelection between 2000 and 2012, and nearly 60% were reelected.

The innovative BF program was specifically designed to reduce local political 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 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 roughly 0.5% of Brazil's GDP, the BF transfers to small municipalities represent nearly 5% of the local budget.

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.[13] Each team is responsible for a geographic area 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 to the broader health care system. Given that the majority of 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 (Fujiwara, 2015).

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 attention in municipalities is co-financed by federal, state, and local resources. Federal transfers that finance FHP teams are paid monthly as a fixed amount per team.[14] These payments were uniform across the country until August 2004, when municipalities with population below 30,000, and HDI below 0.70,[15] started to receive an extra 50% funding per team (see the timeline of events in Figure I). The HDI for eligibility was calculated based on the 2000 census, and the population was referenced using the 2003 official population estimates. Both these values and the list of eligible municipalities have not been updated since.[16]

2 Theoretical Framework

I propose a probabilistic voting model that shows how income increases for the poor affect incumbency advantage, budget allocation, and clientelism in local elections.[17] Intuitively, incumbents can allocate effort to either the implementation pro-poor public goods or targeted clientelistic exchanges. Opportunities for clientelism arise when a subset of the voting poor becomes extremely vulnerable due to a temporary negative income shock. These targeted benefits consist of either cash payments or exclusive access to goods and services redirected from public resources (e.g. medicine, food, cement). The share of poor population that becomes temporarily vulnerable always benefits from immediate aid coming from these clientelistic offers. Incumbency advantage therefore comes from the uncertainty about the willingness of a challenger to engage in such exchanges vis-àvis the incumbent, whose policy is observed.

In this context, an income increase reduces the probability of poor voters becoming extremely vulnerable and therefore susceptible to clientelism, which makes it a less attractive electoral strategy. This leads incumbents to shift effort towards public goods, and reduces incumbency advantage, given that the uncertainty regarding the clientelistic practices of a potential challenger becomes less salient to poor voters vis-à-vis other considerations as ideology and public goods implementation.

Consider a two-period model, in which the first period incumbent implements policy and faces a challenger in the election at the period's end. If elected, it becomes her last period in office due to a two-term limit. If the challenger wins, he can run for reelection at the end of the second period. The rents of office in every period are given by R. Incumbents always exercise an amount E of effort on pro-poor policy, which is allocated between public good implementation (e) or clientelistic redistribution ($1 - e$). Incumbents also incur an utility cost of determining effort allocation, given by κe.[18] They only incur this cost when they first established effort allocation, so that if reelected to a second term, they face $\kappa = 0$ if they keep e unchanged.

Politicians can differ in two dimensions: their willingness to exercise programmatic or clientelistic effort, and their ability to do so, which is measured by $\theta \in [0, 1]$. Extreme politicians (type x) never allocate effort to clientelism, with $e = E$, and $\theta = \bar{\theta}$. Opportunistic politicians allocate clientelistic effort if it enhances their reelection chances, and can be of two types: clientelistic types (type c) have a relative disadvantage in the delivery of public goods, with $\theta = \underline{\theta}$. Programmatic types (type p) have $\theta = \bar{\theta}$, where $\bar{\theta} > \underline{\theta}$. The ex-ante probabilities of each type of politician in the population are given by (μ_x, μ_c, μ_p).

The parameter σ^i denotes the idiosyncratic preference of voter i for the challenger. This parameter has a uniform distribution on $(-1/2\phi, 1/2\phi)$, with density ϕ. The challenger's relative popularity in the entire population is given by λ, with an uniform distribution on $(-1/2\psi, 1/2\psi)$. Wealthy voters represent a share $(1 - n)$ of the population, and vote only based on their idiosyncratic preference and relative popularity of candidates, choosing the incumbent when $\sigma_i + \lambda < 0$. Poor voters represent a share n of the population, and have income y. In keeping with the literature on clientelism (e.g. Bobonis et al. (2017)), opportunities for clientelistic exchanges come from idiosyncratic income shocks to poor households that lead them to become vulnerable and susceptible to vote buying. Let $\delta(y)$ be the share of poor voters subject to such shocks at every period. This is a random draw from the poor population, and increases as income falls ($\delta_y < 0$).

When voting, poor citizens also consider the prospective utility to be received under each candidate in period 2, which is quasi-linear and has two components. The utility coming from pro-poor public goods is given by the concave function $\log(e)$. The utility coming from consumption of private clientelistic transfers is simply given by $(E - e)$. Every period, a share $\delta(y)$ of poor voters is made vulnerable by income shocks and becomes a target for clientelism. Thus, the expected utility of a poor voter after the election is given by $U^P(g, c) = \theta^P \log(e) + \delta(y)(E - e)$, where P refers to the politician that wins the election, i.e., incumbent (I) or challenger (C).

The allocation of effort is not perfectly observable by poor voters, therefore they are not able to identify a programmatic from a clientelistic type. However, if positive effort is put towards vote buying ($e < E$), voters infer that the incumbent is not an extreme type. Notice that if the income y is low enough, there is always a significant part of the poor population that would find some amount of clientelism optimal, and prefer not to elect an extreme politician.[19] This is the source of incumbency advantage of opportunistic politicians in this environment. Accordingly, when voters observe $e < E$, they know that under a reelected incumbent in period 2 they expect the following: $U^I(g, c) = \theta \log(e) + \delta(y)(E - e)$, where both e is the same as it was in period 1.[20]

There is a positive probability that an elected challenger is an extreme type (μ_x), providing utility equal to $g(E)$. With $(1-\mu_x)$ probability, the challenger is opportunistic and behaves like an incumbent seeking reelection, therefore providing $\theta\log(e)+\delta(y)(E-e)$. Thus, the differential in expected utility from an incumbent and a challenger is given by $\mu_x[\theta\log(e)+\delta(y)(E-e)-\theta\log(E)]$.[21]

The timing of events is as follows: (1) an incumbent is elected for her first tenure in office, with the possibility of running for reelection at the end of period 1; (2) with probability $\delta(y)$, poor voters receive a negative shock that makes them vulnerable; (3) incumbent allocates effort between public goods and clientelistic payments, incurring cost κ; and (4) elections happen. If the incumbent wins, she enjoys the rents of office in the second period and only minimizes the allocation cost κ, which implies maintaing effort constant from period 1. If the challenger wins, he behaves like a first period incumbent. Thus, poor citizens vote for the incumbent whenever $\mu_x[\theta\log(e)+\delta(y)(E-e)-\theta\log(E)]>\sigma_i+\lambda$, in which case the incumbent's vote share is:

$$\pi=\phi n\mu_x[\theta\log(e)+\delta(y)(E-e)-\theta\log(E)]+\phi(\frac{1}{2\phi}-\lambda) \tag{1}$$

Under the model's distributional assumptions, incumbents maximize reelection probability subject to the utility cost of effort in period 1, as shown by the equation below.

$$\max_e \frac{1}{2}+\psi n\mu_x[\theta\log(e)+\delta(y)(E-e)-\theta\log(E)]-\kappa e \tag{2}$$

The comparative static of interest here is how the model equilibrium quantities respond to a permanent change in income, which in this context comes from the implementation of a CCT program. Accordingly, the main predictions of the model after an exogenous increase in y are summarized below, and all proofs for the following propositions and conditions under which they hold are shown in the appendix. In Section 5.2, I detail how each one of these propositions can be tested combining the data described in Section 4 with the empirical strategy described in Section 5.

An exogenous increase in y generates the following:

PROPOSITION 1. *Opportunistic incumbents shift effort from clientelism to pro-poor public goods.* The increase in income reduces the vulnerability of poor voters measured by $\delta(y)$, and therefore the opportunities for clientelistic exchanges. The implementation of public goods becomes a relatively more efficient electoral strategy.

PROPOSITION 2. *The vote share of the incumbent falls.* In this model, the incumbency advantage only exists because vulnerable voters desire some extent of clientelism and avoid an extreme public good type politician. As income increases, the optimal allocation of effort becomes closer to a pure public good allocation, and incumbency advantage becomes narrower for all politicians.

PROPOSITION 3. *Clientelistic-type incumbents face a larger decline in vote shares, but are* *less likely to shift effort towards public goods implementation.* Every unit of effort that is allocated to public goods generates less utility to poor citizens under clientelistic politicians when compared to programmatic types. Thus, in equilibrium they shift less their allocation. Also, because clientelistic incumbents rely relatively more on vote buying to sustain their incumbency advantage, they lose more votes when this strategy becomes less attractive.

PROPOSITION 4. *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.* All incumbency advantage comes from exploiting the vulnerability of the poor population in clientelistic exchanges. Municipalities with a large share of poor voters will see higher effects when income increases.

3 Using a Discontinuity as an Instrument

Bolsa Família 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 III).[22] 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, I do not treat this variation as exogenous.

This paper uses a discontinuity in the funding for the Family Health Program (FHP) as an instrument for CCT coverage. I argue that the allocation of FHP funds prompted a 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),[23] BF's coverage target was 11.1 million families (Figure III). 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 funds 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 (AIBF - *Avaliação de Impacto do Bolsa Família*) 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.[24] 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. Additionally, Table A.XI (appendix) shows that the increase in FHP funds contributed to increases in both the number of families visited by local doctors and the number of visits per family.

There are two main reasons that account for the persistence of early CCT coverage differentials. 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.[25] Second, although beneficiaries are required to update their information every two years under penalty of losing the benefit, this rule was not properly enforced prior to 2009, when the government created a permanence rule,[26] 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 FHP funding differentials do not have a significant direct effect on political outcomes (exclusion restriction). These direct effects cannot be tested separately from the effects happening through CCT, given that the discontinuity was introduced at the early stages of BF (Figure I). Instead, I rely on three pieces of evidence indicating that the political outcomes are predominantly generated by the CCT variation, and not by the FHP (i.e. supporting the exclusion restriction).

First, The FHP is the best rated health program in Brazil (IPEA, 2011), and it is run by the mayors. It is highly unlikely that it triggers a massive negative effect on the support for politicians that implement it. Second, similar evidence emerges from the multivariate regression discontinuity design (MRDD), which allows us to observe the heterogeneity of FHP funding effects on both CCT coverage and the political outcomes, over different values of population and HDI (details in Section 5 and Figure V, Figure VII and Figure VII). The idea is that, although the FHP funding differential is applied to the entire 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 additional FHP funding on information dissemination. If CCT coverage differentials are only observed in these municipalities, the same should happen to the political effects.

Third, using a different dataset with the date of introduction of the FHP in each municipality, I regress the political outcomes of interest on the FHP arrival date interacted with time dummies.

Even with municipality fixed effects, this regression might suffer from omitted variable bias (i.e. the arrival date of FHP is not random). However, under the benign assumption that this bias is constant over time, changes in the effects of FHP on politics over time can be identified through the interaction with time dummies. Accordingly, I observe that the FHP effects on political outcomes significantly changes after the introduction of BF, and they change in the same direction of the effects observed in the core MRDD analysis in this paper. See Section 6.3 for a more detailed discussion of this alternative test, and on the overall evidence supporting the exclusion restriction.

The identification also 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 frontier of 30,000 in this paper. Nevertheless, it seems unlikely that the FPM is driving the results here. Although the FPM policy thresholds pre-date the BF for more than a decade, no effects around the discontinuity are observed for electoral cycles pre-BF, i.e., 2000 and 2004, (see Table II). Also, when the equations are estimated using FPM discontinuities at different thresholds (right below and above 30,564), no political effects are observed (see Table A.VII). Section 6.4 discuses this evidence in more details.

Finally, I perform two alternative tests to rule out the impact of potential omitted variable biases. First, the effects are estimated for all outcome variables in two periods, pre-treatment, but under the same thresholds of HDI and population (see Table II). Second, the paper shows the balance of fixed and pre-determined municipality characteristics around the discontinuity, also before treatment (Table A.II).

4 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.[27] In addition to *Bolsa Família*, I also include households receiving the older CCT programs from the Ministry of Education (*Bolsa Escola*), the Ministry of Health (*Bolsa Alimentação*) and the MDS (*Cartão Alimentação*). However, because nearly all beneficiaries of these programs migrated to BF between 2003 and 2006, their number is not meaningful on or after 2008.

The federal government provides monthly data on basic health transfers to municipalities, including the FHP. It also makes data available with respect to coverage by health teams, and health outcomes measured within the scope of the public 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 services provided to the poor 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).[28] 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.[29] 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 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; and the number of candidates. The data also records the education level of all candidates. The main specification estimates the effects using the elections in 2008 and 2012. The 2004 election happened two months after the introduction of the discontinuity, so any effects are unlikely to be observed. The pre-treatment results (2000 and 2004) are shown as a placebo test. Political parties are classified according to their level of clientelism following the Democratic Accountability and Linkages Project (DALP).[30]

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.[31] 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 municipalities where the incumbent is actually running can be used.[32]

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. Again, in keeping with the theory, the main specification only uses municipalities with at least 25% of poor households (BF-eligible), roughly 60% of the sample. Results including low-poverty municipalities are provided in the appendix (Table A.IV).

5 Empirical Strategy

5.1 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, 2015). An extension of the RD approach is the case where the treatment eligibility is determined by two running variables, e.g., latitude and longitude (Dell, 2010; Gerber, Kessler, and Meredith, 2011; Keele and Titiunik, 2015) 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 cutoffs at 30,000 and 0.7, has the ATE defined over the frontier: $F = (p_m, h_m) : (p_m <= 30, h_m = 0.7) \cup (p_m = 30, h_m <= 0.7)$ (Figure IV).

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.

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[33] function 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.[34] 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 IV).

There are two good reasons to explore the heterogeneity of the effects in this paper. First, the two scores have a distinct nature. The sub-populations being compared along the frontier, by way of either a minimum distance or a parametric approach, might differ considerably. This would defeat the spirit of the RDD.[35] Second, the estimation of the effect of CCT on political outcomes depends on having a strong instrument for CCT coverage, which in this paper is found by teh way of exploiting the heterogeneity along the population x HI frontier. 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 segment is derived by averaging the CATEs. Thus, the treatment heterogeneity along the frontier becomes 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:

$$\tau_{Con} = \lim_{\epsilon \to 0} \mathbb{E}[y_m \mid (p_m, h_m) \in N_{\epsilon}^{+}(p, h)] \; - \; \lim_{\epsilon \to 0} \mathbb{E}[y_m \mid (p_m, h_m) \in N_{\epsilon}^{-}(p, h)] \tag{3}$$

where $N_{\epsilon}^{+}(p, h)$ 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 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.

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,[36] CCT coverage is defined as cct_{mt}. The first stage of the instrumental variable estimation is shown in equation 4 below.

$$cct_{mt} = \alpha_0 + \alpha_1 \delta_m + \alpha_2 p_m^c + \alpha_3 h_m^c + \alpha_4 \delta_m p_m^c + \alpha_5 \delta_m h_m^c + \eta_t + \gamma_s + \theta_m + \mu_{mt} \tag{4}$$

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).[37] This is usual in RDDs to reduce the sample variability (Lee and Lemieux, 2010). The local linear regression is weighted by the edge kernel in the main specification, but the robustness to this choice is shown in the results. All kernels are two-dimensional, defined as the following product: $K(u_1, u_2) = [K(u_1) \cdot K(u_2)]$.

This equation can be estimated for any point of the frontier, by centering the score variables at the desired point. The effects are estimated for a total of 19 points, limiting the data to a bandwidth defined over the two score variables.[38] 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}) is 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 5 below).

$$\hat{\tau}_{avg} = \frac{\sum_{k=1}^{K} \alpha_{1k}(p_k, h_k)\,\hat{\lambda}(p_k, h_k)}{\sum_{k=1}^{K} \hat{\lambda}(p_k, h_k)} \tag{5}$$

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), with a bootstrap sample of size 5,000.

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 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 C). 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 6 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.[39] 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 7 below.

$$out_{mt} = \beta_0 + \beta_1 \widehat{cct}_{mt} + \beta_2 p_m^c + \beta_3 h_m^c + \beta_4 \delta_m p_m^c + \beta_5 \delta_m h_m^c + \eta_t + \gamma_s + \theta_m + \epsilon_{mt} \tag{6}$$

$$\hat{\tau}_{IVavg} = \frac{\sum_{k=1}^{K} \beta_{1k}(p_k, h_k)\,\hat{\lambda}(p_k, h_k)}{\sum_{k=1}^{K} \hat{\lambda}(p_k, h_k)} \tag{7}$$

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,[40] which corresponds to the following:

$$Segment = (p_m, h_m) : (p_m >= 27.5, h_m = 0.7) \cup (p_m = 30, h_m >= 0.65) \tag{8}$$

Section 6.1, considers the reasons why the instrument is strong in only selected parts of the sample. The average optimal bandwidths for each outcome variable in this segment are shown in the appendix (Table A.I).

5.2 Mapping the Theoretical Results to the Data

The model in Section 2 presents four predictions for the effect of an income increase on budget allocation, incumbency advantage, quality of politicians, and clientelism. These predictions are tested using the cross-municipal variation in CCT coverage, which is instrumented by the discontinuity in the FHP funding, as described in Section 3. The effect is always calculated using equation 7 for the treatment frontier segment defined in Section 5. The actual level of clientelistic exchanges cannot be directly measured within the scope of this study. I use instead alternative tests where I observe the effects of CCT on boh the electoral performance and pro-poor public allocation of different types of politicians, according to educational background and clientelism score of their parties. These two measures act as a proxy for the candidate's comparative advantage in clientelism.

Proposition 1 is tested using the local budget allocation. The data categorizes municipal spending on the basis of function, and the following six categories represent nearly 90% of the total spending: education, health, administration, urbanization and housing, social security, and transportation. I define the pro-poor public good in terms of spending in education and health services, and the effects of CCT coverage on pro-poor spending are used as a proxy for effort allocated to pro-poor public goods.

The variable that maps the data to Proposition 2 is the share of votes of the incumbent, which is measured only for incumbents that decide to run for reelection. To lend support to these results, I also examine secondary political outcomes that measure the competition in local elections (e.g. margin of victory and number of candidates in elections).

For Proposition 3, I proxy the type of incumbents using their education level and the clientelism score of their parties. The main assumption is that more educated politicians are also relatively more efficient at public good distribution (and less efficient at clientelism, with $\bar{\theta}$). The effect of CCT on the incumbent's vote share and pro-poor allocation is then observed for the different subsamples based on both the incumbent's education and the clientelism score of parties. Additionally, if we assume that voters could observe the true type of challengers before the election, an income increase should make clientelistic challengers less competitive in elections. This papers measures this effect using the education and party of challengers, through four different dependent variables: (1) the share of less-educated challengers entering the race; (2) the share of less-educated challengers finishing top 2 in the race; (3) the share of challengers from clientelistic parties entering the race; and (4) the share of challengers from clientelistic parties finishing top 2 in the race.

Although Proposition 4 does not provide new insights on political outcomes, the alignment of the empirical results with the prediction provides support for the mechanism under examination. This paper uses the subsample of municipalities with a high share (at least 25%) of poor families as the paper's main specification, and all main results are published for this sample. Table A.IV (appendix) shows the main results for all municipalities. The construction of all variables referenced above is fully described in Section 4.

6 Results

6.1 First Stage: Regression Discontinuity Results

Table A.II in the appendix shows that the municipalities on the two sides of the treatment frontier are comparable in many covariates that are fixed or measured before treatment. The coefficients are shown for the preferred frontier segment, and are neither statistically significant at the optimal bandwidth (column 3), nor at alternative bandwidths (columns 1 and 2).

Table I shows the first stage results for health funding and CCT coverage, for the preferred frontier segment. The coefficients estimated for individual points along the entire frontier can be seen in Figure VI. Column (1) is the preferred specification. It shows that a municipality at the discontinuity would have annual basic health transfers of R$1.75bn (pre-treatment), with the treatment triggering an increase of 24% (R$0.43bn). At the same time, CCT coverage is 7.9pp (percentage points) higher, and the additional amount of resources received by voters in treated municipalities is R$0.7bn, which is 60% higher than the differential in health funding generated by the discontinuity.

The remainder of the table shows the robustness of results to the choice of bandwidth, and the inclusion of controls. Robustness to the choice of kernel is shown in Table A.V in the appendix. There is no significant variation in the magnitude of 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 in this context, although seemingly small, would work in favor of the results.

Figure V shows the heterogeneity of the first stage results. The first plot has the treatment effects on CCT coverage at the discontinuity, for different points along the population x HDI frontier. The second plot shows the pre-treatment average of a few variables that measure the quality of both pre-existing infrastructure and old CCT coverage, also for the same points in the treatment frontier. As expected, the results show that the CCT coverage only responds to the FHP discontinuity in locations with the following characteristics: low infrastructure of public services, and a low coverage of old CCT programs (see the discussion in Section 3). The treatment effect varies significantly across municipal characteristics. Wealthier and smaller municipalities (left side of the figure) have a coefficient that is barely statistically significant, and weak in magnitude, indicating that health funds are a weak instrument 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 instrument is only strong at the filled blue circles (2 points on the left and 5 points on the right), which coincide with low coverage of public services and Old CCT.

6.2 Second Stage: Political Outcomes

Table II shows the main results for political outcomes. Column (1) shows the reduced form coefficients from the regression discontinuity in the treatment period (2008,2012); column (2) shows the same coefficients for elections in the pre-treatment period (2000,2004); and column (3) shows the IV coefficients for the effect of CCT on political outcomes, post-treatment. Table A.V and Table A.VI in the appendix show the robustness of the results to kernel and bandwidth choice, and the inclusion of municipal controls. Figure VI and Figure VII show the individual treatment effects at every one of the 19 points on the population x HDI frontier for all relevant variables.

Proposition 1 The main result of interest is the budget allocated to pro-poor spending. From Table II we learn that these services represent 52% of the budget, pre-treatment. In line with the theory, column (1) shows an increase of 4.0 pp (percentage points) in the share of budget allocated to these areas. From the IV regression (column 3), for a 10pp increase in CCT coverage, the pro-poor spending increases by 4.8pp. Column (2) provides a placebo test of the identification strategy, as the effects are estimated for outcomes measured in 2000 and 2004, before the policy of discontinuous funding for the FHP was implemented.[41] The positive effect for pro-poor spending was not present at the discontinuity in electoral tenures prior to 2004, for which the result was negative and insignificant.

Table A.III in the appendix shows the detailed coefficients for all six individual budget categories, as well as for total budget and its breakdown by expense type (i.e. capital investment, personnel or others). No other functional budget category had relevant coefficients,[42] which implies that the cost of re-allocating budget to pro-poor spending was divided across most other categories.

However, the examination of the budget breakdown by expense type indicates both a decrease in allocation to capital investment and an increase in expenses with personnel. This might suggest that incumbents reallocate resources to redistributive spending from infra-structure investment. In line with this dynamic, higher personnel expenses were most likely due to the shift from capital investment to labor-intensive spending in education and health. While the potential implications of this budget shift in the profile of public employees is outside the focus of this study, it is briefly discussed in the appendix (Section B).

Proposition 2 The main result of interest here is the loss of support by the incumbent, which is a direct model prediction. From Table II, the pre-treatment average vote share for incumbents is 51% in the elections of 2008 and 2012. In line with the theory, columns (1) and (3) 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 8.3pp vote loss for the incumbent. This result is robust to the choice of kernel, bandwidth, and the exclusion of municipal controls (Table A.V and Table A.VI). Column (2) also 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. These variables measure different dimensions of electoral competition (margin of victory and number of candidates in elections). From the IV results in column (3) of Table II, for a 10pp increase in CCT coverage, there are 0.4 more candidates running for mayor, and a 6.4pp lower margin of victory. Again, these results are also robust to the choice of kernel, bandwidth, and the exclusion of municipal controls; and no significant effect on those variables are observed in past elections.

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. All in all, this study's identification strategy shows evidence that the loss of support by the incumbent goes beyond the affected poor families, but it cannot identify which of the propagating effects is more relevant.

Proposition 3 Are these effects heterogeneous by the type of politician? The first test is shown in Table II, where I examine the behavior and electoral performance of challengers by their levels of education and party profile. Without treatment, the share of challengers without high school that enter the race is 11%, whereas the share that rank top 2 in elections is 14%. At the discontinuity, less educated challengers are no more or less likely to enter the race than more educated ones, but are less likely to rank top 2. From the IV results in column (3), for a 10pp increase in municipal CCT coverage, the share of less educated challengers that rank top 2 in the election is 14pp lower.

As for party profile, without treatment, the share of challengers from clientelistic parties that enter the race is 37%, whereas the share that rank top 2 in elections is 46%. At the discontinuity, challengers from clientelistic parties are less likely to enter the race in treated municipalities, but there is no difference in the share that rank top 2. From the IV results in column (3), for a 10pp increase in municipal CCT coverage, the share of challengers from clientelistic parties that run for mayor is 14pp lower. It is possible that when clientelistic parties expect to be less competitive in the election, they decide to run in a coalition supporting a more competitive candidate from another party. In this case, only clientelistic parties with high-valence candidates would decide to have a mayoral candidate, which would explain why we do not observe any effects by party for challengers that rank top 2 in the election. All these results are also robust to the choice of kernel, bandwidth, and the exclusion of municipal controls; and no significant effect on those variables are observed in past elections.

This model prediction can also be tested by splitting the sample by both the party and education of incumbents. The results for the incumbent's vote share and pro-poor spending are shown in Table III. The magnitude of the effects follows the theoretical prediction, but some coefficients lose statistical significance due to small sample size. As expected, the loss in vote shares is higher an statistically significant for the subsample of incumbents with less than a college degree, but the shift towards pro-poor redistributive spending in lower for these incumbents (it is high and significant for the high-education sample).[43] A similar pattern is observed when comparing the vote shares of incumbents from programmatic an clientelistic parties, the loss is high and significant for the clientelistic subset.[44] Finally, the differences by party in the effects on pro-poor spending are less pronounced. One possible explanation here is that the shift towards programmatic pro-poor effort should be higher if the initial levels of programmatic effort are low, as predicted by the model. This could be the case in places where clientelism is more attractive due to the unique characteristics of the municipality, in which case these places are also more likely to have a clientelistic incumbent.

Proposition 4 All the estimated effects are expected to be stronger for municipalities with a higher share of poor population (BF-eligible). All tables in the paper report results for a sample where municipalities have at least 25% of their population in this category. Table A.IV shows the estimation including all municipalities. The results remain highly significant for the incumbent's vote share, but the coefficient has a slightly lower magnitude. For the pro-poor spending, the coefficient's magnitude drops by nearly 40% and remains only significant for a 90% confidence level.

Anti-Clientelism or Leftism? In Brazil, the line that separates programmatic and clientelistic parties is similar to the one separating left an right-wing parties.[45] One could argue that the effects estimated here are just reflecting a general shift of poor voters towards supporting left parties as a consequence of higher CCT coverage, especially since the left-wing President Lula (PT Party, 2003-2010) is widely seen by the poor population as the responsible for the spread of CCT benefits (Zucco, 2013). In turn, left parties would also focus more effort on pro-poor public goo distribution.

However, this hypothesis does not fit the results when we examine the performance of both incumbents an challengers by ideology. Table III shows the main results for subsamples of incumbents from left or right-wing parties. Both the loss in vote shares and the shift towards pro-poor goods is nearly the same for left and right-wing incumbents, with right-wing parties actually providing slightly higher level of pro-poor spending than their leftist counterparts. In Figure VII we look at entry and performance of challengers from left parties, in similar exercise to the one conducted for their education level. Here, left-party challengers are neither more likely to enter nor to succeed in electoral races than their right-wing counterparts.

6.3 Exclusion Restriction

The existence of FHP direct effects on political outcomes cannot be tested separately from the effects happening through CCT. Instead, I rely on three pieces of evidence indicating that the political outcomes are predominantly generated by the CCT variation, and not by the FHP (i.e. supporting the exclusion restriction).

First, surveys indicate that the FHP is rated as the best health program in Brazil (see IPEA (2011)), with approval rate above 80%. Given that the FHP is run by municipalities, if the results observed in this paper are driven predominantly by direct FHP effects, we would have to conclude a highly rated local health policy triggers a significant loss of support for politicians that implement it, which seems unlikely. As for budget allocation, even if the entire FHP extra funding is actually spent in health services,[46] the predicted increase in the budget share of health services would only 0.2pp, nearly 9 times lower than the point estimate found in the paper.

Second, the heterogeneity of first stage average treatment effects (i.e. effects of FHP funding on CCT coverage) along the treatment frontier provides additional evidence in favor of the exclusion restriction. The coefficients for political outcomes show a weak response in areas of the frontier that coincide with a weak instrument (Figures VI and VII). In other words, the heterogeneity of the first-stage results matches the heterogeneity of the reduced-form estimation for most outcomes. Although this does not rule out the presence of FHP direct effects, it is strong evidence that they happen through the changes in CCT coverage.

To further investigate this, I compare two equal size ranges of the frontier (6 bins): the first was the preferred segment for this paper, where the instrument is statistically strong (see page 16). The second is a segment with population in the range 7,500-17,500, and HDI of 0.7, where the instrument is weak.[47] The coefficients for the weak-IV segment are shown in Table A.IX in the appendix. All variables that were significant in the preferred segment become statistically insignificant.

Third, I attempt to identify changes in the direct effects triggered by the arrival of BF using a dataset containing the date of FHP implementation in each municipality. Here, I regress the various political outcomes (y_{mt}) on a dummy that indicates the presence or coverage of FHP in municipality m at time t (FHP_{mt}), interacted with time dummies. Because the FHP arrival is not random, this regression might suffer from omitted variable bias even with the inclusion of municipality fixed-effects, i.e., political outcomes could be affected by the same time-municipality unique factors influencing the decision to implement the FHP. Nevertheless, under the assumption that this omitted variable bias is constant over time, the interaction between a time dummy for before and after the consolidation Bolsa Família ($postBF_m$), and FHP presence (FHP_{mt}), should not be statistically significant unless there is another event changing the program's impact on politics. This effect is measured by α_1 in the equation below.

$$y_{mt} = \alpha_0 + \alpha_1 FHP_{mt} * postBF_m + \alpha_2 FHP_{mt} + postBF_m + \gamma_{Muni} + \mu_{mt} \tag{9}$$

Table A.VIII (appendix) shows the coefficients α_2 and α_1 for two different definitions of FHP presence: while specification (A) defines FHP as a dummy indicating if the program has been implemented or not in the municipality, specification (B) defines FHP as a dummy indicating if the program covers at least 50% of the targeted population in the municipality, or not. Post-BF (2008 an 2012 electoral cycles), the effect of FHP on political outcomes changes always in the direction predicted by theory, and in line with the main MRDD results in this paper. These effects are also statistically significant for the main variables of interest in at least one specification, namely pro-poor spending and the incumbent's vote share (albeit only at a 90% confidence for the latter).[48]

6.4 *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.[49] 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 paper.

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 (appendix, Table A.III). This is not surprising, given that the theoretical FPM differential is much higher at lower population thresholds.[50]

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

Fourth, Table A.VII (appendix) shows the reduced form coefficients for political outcomes in 2008 and 2012, setting the population cutoff 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 in this paper.

7 Conclusion

This paper employs a multivariate regression discontinuity design and a novel identification strategy to estimate the effects of a CCT program on the local politics of Brazilian municipalities. It shows that income transfers reduce incumbency advantage, increase both electoral competition and the quality of candidates, and weaken 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 here reconciles these empirical findings by showing that cash transfers reduce the ability of incumbents to raise support with clientelism.

These results have policy implications. First, they suggest that policy spillovers matter. In addition to the political impacts of the CCT program, this paper also shows that a small differential in funding for a health program generated a much larger impact on the CCT distribution.

Second, they shed light on what are the incentives for national politicians to implement similar income transfers programs. If properly shielded from political manipulation, CCT programs may increase the ability of national governments to shape local politics in their favor by weakening clientelistic machines of opposition parties. 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, these findings are useful to inform other poverty reduction policies. The results here provide evidence that reducing poverty in a way that limits credit claiming by political actors might have broader consequences to both the political and economic environments.

Finnaly, this paper shows that the CCT program is positive for the poor sectors of the population, given the budget shift in their favor. More electoral competition and better politicians are also associated with more accountability. However, the overall welfare effects of the program remain uncertain. Anticipating the long-term consequences of less availability of capital-intensive public goods are beyond the ability of the identification strategy employed here. Also, the magnitude of effects has to be treated carefully when applied to a more general context, for two reasons. First, the results magnitude may well depend on institutional features that are specific to the Brazilian case. Second, the effects may still be affected by a residual direct impact of the FHP funding, even though the evidence on the exclusion restriction supports the direction of the results presented here.

References

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* .

Bardhan, Pranab and Dilip Mookherjee. 2005. "Decentralizing antipoverty program delivery in developing countries." *Journal of Public Economics* 89 (4):675 – 704.

Bobonis, Gustavo, Paul Gertler, Marco Gonzalez-Navarro, and Simeon Nichter. 2017. "Vulnerability and Clientelism." *Working Paper* .

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

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

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

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* .

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* 58:1132–1155.

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

Ferraz, Claudio and Frederico Finan. 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.

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.

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

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. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23 (1):127–155.

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

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.

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 and Michael Peress. 2017. "Request Fulfilling: When Citizens Demand Clientelist Benefits." *Comparative Political Studies* 50 (8):1086–1117.

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.

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.

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

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).

Notes

[1]I use the terms clientelism and vote buying interchangeably to mean the exchange of a private transfer to a select group of voters, using public resources, with the expectation that it will influence electoral support.

[2]See Brusco, Nazareno, and Stokes (2004); Vicente and Wantchekon (2009); Fujiwara and Wantchekon (2013); Anderson, Francois, and Kotwal (2015); Cruz, Keefer, and Labonne (2015); Bobonis et al. (2017).

[3]Stokes et al. (2013) discuss the role of economic development in reducing the efficacy of local political brokers, and Bobonis et al. (2017) show with an experiment of cisterns distribution in Northeastern Brazil that voters depend less on the public sector for their subsistence once they become less vulnerable.

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

[5]The press, academics, and politicians have engaged in debates about the political impact of BF since its inception. President Dilma Rousseff claims that *Bolsa Família* swept clientelism off the country (http://glo.bo/1f46NiQ). This view is supported by Rego and Pinzani (2013). Nevertheless, criticisms of the program's political impacts are abundant. The press has often pointed out that the program is subject to manipulation by local politicians (http://goo.gl/COMxng); or that the program serves as a vote buying machine for the federal government in national elections (http://goo.gl/Jg96Mi and http://goo.gl/SNp6kr).

[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]Parties are labeled programmatic or clientelistic based on the ALP survey. Full details are available in Section 4.

[8]Unlike this paper, Rodriguez-Chamussy (2015) focuses on party incumbency rather than candidate incumbency.

[9]See Fried (2012) and Zucco (2013) for historical accounts

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

[11]For example: annual budget, total cap on the benefits, eligibility thresholds, and municipal coverage targets.

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

[13]Teams include a minimum of one family doctor, one nurse, one assistant nurse, and six health agents.

[14]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.

[15]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 therefore excluded from our sample.

[16]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.

[17]The mechanism respects the traits of vote buying in Brazil, described by surveys and press reports. It characteristically involves the use of public resources and targeting of the poor and vulnerable. 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. Nichter and Peress (2017) show that in Brazil poor voters demand clientelistic transfers from local politicians when they are exposed to negative shocks. In addition, the following examples of media coverage (in Portuguese) provide insights on the mechanics of this practice: *http://goo.gl/GqdkWf*;*http://goo.gl/XecDEu*; and *http://goo.gl/VSsvO2*.

[18]If $\kappa > 0$ ($\kappa < 0$), it means that the cost related to effort in implementing public goods is relatively higher (lower) than the cost of effort allocated to clientelism.

[19]This is in keeping with the context of many developing democracies.

[20]Opportunistic incumbents always maximize their allocation of effort based on their type-specific cost in period 1. In period 2, they have no incentive to change this allocation given that there are no reelection incentives and no allocation cost if they keep effort constant across periods.

[21]$U^I(g,c) - U^C(g,c)$

[22]The global coverage target for BF is the sum of local targets, and is binding. The local coverage targets are not binding, i.e., the sample contains a large number of municipalities with coverage higher than 100%.

[23]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.

[24]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.

[25]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 again in 2011, in order to incorporate data from the 2006 PNAD survey and the 2010 census, respectively.

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

[27]The number of eligible families per municipality was estimated three times by the MDS, based on the PNAD surveys from 2004 and 2006, and the 2010 census. For the CCT coverage in 2008 and 2012, I use the estimates of poor families from 2006 and 2010, respectively. For the pre-treatment CCT coverage calculated in June 2004, I use the 2004 estimate.

[28]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.

[29]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 (sanitation was on average only 6% of the aggregated expenses in 2004-2012). For 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.

[30]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 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 higher or equal than 3 (out of 4 in an increasing scale) are identified as clientelistic parties. All small parties that were not evaluated by DALP are identified as non-clientelistic. Between 2008 and 2012, 95% of municipalities were governed by a party that was represented in the DALP survey.

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

[32]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.

[33]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 the cut-offs, in contrast to the average treatment effect for the entire frontier.

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

[35]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.

[36]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).

[37]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 also included as a control.

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

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

[40]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.

[41]Municipalities started to receive the extra funding under the new policy in August 2004, a few weeks before the election that happened in early October. This paper works under the very plausible assumption that there was not enough time to consistently observe the mechanism proposed here in such a short period of time. In fact, in August 2004 the budget allocation for the year was already defined, and the candidates for the municipal election were already campaigning.

[42]Transportation spending (2.7% of the total spending) was negative and significant. However, this result has to be treated with skepticism, given that the effect is not robust to the IV, and it also pre-dates the treatment (it was already present in 1997-2004). Transportation spending in small municipalities is mostly infra-structure spending in road transport, as opposed to a public system of transportation.

[43]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.

[44]Notice that the confidence intervals for the two coefficients overlap in many of these cases, since the subsamples have a much lower number of observations the estimation has high variance.

[45]In the DALP survey, the differences arise because of two parties: PSDB is a right-wing programmatic party, while PDT is a left-wing clientelistic party. All other 9 parties are either left-wing and programmatic or right-wing and clientelistic.

[46]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.

[47]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.

[48]The results are also significant for the entry of challengers from clientelistic parties.

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

[50]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.

Table I: Regression Discontinuity Results

Dependent Variable	Mean (Pre-Treat.)	Coefficient (1)	(2)	(3)	(4)
Health funds, R$bn	1.752	0.219***	0.220***	0.249***	0.213***
CI	[1.64,1.86]	[0.14,0.29]	[0.14,0.31]	[0.18,0.31]	[0.13,0.29]
Obs. per bin		329	329	476	314
Bandwidths		(0.71,0.99)	(0.71,0.99)	(0.90,0.90)	(0.75,0.75)
CCT cov, p.p. over target	0.515	7.897***	6.969**	8.439***	8.935***
CI	[-2.64,3.07]	[3.84,11.60]	[2.15,11.38]	[4.11,12.43]	[3.91,13.41]
Obs. per bin		612	612	476	314
Bandwidths		(1.00,1.00)	(1.00,1.00)	(0.90,0.90)	(0.75,0.75)
Bandwidth type	Optimal	Optimal	Optimal	0.90	0.75
Municipal controls	Yes	Yes	No	Yes	Yes

Health Funds are measured in log(Variable). Significant at: 99% ***, 95% **, 90% *. All variables are measured at the end of the election years, 2008 and 2012. Confidence intervals in square brackets are clustered by municipality. Optimal bandwidths in standard deviations of the population and HDI averages of 16,000 and 0.07, respectively. The coefficients represent the average effect for the preferred frontier segment (6 bins). Regressions include year and state effects. The first (non-numbered) column corresponds to the predicted values for a municipality at the discontinuity before treatment. The list of included controls is described in the text. Column (1) has the main specification under the optimal bandwidth. Column (2) is the same as (1) but without municipality-level controls. Columns (3) and (4) are the same as (1), but with constant bandwidths of 0.90 and 0.75, respectively.

Table II: Main Results: Political Outcomes

	Coefficient, [90% CI]				Avg Band.
	Mean	RDD	RDD past	IV	(Pop. HDI)
	(Pre-Treat.)	(1)	(2)	(3)	Obs. per bin
Incumbent's vote share	51.010	-7.750***	1.075	-0.827**	(0.99,0.96)
(%)	[47.90,53.85]	[-12.46,-3.66]	[-2.96,5.71]	[-1.91,-0.28]	434
Margin of victory	16.715	-6.015**	-0.725	-0.638**	(1.00,1.00)
(p.p.)	[13.77,20.82]	[-11.73,-1.77]	[-4.17,3.28]	[-1.70,-0.10]	470
Candidates	2.312	0.392***	0.118	0.042***	(1.00,0.98)
(number)	[2.24,2.40]	[0.20,0.63]	[-0.06,0.32]	[0.02,0.10]	452
Pro-poor spending	51.816	4.022***	-0.981	0.483**	(1.00,1.00)
(% share)	[49.67,52.48]	[1.73,7.25]	[-3.33,1.11]	[0.12,1.90]	350
Challenger's entry	0.112	-0.001	-0.020	0.000	(0.97,0.98)
(share without HS)	[0.08,0.17]	[-0.08,0.09]	[-0.11,0.07]	[-0.01,0.01]	795
Challenger's entry	0.370	-0.138**	-0.049	-0.014*	(0.97,0.95)
(share Clientelistic)	[0.32,0.42]	[-0.23,-0.02]	[-0.15,0.06]	[-0.04,0.00]	771
Challenger is top 2	0.139	-0.128**	0.051	-0.014**	(0.98,0.98)
(share without HS)	[0.09,0.20]	[-0.21,-0.04]	[-0.08,0.19]	[-0.03,0.00]	468
Challenger is top 2	0.458	-0.116	-0.069	-0.012	(1.00,0.98)
(share Clientelistic)	[0.38,0.53]	[-0.26,0.04]	[-0.21,0.07]	[-0.04,0.01]	472

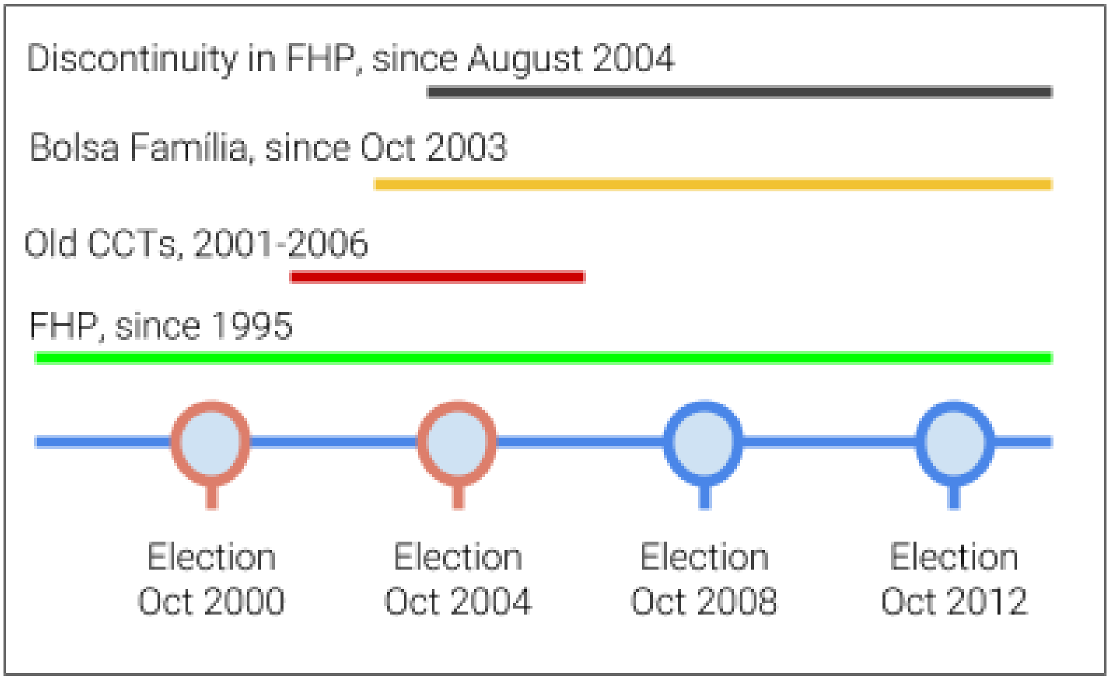
Significant at: 99% ***, 95% **, 90% *. Confidence intervals in square brackets are clustered by municipality. Optimal bandwidths are shown in parenthesis as standard deviations of the population and HDI averages of 16,000 and 0.07, respectively. Number of observations are shown below the bandwidths. The coefficients represent the average effect for the preferred frontier segment (6 bins). Regressions include year and state effects. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment. The list of included controls is described in the text. Column (1) shows the RDD effects of the discontinuity in health funding on political outcomes for the post-treatment period (2008,2012); Column (2) shows the same effects as (1), but for the pre-treatment period (2000,2004); Column (3) shows the IV regression for the post-treatment period (2008,2012).

Table III: Heterogeneity in Political Outcomes

	Incumbent's Vote Share		Pro-poor Spending	
split by: Incumbent's Education	*High*	*Low*	*High*	*Low*
Coefficient	-4.705	-11.127**	8.041***	0.374
CI	[-10.67,0.61]	[-20.06,-3.21]	[4.74,14.07]	[-2.76,3.81]
Obs. per bin	229	205	136	122
split by: Incumbent's Party	*Programmatic*	*Clientelistic*	*Programmatic*	*Clientelistic*
Coefficient	-2.537	-8.002*	3.150*	2.992*
CI	[-11.65,4.34]	[-15.56,-0.19]	[0.55,8.45]	[0.18,7.12]
Obs. per bin	152	219	112	187
split by: Incumbent's Party	*Left*	*Right*	*Left*	*Right*
Coefficient	-7.007*	-7.472**	3.316*	3.624**
CI	[-15.25,-0.38]	[-13.70,-1.97]	[0.17,8.84]	[1.17,7.40]
Obs. per bin	130	304	90	259

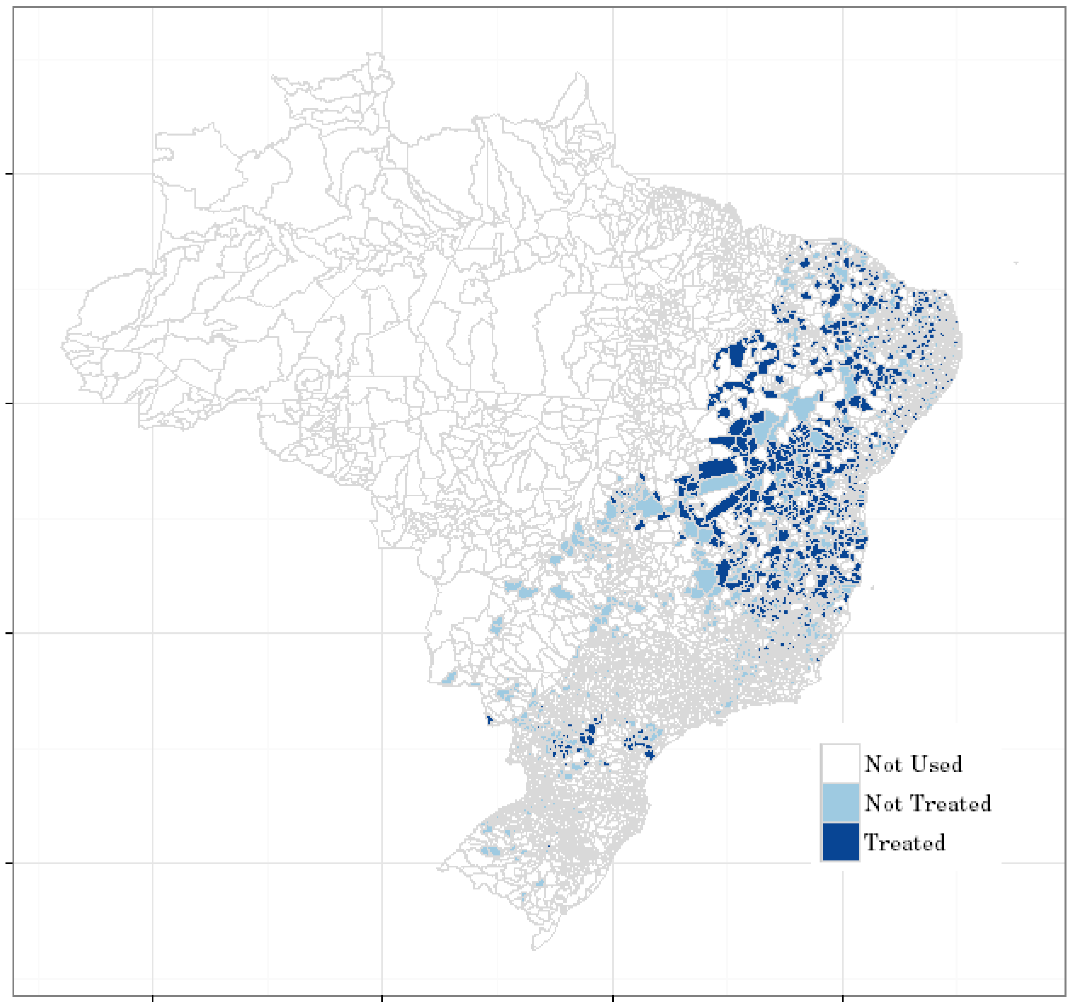
Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality. The coefficients represent the ATE for the preferred frontier segment (6 bins). All regressions include year and state effects, and the municipal-level controls described in the text. Highly educated incumbents have a college degree; less educated ones have at most some post secondary education. High clientelism parties are: PP, PDT, PTB, MDB, PR, DEM. Low clientelism parties are: PT, PSB, PPS, PSDB and PC do B. The subsamples by clientelism only include parties with a DALP clientelism score. Left parties are: PT, PDT, PSB, PPS, PSTU, PSOL, REDE, PCO, PMN, PCB and PC do B.

Figure I: Timeline of Events



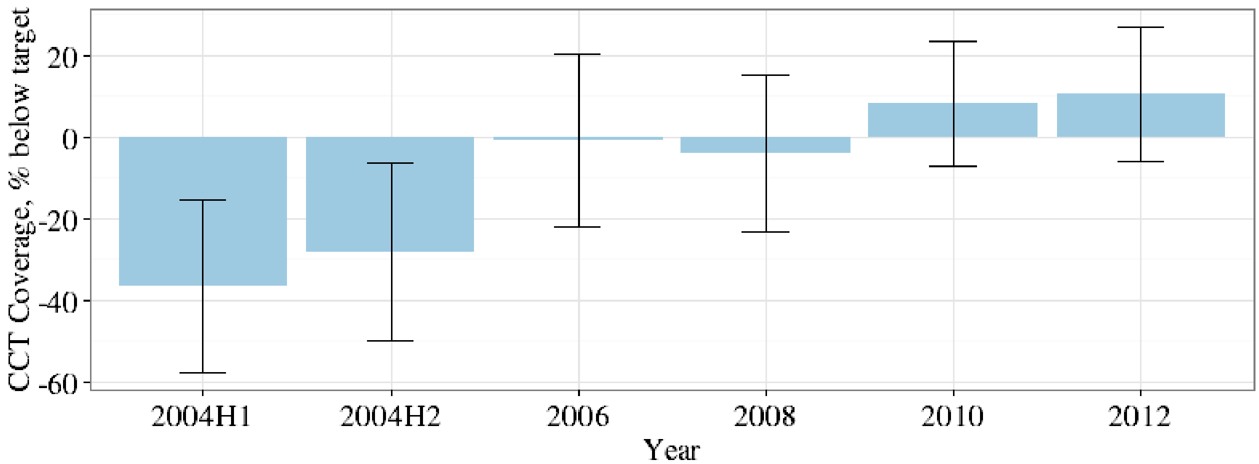
The FHP started in 1995. Bolsa Família was created in Oct 2003, and the FHP funding discontinuity in Aug 2004. All elections happened in early October.

Figure II: Map of Municipalities in the Sample



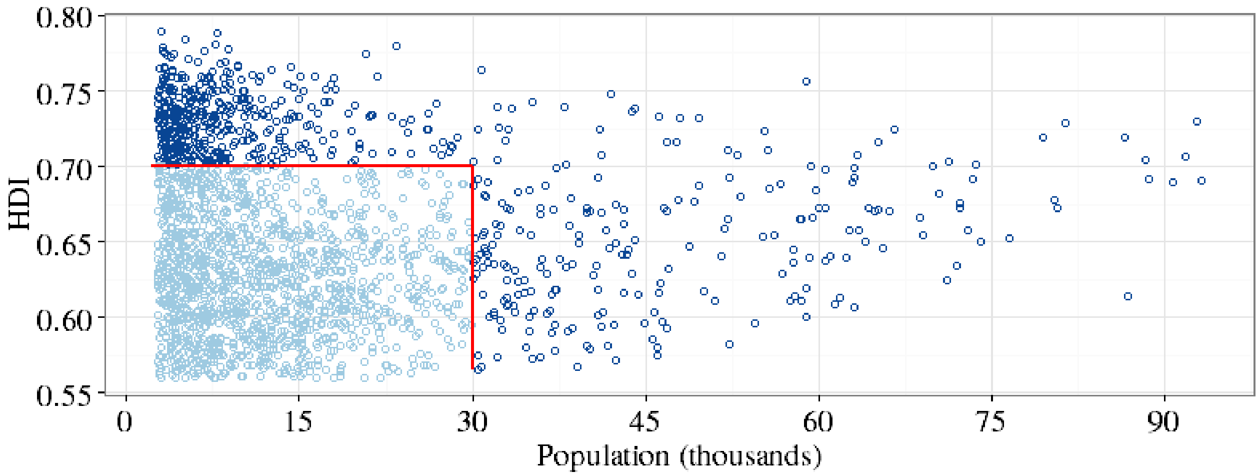
Municipalities in white are either outside of the bandwidth used for the RDD, located in the legal Amazon (high-left side) or have a share of poor population below 25% (most of the southern municipalities). The map shows a total of 1,441 colored municipalities.

Figure III: CCT Coverage vs. Number of Poor Families



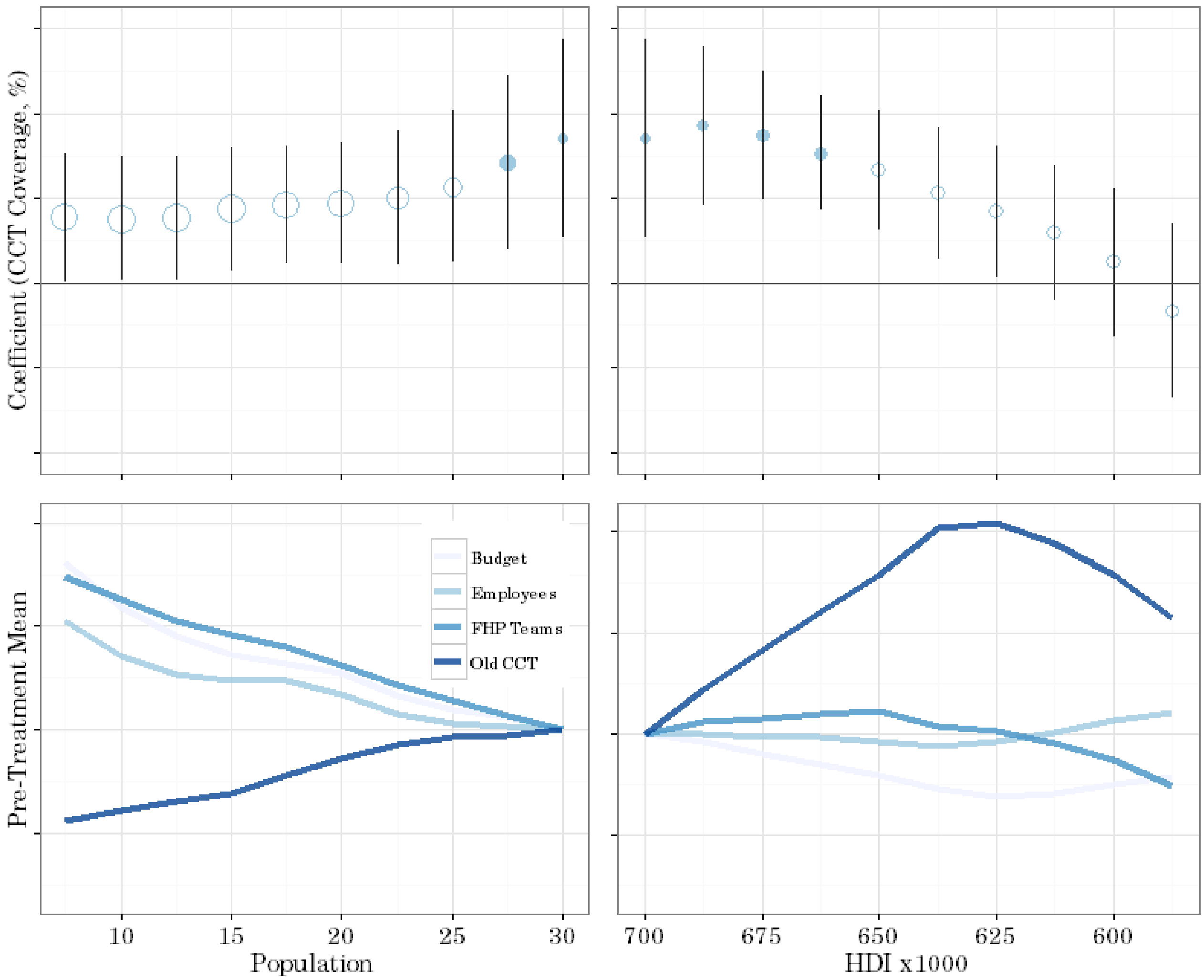
The y-axis shows the gap between CCT coverage and the target of poor families in each municipality, in percentage points. For example, in the first half of 2004, CCT coverage was nearly 40% below the government's target. The vertical lines show the standard deviation of this coverage gap, across municipalities. CCT coverage includes *Bolsa Família*, *Bolsa Escola*, *Bolsa Alimentação* and *Cartão Alimentação*. **The 2004 coverage is adjusted to account for 974,000 households receiving both *Bolsa Família* and *Bolsa Escola* at the same time, as disclosed to the press by MDS. If there were other families receiving duplicated benefits before 2006, the coverage gaps could have been even higher. After 2006 the vast majority of beneficiaries of old CCT programs migrated to Bolsa Família, so the coverage is measured precisely.

Figure IV: Potential Sample and Treatment Frontier



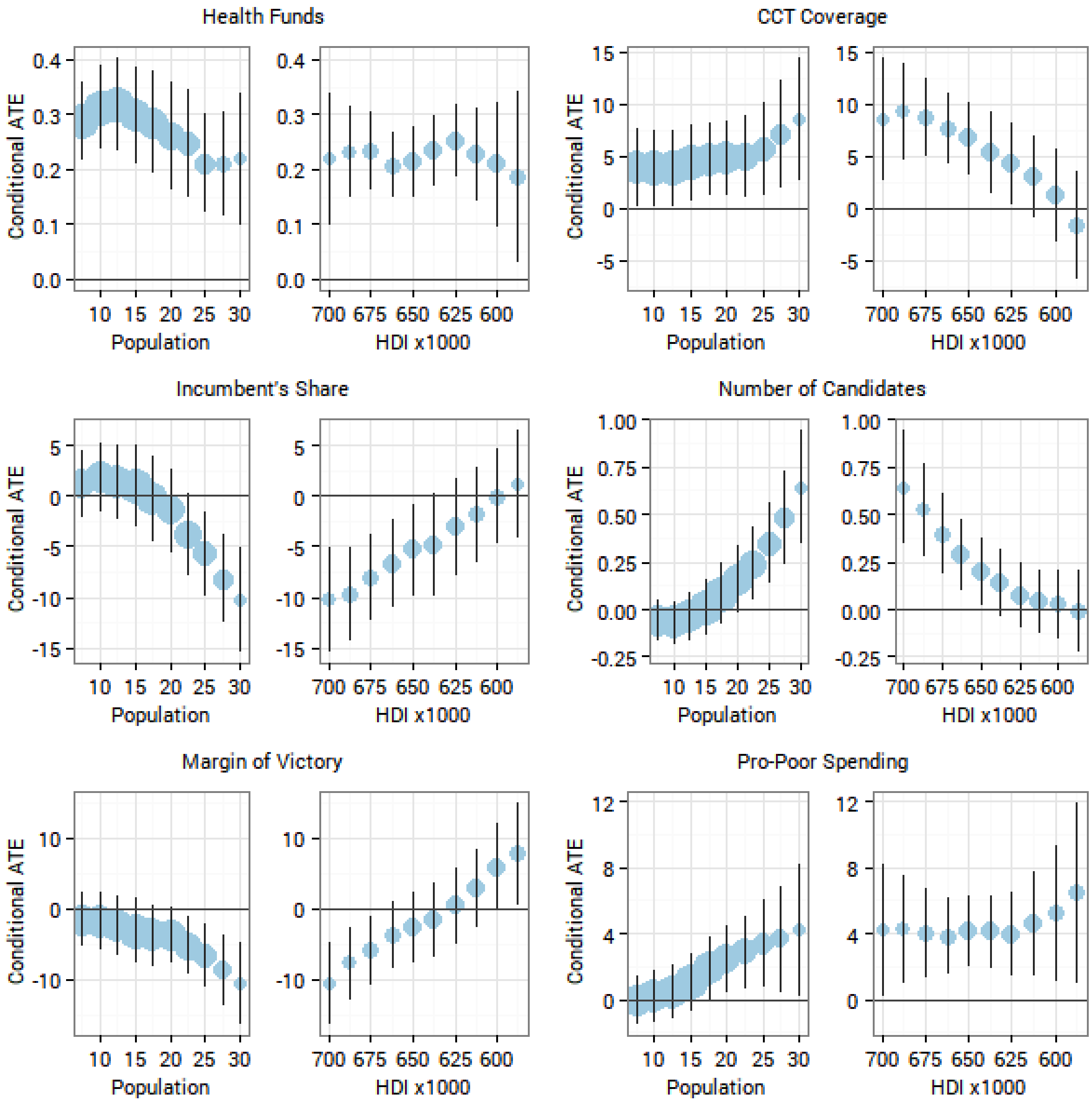
The potential sample includes all municipalities within the central 95% percentile in population and HDI. The treatment frontier is the red line. Light blue dots represent municipalities eligible to treatment.

Figure V: Heterogeneity of the CATE for CCT Coverage Along the Frontier



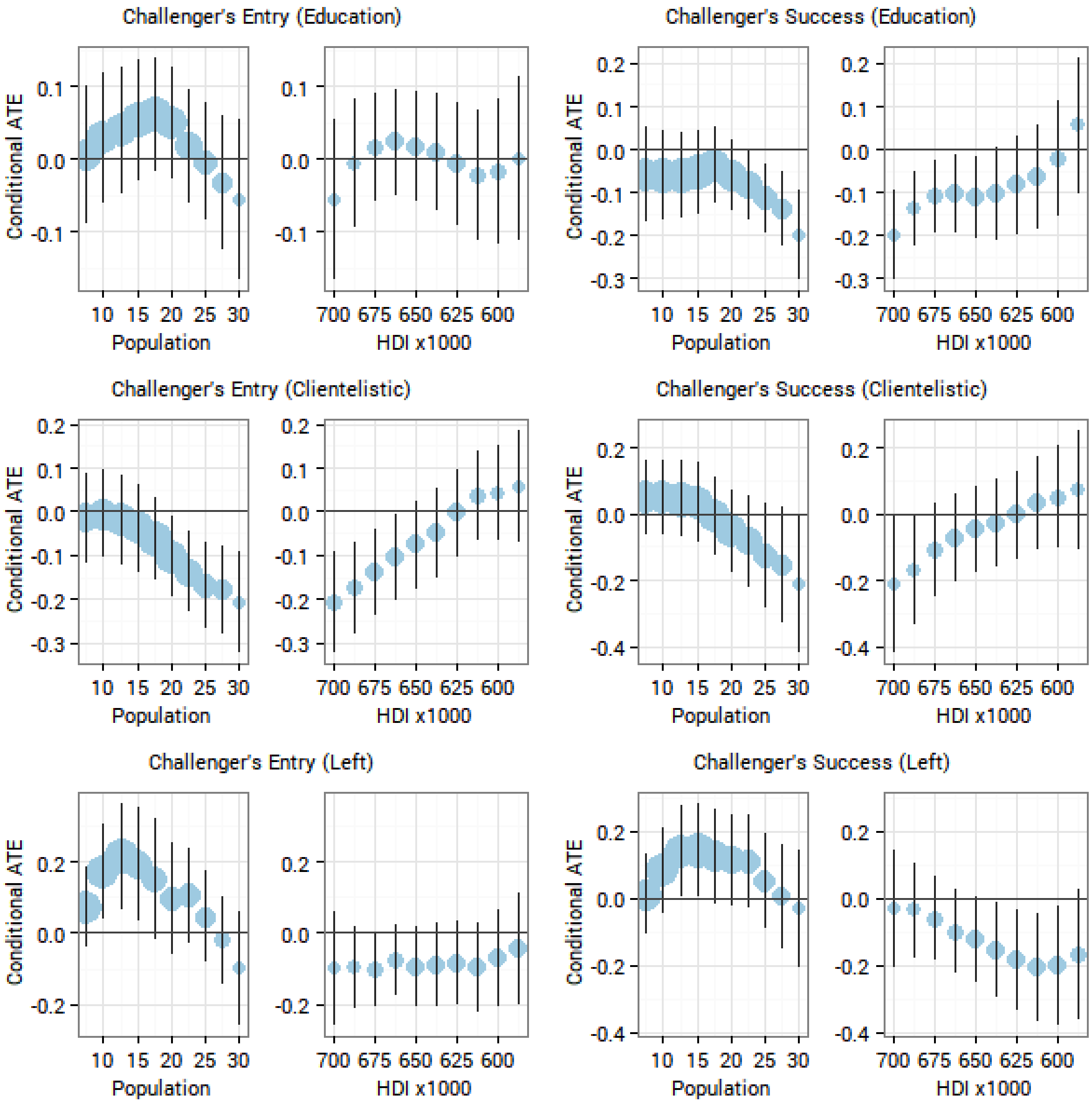
The y-axis shows in the first plot the conditional ATEs for CCT Coverage. The y-axis on the other plots shows the pre-treatment average of the variables, converted to a common scale. For all charts, the left side has HDI fixed at 0.7 and population between 7,500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. All coefficients are from a local linear regression (edge kernel) including year and state effects, and the municipal controls. The bandwidths are set as one standard deviations of 16,000 for population and 0.07 for HDI. The size of the dots represent the number of observations in each bin. There are 19 different bins, I repeat the bin located at the origin in both sides (Pop=30,000, HDI=0.70).

Figure VI: Heterogeneity of the CATE



The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI fixed at 0.7 and population between 7,500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. All coefficients are from a local linear regression weighted by the edge kernel including year and state effects, and the municipal controls listed in the text. The bandwidth is the estimated optimal for each variable (see Table 1A). The size of the dots represent the number of observations in each bin. There are a total of 19 different bins, I repeat the bin located at the origin in both sides (Pop=30,000, HDI=0.70).

Figure VII: Heterogeneity of the CATE



The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI fixed at 0.7 and population between 7,500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. All coefficients are from a local linear regression weighted by the edge kernel including year and state effects, and the municipal controls listed in the text. The bandwidth is the estimated optimal for each variable (see Table 1A). The size of the dots represent the number of observations in each bin. There are a total of 19 different bins, I repeat the bin located at the origin in both sides (Pop=30,000, HDI=0.70).

A Mathematical Appendix

The equation to be maximized by incumbents is the following:

$$\max_{e} \frac{1}{2} + \psi n \mu_x [\theta \log(e) - \theta \log(E) + \delta(y)(E - e)] - \kappa e \tag{10}$$

The first order condition is then:

$$\psi n \mu_x [\theta \frac{1}{e} - \delta(y)] - \kappa = 0 \tag{11}$$

We can solve for e as follows:

$$e = \frac{\theta}{\delta(y) + \mathrm{K}} \tag{12}$$

where $\mathrm{K} = \kappa/(\psi n \mu_x)$. For an interior solution we must have $(\delta(y) + \mathrm{K}) > \theta/E$.

Proof of Proposition 1 As income y increases, $\delta(y)$ decreases by assumption ($\delta_y < 0$), which requires a new allocation of effort that reduces the total value of the term $1/e$. Thus, effort must be directed towards public goods. This can also be seen when we solve directly for e_y below:

$$e_y = \frac{-\theta \delta_y}{(\delta(y) + \mathrm{K})^2} > 0 \tag{13}$$

Proof of Proposition 2 The vote share of the incumbent is given by:

$$\pi = \phi n \mu_x [\theta \log(e) - \theta \log(E) + \delta(y)(E - e)] + \phi(\frac{1}{2\phi} - \lambda) \tag{14}$$

The proposition requires that $\pi_y < 0$. After substituting for the value of e, this term can be decomposed as:

$$\pi_y = \phi n \mu_x \delta_y [E - (e + \frac{\theta \mathrm{K}}{(\delta(y) + \mathrm{K})^2})] \tag{15}$$

The terms is negative when the component inside the brackets is positive. We know that $E \geq e$. Thus, the result holds as long as the value of $\mathrm{K} < 0$. If $\mathrm{K} > 0$, the result holds if the cost of allocation κ is low enough. The intuition is that incumbents face a trade off between reducing their utility cost of effort and increasing their probability of reelection. With an income increase, they will shift effort to public goods until the marginal gains in reelection chances equal the marginal utility cost of effort (for a positive κ). If *kappa* is too high, they might drastically increase effort in public goods and potentially improve their incumbency advantage. From here on, we assume that $\kappa > 0$.

Proof of proposition 3 First I show that clientelistic types allocate less effort to public goods when y increases. Starting with e_y from equation 13, the derivative of this term with respect to θ is shown below, giving us the desired result:

$$e_{y\theta} = \frac{-\delta_y}{(\delta(y) + \mathrm{K})^2} > 0 \tag{16}$$

Now as for $\pi_{y\theta}$, we start by substituting e in the equation for π_y:

$$\pi_y = \phi n \mu_x \delta_y [E - (\frac{\theta}{(\delta(y) + \mathrm{K})} + \frac{\theta \mathrm{K}}{(\delta(y) + \mathrm{K})^2})] \tag{17}$$

Now, derivating it with respect to θ we have the result:

$$\pi_{y\theta} = -\phi n \mu_x \delta_y [\frac{\delta(y) + 2\mathrm{K}}{(\delta(y) + \mathrm{K})^2} > 0] \tag{18}$$

Proof of Proposition 4 We need to show that $e_{yn} > 0$, and $\pi_{yn} < 0$. First we take the value of e_y from before and derivate it with respect to n, as follows:

$$e_{yn} = e_{y\mathrm{K}} \mathrm{K}_n = \frac{-2(\delta(y) + \mathrm{K})}{(\delta(y) + \mathrm{K})^4} * -\frac{\mathrm{K}}{n} > 0 \tag{19}$$

We can do the same to π_y, shown in equation 17. We know that $\mathrm{K}_n = -\mathrm{K}_n/2 < 0$ when $\kappa > 0$. Now, equation 17 can be decomposed as follows:

$$\pi_y = \phi n \mu_x \delta_y [E - \frac{\theta \delta(y)}{(\delta(y) + \mathrm{K})^2}] - \frac{2\theta \kappa \phi \delta_y}{\psi(\delta(y) + \mathrm{K})^2} \tag{20}$$

Derivating with respect to n and reorganizing we have:

$$\pi_y = \phi \mu_x \delta_y [E - e\frac{\delta(y)}{(\delta(y) + \mathrm{K})} - 2\theta \mathrm{K} \frac{(2\mathrm{K} + \delta(y))}{(\delta(y) + \mathrm{K})^3}] < 0 \tag{21}$$

Again here, assuming that $\mathrm{K} > 0$, the result holds for a low enough value of K, given that the sum of the first two terms in brackets is strictly positive, and $\delta_y < 0$.

B Effects on Hiring

The empirical evidence 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[51] to provide insight on how this budget reallocation can affect public services in the long term. The estimation results are shown in Table A.VII.

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 A.VII 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 shown for both the level and rate of change in these variables. The only noteworthy effect is a 8.3pp higher change in the share of temporary hiring in 2008-12. In the data, temporary hiring is generally positively correlated with budget changes.[52] 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 18.0pp 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 are not replaced by permanent positions[53], thus outliving the budget changes, there are at least two potential issues for future discussion here. First, there is a potential 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.

C 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,[54] 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.[55] 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 (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.[56] 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.[57]

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 expression, 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:

$$MSE = (Bias_1 + Bias_2 + Bias_3)^2 + 2 * Variance \tag{22}$$

where,

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

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.

Table A.I: Optimal Bandwidths

Kernel	Edge	Epanechnikov	Normal	Uniform
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)
Challenger's entry (share without HS)	(0.97,0.98)	(0.93,0.96)	(0.88,0.93)	(0.89,0.93)
Challenger's entry (share Clientelistic)	(0.97,0.95)	(0.95,0.94)	(0.91,0.91)	(0.90,0.90)
Challenger is top 2 (share without HS)	(0.98,0.98)	(0.95,0.96)	(0.89,0.93)	(0.89,0.92)
Challenger is top 2 (share Clientelistic)	(1.00,0.98)	(0.99,0.95)	(0.97,0.92)	(0.96,0.90)
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)
Pro-poor spending	(1.00,1.00)	(0.98,0.99)	(0.97,0.95)	(0.96,0.94)
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 A.II: Balance of Fixed and Pre-determined Variables

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

[a]Estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality and shown in square brackets. Optimal bandwidths are shown in parenthesis as standard deviations of the population and HDI averages of 16,000 and 0.07, respectively. The coefficients represent the average effect for the preferred frontier segment (6 bins). Regressions include year and state effects. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment.

Table A.III: Main Results: Other Budget Categories

	Coefficient, [90% CI]				Avg Band.
	Mean	RDD	RDD past	IV	(Pop. HDI)
	(Pre-Treat.)	(1)	(2)	(3)	Obs. per bin
Total budget	149.135	-0.042	0.085	-0.003	(0.94,1.00)
(RSbn)	[137.78,160.16]	[-0.13,0.03]	[-0.02,0.20]	[-0.01,0.00]	325
Capital investment	10.558	-1.726*	0.773	-0.150*	(1.00,1.00)
(% share)	[9.70,11.73]	[-3.34,-0.22]	[-1.18,3.32]	[-0.35,-0.02]	352
Expenses w. personnel	46.577	2.852**	0.201	0.248**	(1.00,1.00)
(% share)	[44.82,47.97]	[0.66,5.20]	[-1.65,2.21]	[0.07,0.60]	352
Other	42.750	-1.149	-0.812	-0.104	(1.00,1.00)
(% share)	[41.59,43.91]	[-3.05,0.82]	[-3.08,1.35]	[-0.34,0.05]	351
Health	22.225	2.174***	0.153	0.268**	(1.00,0.96)
(% share)	[21.10,23.39]	[1.05,3.73]	[-1.76,1.81]	[0.07,1.28]	329
Education	29.644	1.842**	-1.202	0.234	(1.00,1.00)
(% share)	[28.33,30.90]	[0.34,3.91]	[-2.56,0.09]	[-0.02,1.04]	350
Urbanization	9.750	-0.121	2.047*	-0.031	(1.00,1.00)
(% share)	[9.10,10.40]	[-1.76,1.24]	[0.13,4.48]	[-0.50,0.19]	353
Administration	14.359	-1.146	0.301	-0.127	(1.00,1.00)
(% share)	[13.42,15.40]	[-3.16,0.96]	[-1.94,2.83]	[-0.67,0.25]	353
Social Security	6.251	-0.061	-0.445	-0.003	(0.98,1.00)
(% share)	[5.59,7.10]	[-1.50,1.12]	[-1.63,0.60]	[-0.23,0.18]	347
Transportation	2.711	-0.794**	-1.089**	-0.095	(0.99,0.98)
(% share)	[2.28,3.21]	[-1.34,-0.19]	[-2.03,-0.19]	[-0.29,0.05]	344

Total budget estimated in log(Variable), except from the pre-treatment average. Significant at: 99% ***, 95% **, 90% *. Confidence intervals in square brackets are clustered by municipality. Optimal bandwidths are shown in parenthesis as standard deviations of the population and HDI averages of 16,000 and 0.07, respectively. Number of observations are shown below the bandwidths. The coefficients represent the average effect for the preferred frontier segment (6 bins). Regressions include year and state effects. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment. The list of included controls is described in the text. Column (1) shows the RDD effects of the discontinuity in health funding on political outcomes for the post-treatment period (2008,2012); Column (2) shows the same effects as (1), but for the pre-treatment period (2000,2004); Column (3) shows the IV regression for the post-treatment period (2008,2012).

Table A.IV: Sample Including also Low-Poverty Municipalities

	Coefficient, [90% CI]				Avg Band.
	Mean	RDD	RDD past	IV	(Pop. HDI)
	(Pre-Treat.)	(1)	(2)	(3)	Obs. per bin
Incumbent's vote share	50.524	-6.261**	1.680	-0.685**	(0.99,0.97)
(%)	[47.74,53.77]	[-11.07,-1.31]	[-2.36,6.05]	[-1.72,-0.15]	537
Margin of victory	17.236	-3.687	-1.463	-0.409	(1.00,1.00)
(p.p.)	[14.70,20.63]	[-9.10,3.06]	[-4.94,2.48]	[-1.32,0.31]	573
Candidates	2.356	0.370***	0.112	0.040***	(0.99,0.98)
(number)	[2.29,2.43]	[0.18,0.60]	[-0.06,0.31]	[0.02,0.10]	555
Pro-poor spending	52.277	2.512*	-1.834	0.326	(0.99,0.99)
(% share)	[50.32,52.82]	[0.18,5.72]	[-4.10,0.31]	[-0.08,1.87]	458
Challenger's entry	0.133	-0.029	-0.030	-0.003	(0.97,0.98)
(share without HS)	[0.10,0.20]	[-0.11,0.05]	[-0.12,0.06]	[-0.01,0.01]	994
Challenger's entry	0.379	-0.118*	-0.021	-0.013*	(0.97,0.96)
(share Clientelistic)	[0.34,0.43]	[-0.21,-0.01]	[-0.12,0.10]	[-0.03,0.00]	969
Challenger is top 2	0.142	-0.137***	0.047	-0.015**	(0.98,0.99)
(share without HS)	[0.10,0.19]	[-0.22,-0.06]	[-0.09,0.18]	[-0.03,-0.01]	579
Challenger is top 2	0.483	-0.105	-0.026	-0.011	(1.00,0.98)
(share Clientelistic)	[0.42,0.55]	[-0.25,0.05]	[-0.18,0.12]	[-0.04,0.01]	587

Significant at: 99% ***, 95% **, 90% *. Confidence intervals in square brackets are clustered by municipality. Optimal bandwidths are shown in parenthesis as standard deviations of the population and HDI averages of 16,000 and 0.07, respectively. Number of observations are shown below the bandwidths. The coefficients represent the average effect for the preferred frontier segment (6 bins). Regressions include year and state effects. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment. The list of included controls is described in the text. Column (1) shows the RDD effects of the discontinuity in health funding on political outcomes for the post-treatment period (2008,2012); Column (2) shows the same effects as (1), but for the pre-treatment period (2000,2004); Column (3) shows the IV regression for the post-treatment period (2008,2012).

Table A.V: RDD: Robustness to Kernel Choice

Dependent Variable	(1)	(2)	(3)
Health Funds	0.219***	0.215***	0.198***
(R$bn)	[0.14,0.29]	[0.13,0.29]	[0.10,0.30]
Obs. per bin	329	269	180
CCT Coverage	7.897***	7.645***	7.292***
(p.p. over target)	[3.84,11.60]	[3.60,11.27]	[3.32,10.91]
Obs. per bin	612	590	544
Incumbent's vote share	-7.750***	-7.435***	-7.118***
(%)	[-12.46,-3.66]	[-11.87,-3.51]	[-11.24,-3.27]
Obs. per bin	434	417	388
Margin of victory	-6.015**	-5.504**	-4.916**
(p.p.)	[-11.73,-1.77]	[-10.59,-1.45]	[-9.49,-1.15]
Obs. per bin	470	469	467
Candidates	0.392***	0.370***	0.339***
(number)	[0.20,0.63]	[0.18,0.59]	[0.15,0.54]
Obs. per bin	452	436	410
Pro-poor spending	4.022***	3.801***	3.703***
(% share)	[1.73,7.25]	[1.62,6.71]	[1.59,6.41]
Obs. per bin	350	346	319
Challenger's entry	-0.001	-0.008	-0.009
(share without HS)	[-0.08,0.09]	[-0.08,0.08]	[-0.09,0.08]
Obs. per bin	795	750	684
Challenger's entry	-0.138**	-0.137**	-0.128*
(share Clientelistic)	[-0.23,-0.02]	[-0.23,-0.02]	[-0.22,-0.02]
Obs. per bin	771	740	683
Challenger is top 2	-0.128**	-0.124**	-0.119**
(share without HS)	[-0.21,-0.04]	[-0.21,-0.03]	[-0.20,-0.03]
Obs. per bin	468	444	392
Challenger is top 2	-0.116	-0.107	-0.099
(share Clientelistic)	[-0.26,0.04]	[-0.24,0.05]	[-0.23,0.05]
Obs. per bin	472	453	420
Kernel	Epanech.	Gaussian	Uniform

Health Funds are estimated in log(Variable). Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality and shown in square brackets. The coefficients represent the average effect for the preferred frontier segment (6 bins), under the optimal bandwidth. Regressions include year and state effects. The list of included controls is described in the text.

Table A.VI: RDD: Other Robustness Tests

Dependent Variable	(1)	(2)	(3)
Incumbent's vote share	-7.631***	-7.881***	-7.919**
(%)	[-11.96,-3.04]	[-12.92,-3.37]	[-13.96,-2.14]
Obs. per bin	434	367	244
Margin of victory	-5.753**	-6.498**	-6.804**
(p.p.)	[-11.16,-0.96]	[-12.97,-1.73]	[-14.80,-1.09]
Obs. per bin	470	367	244
Candidates	0.345***	0.423***	0.460***
(number)	[0.14,0.59]	[0.22,0.68]	[0.22,0.77]
Obs. per bin	452	367	244
Pro-poor spending	2.938**	4.225***	4.810**
(% share)	[0.53,5.77]	[1.62,8.04]	[1.51,9.70]
Obs. per bin	350	275	179
Challenger's entry	-0.002	-0.003	0.004
(share without HS)	[-0.08,0.07]	[-0.08,0.09]	[-0.09,0.11]
Obs. per bin	795	673	463
Challenger's entry	-0.127**	-0.148**	-0.146*
(share Clientelistic)	[-0.22,-0.02]	[-0.25,-0.03]	[-0.27,-0.01]
Obs. per bin	771	673	463
Challenger is top 2	-0.129**	-0.144**	-0.164**
(share without HS)	[-0.21,-0.05]	[-0.23,-0.05]	[-0.27,-0.05]
Obs. per bin	468	385	258
Challenger is top 2	-0.118	-0.134	-0.140
(share Clientelistic)	[-0.26,0.04]	[-0.29,0.03]	[-0.34,0.05]
Obs. per bin	472	385	258
Controls	No	Yes	Yes
Bandwidth	Optimal	0.90	0.75

Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality and shown in square brackets. The coefficients represent the average effect for the preferred frontier segment (6 bins), under the optimal bandwidth. Regressions include year and state effects. The list of included controls is described in the text.

Table A.VII: Outcomes at Different FPM Thresholds

Dependent Variable	(1)	(2)
CCT Coverage	-2.998	-4.806
(pp over target)	[-7.00,0.77]	[-10.08,1.61]
Obs. per bin	1053	354
Incumbent's vote share	2.243	-0.596
(%)	[-1.90,6.26]	[-9.48,5.86]
Obs. per bin	744	249
Margin of victory	4.447*	-1.960
(p.p.)	[0.67,8.56]	[-15.21,7.63]
Obs. per bin	808	273
Candidates	-0.166	0.086
(number)	[-0.34,0.01]	[-0.17,0.35]
Obs. per bin	773	259
Pro-poor spending	0.533	-2.515
(% share)	[-1.61,2.16]	[-5.66,1.03]
Obs. per bin	591	203
Challenger's entry	0.014	-0.109
(share without HS)	[-0.08,0.11]	[-0.30,0.03]
Obs. per bin	1259	502
Challenger's entry	0.116*	0.223
(share Clientelistic)	[0.01,0.23]	[-0.02,0.51]
Obs. per bin	1207	485
Challenger is top 2	0.001	-0.130
(share without HS)	[-0.10,0.10]	[-0.38,0.05]
Obs. per bin	795	271
Challenger is top 2	0.120	0.112
(share Clientelistic)	[-0.03,0.27]	[-0.20,0.43]
Obs. per bin	800	271
Pop. cutoff	23,772	37,356

Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality and shown in square brackets. The coefficients represent the average effect for the preferred frontier segment (6 bins), under the optimal bandwidth. Regressions include year and state effects. The list of included controls is described in the text. For this table, treatment eligibility is solely determined by the population threshold.

Table A.VIII: Change in the Political Impact of the FHP post-2004

	Specification (A)		Specification (B)	
	FHP	FHP x Post-BF	FHP	FHP x Post-BF
Incumbent's vote share	3.755***	-2.786*	3.474***	-2.532
(%)	[0.857]	[1.524]	[0.899]	[2.278]
Obs	(7654)	(7654)	(7654)	(7654)
Pro-poor spending,	0.944***	0.098	0.704***	1.323**
(% share)	[0.236]	[0.386]	[0.238]	[0.572]
Obs	(7124)	(7124)	(7124)	(7124)
Challenger's entry	-0.030	-0.041	-0.033*	-0.078*
(share without HS)	[0.019]	[0.029]	[0.019]	[0.043]
Obs	(12158)	(12158)	(12158)	(12158)
Challenger's entry	-0.033	-0.041	-0.035*	-0.104**
(share Clientelistic)	[0.022]	[0.033]	[0.021]	[0.047]
Obs	(12158)	(12158)	(12158)	(12158)
Challenger is top 2	-3.332	-2.149	-1.953	-6.950
(share without HS)	[2.605]	[4.079]	[2.624]	[6.130]
Obs	(8027)	(8027)	(8027)	(8027)
Challenger is top 2	-1.263	2.435	-2.011	-7.630
(share Clientelistic)	[2.971]	[4.933]	[2.945]	[6.999]
Obs	(8027)	(8027)	(8027)	(8027)

Significant at: 99% ***, 95% **, 90% *. Specification (A) defines the variable FHP as a dummy indicating if the program has been implemented or not in the municipality. Specification (B) defines the variable FHP as a dummy indicating if the program covers at least 50% of the targeted population in the municipality, or not. In both cases, post-BF is a dummy indicating the years 2008 and 2012 (as opposed to 2000 and 2004). The sample includes only smaller municipalities (below 60 thousand) but it is not limited by poverty levels in order to allow enough variation in the FHP presence is 2008-12. All regressions include Municipality fixed-effects. The regression for the vote shares of the incumbent also includes the past vote share of the candidate as a control.

Table A.IX: Results for the Weak IV segment

	Coefficient, [90% CI]				Avg Band.
	Mean	RDD	RDD past	IV	(Pop. HDI)
	(Pre-Treat.)	(1)	(2)	(3)	Obs. per bin
Incumbent's vote share	49.030	0.798	0.190	0.147	(0.71,0.93)
(%)	[46.25,51.94]	[-2.27,4.30]	[-3.53,3.60]	[-0.87,1.68]	695
Margin of victory	16.819	-2.644	-1.965	-0.514	(1.00,0.87)
(p.p.)	[14.16,20.22]	[-6.63,1.24]	[-5.39,0.84]	[-3.97,0.34]	945
Candidates	2.307	-0.001	0.110	-0.001	(0.77,1.00)
(number)	[2.22,2.42]	[-0.12,0.12]	[-0.03,0.27]	[-0.10,0.06]	791
Pro-poor spending	49.149	0.875	-1.777*	-0.467*	(0.79,0.78)
(% share)	[47.68,50.21]	[-0.58,2.46]	[-3.22,-0.26]	[-73.80,-0.04]	614
Challenger's entry	0.244	0.039	-0.006	0.007	(0.90,1.00)
(share without HS)	[0.17,0.30]	[-0.04,0.12]	[-0.09,0.07]	[-0.01,0.04]	1419
Challenger's entry	0.400	-0.038	-0.138***	-0.006	(0.84,1.00)
(share Clientelistic)	[0.33,0.47]	[-0.13,0.05]	[-0.21,-0.05]	[-0.04,0.02]	1291
Challenger is top 2	0.307	-0.055	-0.042	-0.011	(1.00,1.00)
(share without HS)	[0.24,0.38]	[-0.14,0.04]	[-0.14,0.06]	[-0.06,0.01]	1105
Challenger is top 2	0.462	0.025	-0.232***	0.006	(1.00,1.00)
(share Clientelistic)	[0.39,0.54]	[-0.08,0.13]	[-0.31,-0.14]	[-0.03,0.06]	1107

Significant at: 99% ***, 95% **, 90% *. Confidence intervals in square brackets are clustered by municipality. Optimal bandwidths are shown in parenthesis as standard deviations of the population and HDI averages of 16,000 and 0.07, respectively. Number of observations are shown below the bandwidths. The coefficients represent the average effect for the preferred frontier segment (6 bins). Regressions include year and state effects. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment. The list of included controls is described in the text. Column (1) shows the RDD effects of the discontinuity in health funding on political outcomes for the post-treatment period (2008,2012); Column (2) shows the same effects as (1), but for the pre-treatment period (2000,2004); Column (3) shows the IV regression for the post-treatment period (2008,2012).

Table A.X: Health Outcomes at Different Frontier Segments and Periods

	Mean	Coefficient, [90% CI]	
	[90% CI]	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]
Obs per bin		{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]
Obs per bin		{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]
Obs per bin		{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]
Obs per bin		{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]
Obs per bin		{470}	{529}
Pre-Natals	66.27	0.49	3.20
(% of pregnancies)	[63.20,69.26]	[-5.01,6.19]	[-3.12,10.17]
Obs per bin		{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]
Obs per bin		{476}	{542}

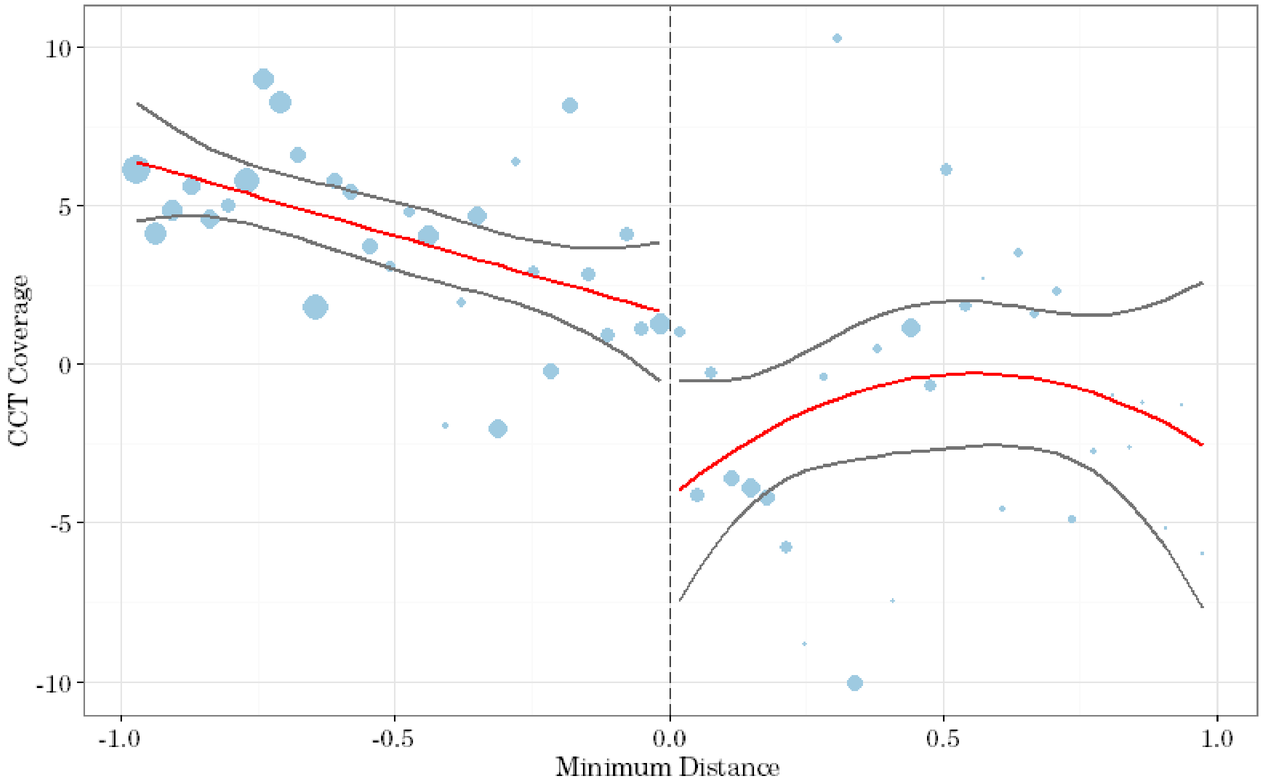
[a] log (Variable). Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality and shown in square brackets. The coefficients represent the average effect for the preferred frontier segment (6 bins). All regressions include year and state effects and are estimated using the edge kernel and a bandwidth of 0.9 standard deviations of population and HDI. The list of controls is in the text. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment.

Table A.XI: Hiring Outcomes

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

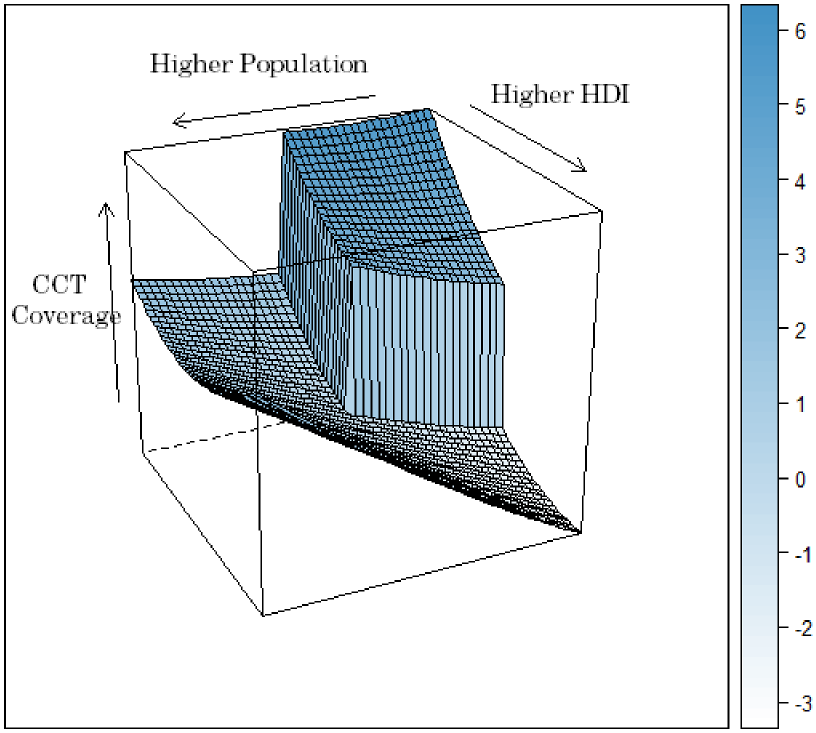
[a] log(Variable). Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality and shown in square brackets. The coefficients represent the average effect for the preferred frontier segment (6 bins). All regressions include year and state effects and are estimated using the edge kernel and a bandwidth of 0.9 standard deviations of population and HDI. The list of controls is in the text. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment.

Figure A.I: CCT Coverage: One-dimension RDD



The vertical line represents the treatment frontier (Pop = 30,000 and HDI = 0.7). The blue dots represent 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, measured by the lowest value of population and HDI (population and HDI are normalized to a common scale by their standard deviations). The red lines are fitted by local linear regression on the unbined data. The grey lines are the 90% confidence level. The sample used includes the entire treatment frontier and all observations within 1 distance unit from treatment.

Figure A.II: CCT Coverage: Three Dimensional View of The Parametric Fit



The raised area represents observations eligible to treatment (Pop <= 30,000 and HDI <= 0.7). The three dimensional surface is fitted regressing CCT coverage on a quadratic polinomial on population and HDI. The sample used includes the entire treatment frontier and all observations within 2 distance units from treatment. Population and HDI are normalized to a common scale by their standard deviations.