Journal of Public Economics 176 (2019) 1–17

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Cash transfers, clientelism, and political enfranchisement: Evidence from Brazil☆

Anderson Frey

Department of Political Science, University of Rochester, Harkness Hall, 320B, Rochester, NY 14627, United States of America

ARTICLE INFO

Article history:
Received 9 May 2018
Received in revised form 8 January 2019
Accepted 9 May 2019
Available online 6 June 2019

Keywords:
Political economy
Clientelism
Redistribution
Cash transfers

ABSTRACT

This paper uses Brazil's *Bolsa Família* to show that redistributive policies that are shielded from the influence of political intermediaries can reduce incumbency advantage for mayors, 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 clientelism a less attractive strategy to incumbent mayors. Consequently, incumbents reallocate effort away from the practice into public good distribution.

1. Introduction

Political leaders in many developing democracies are able to reap electoral rewards by replacing public good distribution with private transfers targeting 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 electoral competition, perpetuation of political machines, and inefficient public good provision.[3] Despite the established correlation between clientelism and poverty, the causal channels linking redistribution to a reduction in the practice remain vastly unexplored.[4]

Can redistributive policies reduce clientelism? The answer depends on their institutional design. Whenever the policy can be subject to capture by politicians insomuch that the poor sees the benefits as contingent on electoral support, it is likely to perpetuate the existing structures of political power and clientelism.[5] However, effective redistributive policies could also have an indirect 'income' effect on political behavior, as permanent reductions in poverty can break the voter's dependence on clientelism. This income effect is more likely to prevail, and therefore reduce the power of dominant political elites, when the policy is designed to shield the resources from capture.

Accordingly, this paper examines this income effect of policy in the context of municipal leadership selection and public good distribution in Brazil, using the largest conditional cash transfers (CCT) program in the world, *Bolsa Família* (BF). CCT programs are one of the main poverty alleviation policies adopted by developing nations. For the most part, they are designed to provide a stable income for the extremely poor, and to prevent capture by political intermediaries.[6] 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

☆ 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 Vancouver School of Economics, Vanderbilt, Rochester, York, Ottawa and Harris (Chicago) for the comments and suggestions. All errors are my own.

E-mail address: anderson.frey@rochester.edu.

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 et al. (2004), Vicente and Wantchekon (2009), Fujiwara and Wantchekon (2013), Anderson et al. (2015), Cruz et al. (2018) and Bobonis et al. (2017).

3 See reviews in Boix and Stokes (2009) and Hicken (2011).

4 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.

5 The literature has several examples of politicians claiming credit for redistributive policies, and electorally benefiting from them. Cruz and Schneider (2017) study international aid in the Philippines. Others examine the implementation of pro-poor transfers in Uruguay (Manacorda et al., 2011), Colombia (Baez et al., 2012), Mexico (De La O, 2013), and Brazil (Zucco, 2013).

6 De La Oh (2015) discusses the potential impact of CCT design on clientelism.

https://doi.org/10.1016/j.jpubeco.2019.05.002
0047-2727/

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.[7]

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. CCT reduces the incumbency advantage for mayors, increases both electoral competition and the quality of candidates, weakens support for candidates from parties associated with clientelism, and prompts an increase in health care and education spending.[8]

In order to show how these multiple results are reconciled in a clientelistic environment, I propose a probabilistic voting model where the incumbent mayor chooses to allocate effort to either implementation of pro-poor public goods or clientelism. Clientelism is viable when a share of the poor population becomes extremely vulnerable due to negative income shocks. Uncertainty about the willingness of challengers to engage in clientelism 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 clientelism becomes less attractive, incumbents reallocate effort towards public goods.

The main empirical challenge here is to observe only this indirect, clientelism-reducing, income effect of the policy. For this, municipal CCT coverage needs to be exogenous to the influence of local politicians. BF was created and financed by the federal government, and it has been aggressively marketed as so. This, combined with an environment of candidate-driven elections and weak party-voter linkages, limits the ability of mayors to locally claim credit for the policy benefits. On the other hand, BF's enrollment is jointly run by the central and municipal governments, which gives mayors some leeway to interfere in the selection of beneficiaries. This means that municipal CCT coverage levels could be at least partially endogenous to local political manipulation.

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, and given that the discontinuity is determined by two variables, using a multivariate regression discontinuity design (MRDD). Under similar assumptions to the ones in the single-score RDD, this methodology also generates a quasi-experimental assignment of FHP resources to municipalities.

This identification strategy relies on the FHP having a strong impact on 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 an information channel. Access to information about BF has been key to the enrollment of potential beneficiaries, given that household registration is primarily self-directed. Where municipalities have few resources to promote enrollment, the FHP teams have sufficient capillarity and penetration among the poor to disseminate the information about the policy. A survey with 10,000 poor households in Brazil shows that >10% of BF beneficiaries came to know about the program through their family doctor. In this context, a 50% funding increase for the FHP program was responsible for a significant boost in BF enrollment in the affected municipalities.

Although the exclusion restriction cannot be tested directly, this paper uses multiple falsification tests to provide strong evidence that the FHP affects political outcomes almost exclusively through CCT coverage. The basic mechanics behind these falsification tests relies on the heterogeneity of the first stage estimates, either over time or across municipalities, to simulate a counterfactual world where the FHP program has no effect on CCTs, allowing me to assess its direct effects on politics.

The first class of tests relies on the heterogeneity of the first stage estimates across municipalities. The intuition is that the information channel proposed here does not operate homogenously for all municipalities in the sample, which allows me to observe municipalities where only the direct effects of FHP on politics are operating. This is primarily done with the help of the MRDD method. Contrary to the single-score RDD, the MRDD identifies treatment effects for a frontier of points on a two-variable space (population and HDI), which allows the researcher to systematically observe and assess effect heterogeneity. Using this approach, this paper shows that the reduced-form effects on political outcomes are only statistically significant for parts of the sample where the first stage is also systematically present. This suggests that potential direct FHP effects on politics are not a likely confounding in the main estimation results.

Another such test is conducted by estimating the regressions separately for subsamples with high and low CCT coverage pre-treatment, i.e., before the FHP discontinuity was implemented. As expected, the first stage is much weaker in magnitude for the group of municipalities that were already well covered by the CCT program in 2004. Accordingly, the reduced-form effects on the multiple political outcomes are also much weaker for this sample. Even though this test does not directly rule out potential direct FHP effects, it strongly suggests that this paper's results are indeed a consequence of increases in CCT coverage in municipalities.

The second class of falsification tests relies on examining the effects of FHP on political outcomes in the periods before and after the arrival of BF. Using panel data, the paper estimates the effects of both FHP coverage and presence on political outcomes between 2000 and 2012, focusing on how they change after 2004 (BF started in 2003). The intuition for this exercise is that, even if the estimation in the pre-BF period shows that FHP has a direct impact on politics, significant post-BF changes should provide evidence that CCT coverage is indeed a channel connecting FHP and political outcomes. Accordingly, this paper finds that the effect of FHP in several political outcomes changes significantly after BF arrives, and in the exact same direction of the coefficients estimated with the MRDD method in the main specification.

All-in, 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 10 pp increase in CCT coverage triggers a 8.2 pp vote loss for the incumbent, which is mostly driven by the effect on less educated politicians or candidates that belong to clientelistic parties. While incumbents from clientelistic parties are more likely to lose votes, clientelistic challengers become less likely to even enter mayoral races. Higher CCT coverage also prompts a 6.3 pp 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 4.4 pp (nearly 9%). These results are robust to the choice of kernel and bandwidth in the nonparametric estimation. What is more, the magnitude and direction of the estimated coefficients indicate that these results are not mechanically driven by the 50% FHP funding increase alone. The provision of health services increases seven times more than what would be expected in the absence of the CCT mechanism described here.

This paper also argues that the thread that connects all the empirical results is a reduction in clientelism. One potential alternative

[7] 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). 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).

[8] 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 towards these categories can be seen as redistributive (Fujiwara, 2015).

interpretation is that CCTs generate a shift in electoral support towards left-wing parties, given that the success of BF in Brazil is widely attributed to President Lula (2003−2010) and the leftist party PT (Zucco, 2013). As the argument goes, mayors from PT or other leftist parties might be able to claim 'undeserved' credit for the policy arrival, even if they cannot influence selection of beneficiaries, by trying to associate themselves to Lula. Nevertheless, the results provide enough evidence to rule out this explanation. When parties are aggregated into left- and right-wing groups, the results show that politicians (both incumbents and challengers) from leftist parties are no more likely to lose votes than their right-wing counterparts, and also equally likely to shift effort towards pro-poor public goods. Similar results are observed when the performance of PT candidates is compared to the ones from either left or right wing parties. In short, heterogeneous effects are only observed when parties are categorized by their historical association with clientelism.[9]

This work contributes mainly to three strands of the literature in both economics and political science. First, it adds to the extensive research on the inner workings of clientelistic politics around the world. For example, Finan and Schechter (2012) emphasize the role of reciprocity in the way voters are targeted for vote buying in Peru. Anderson et al. (2015) show how land ownership and cast relations make clientelism viable in India, and Brusco et al. (2004) discuss the role of risk aversion in vote buying in Argentina. Focusing on how information affects the effectiveness of vote buying, Fujiwara and Wantchekon (2013) and Cruz et al. (2018) also examine the result of experiments in Benin and the Philippines, respectively. The key finding in this paper is that redistributive policies can be an effective tool to weaken clientelistic ties, and replace the practice with more programmatic forms of redistribution.

Second, the present results are in stark contrast with those of the extant work on the political effects of CCTs around the world, which typically show that the policy increases support for incumbents. The difference lies in the fact that this literature examines the question in environments where politicians can easily capture the policy, and 'credit claiming' effectively overshadows CCT's indirect income effects. For example, a number of papers examine how CCTs affect the support for the central government that actually creates the policy. Not surprisingly, incumbents benefit from credit claiming in this context, as shown in Uruguay (Manacorda et al., 2011), Colombia (Baez et al., 2012), Mexico (De La O, 2013), and Brazil (Zucco, 2013).

Only a few papers focus on local politics. Labonne (2013) examines the impact of CCT on mayoral incumbency advantage in the Philippines.[10] Here again, the analysis is built on the assumption that mayors can claim full credit for the arrival of transfers in these villages. As the argument goes, mayors benefit from the extra funds brought by the policy by targeting budget resources to other voters. De Janvry et al. (2012) focus on *Bolsa Escola* (BE), a precursor to BF in Brazil. The striking difference between these programs is that mayors were in charge of BE beneficiary selection, which allowed for ample credit claiming, as voters likely saw the benefits as contingent. In contrast, this paper is only concerned with the CCT coverage that is exogenous to mayors, so as to observe the policy's income effects. 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 credit claiming and capture are limited.

Finally, by extending the nonparametric method 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 et al., 2013).

The remainder of the paper is structured as follows: Section 2 has a brief overview of the institutional background; Section 3 presents the theory; Section 4 introduces identification strategy; Section 5 presents the empirical strategy for the MRDD, also describing the data and construction of variables; and Section 6 discusses the results and the multiple robustness and falsification tests, followed by the conclusion.

2. Institutional background

Bolsa Família (BF). This 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.[11] 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.[12] The former determines major guidelines,[13] 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 the families.

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, weak voter-party linkages, 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.[14] 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. Most revenues for small municipalities come in the form of transfers from higher levels of government; and clientelism is abundant (Alston and Mueller, 2006; Fried, 2012). Given that mayors have significant control over the budget allocation (Ferraz and Finan, 2011), clientelism is often financed through public funds and services. 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.

[9] Section 5.2 shows how parties are labeled programmatic or clientelistic based on the DALP survey.

[10] Contrary to this paper, Labonne (2013) does not focus on other variables such as the quality of politicians, or public goods distribution.

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

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

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

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

Surveys indicate that its beneficiaries perceive BF as being more resistant to local political manipulation than other government programs (Sugiyama and Hunter, 2013). BF funds represent the second most important source of federal government transfers to municipalities, comprising >12% of the total transfers. 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, it represents nearly 5% of the local budget in small municipalities.

The Family Health Program (FHP). The FHP was created by the Ministry of Health (MH) in 1994. It finances teams of health professionals that regularly visit households to provide basic care.[15] Each team is responsible for a geographic area and serves a population of up to 4000 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 effectively provides services to the poor population (Fujiwara, 2015).

The identification strategy in this paper 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.[16] These payments were uniform across the country until August 2004, when municipalities with population below 30,000, and HDI below 0.70,[17] started to receive an extra 50% funding per team (see the timeline of events in Fig. 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.[18]

3. Theoretical framework

I propose a probabilistic voting model where incumbents allocate effort to either the implementation pro-poor public goods or targeted transfers that finance the voter's private consumption (clientelism).[19] Voters have convex preferences over public goods and clientelistic transfers. CCTs increase the income of voters, therefore reducing the relative marginal utility of targeted transfers, and consequently the voters' susceptibility to clientelism. This makes public goods a more attractive strategy to reelection seeking politicians.

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 redistribution, which is allocated between public goods (e) or clientelism ($E - e$). The effort allocation cost is given by κe,[20] and applies every period in which e changes, i.e., reelected incumbents face $\kappa = 0$ if they keep effort unchanged.

Politicians differ in two dimensions: their willingness to exercise programmatic effort, and their ability to do so, which is measured by $\theta \in [0,1]$. Extreme politicians (type x) never do clientelism ($e = E$ and $\theta = \theta$). Opportunistic politicians allocate 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 $\theta = \theta$. Programmatic types (type p) have $\theta = \theta$, where $\theta > \theta$. 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 a 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, opportunities for clientelistic exchanges come from idiosyncratic income shocks to poor households that make them susceptible to vote buying. Let $\delta(y)$ be the share of the poor subject to such shocks at every period. This is a random draw from the poor population, and increases as income falls ($\delta_y < 0$). Poor citizens also consider the prospective utility under each candidate in period 2 when voting, which is quasi-linear and given by $U^P(g,c) = \theta^P \log(e) + \delta(y)(E - e)$, where P refers to the politician that wins the election (incumbent (I) or challenger (C)). The utility coming from pro-poor public goods is given by $\log(e)$, and the utility coming from consumption of clientelistic transfers is given by $(E - e)$.

The effort allocation is not perfectly observable by voters, so they cannot 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. If y is low enough, there is always a significant share of the poor that would find some amount of clientelism optimal, and prefer not to elect an extreme type. This is the source of incumbency advantage of opportunistic politicians: 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 e is the same as it was in period 1.[21] As for the challenger, with positive probability (μ_x) an elected challenger is extreme and provides utility equal to $g(E)$ in period 2. With $(1 - \mu_x)$ probability, the challenger is opportunistic and behaves like an incumbent seeking reelection, providing $\theta \log(e) + \delta(y)(E - e)$. Thus, the gap in expected voter's utility from electing an incumbent as opposed to a challenger is: $\mu_x[\theta \log(e) + \delta(y)(E - e) - \theta \log(E)]$.[22]

The timing of events is as follows: (1) an incumbent is elected for her first tenure in office, with the possibility of reelection; (2) with probability $\delta(y)$, poor voters receive a negative shock that makes them vulnerable; (3) incumbent allocates effort e, incurring cost κ; and (4) elections happen. If the incumbent wins, she enjoys the rents of office in period 2 and minimizes allocation cost κ, which implies that effort remains unchanged from period 1. If the challenger wins, he behaves like a first period incumbent. Thus, poor citizens vote for the incumbent whenever μ_x

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

16 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.

17 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.

18 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 N° 1.434/GM, July 14, 2004.

19 The mechanism follows the traits of vote buying in Brazil, described by surveys and the press. It characteristically involves the use of public resources and targeting of the poor. 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.

20 If $\kappa > 0$ ($\kappa < 0$).

21 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.

22 $U^I(g,c) - U^C(g,c)$.

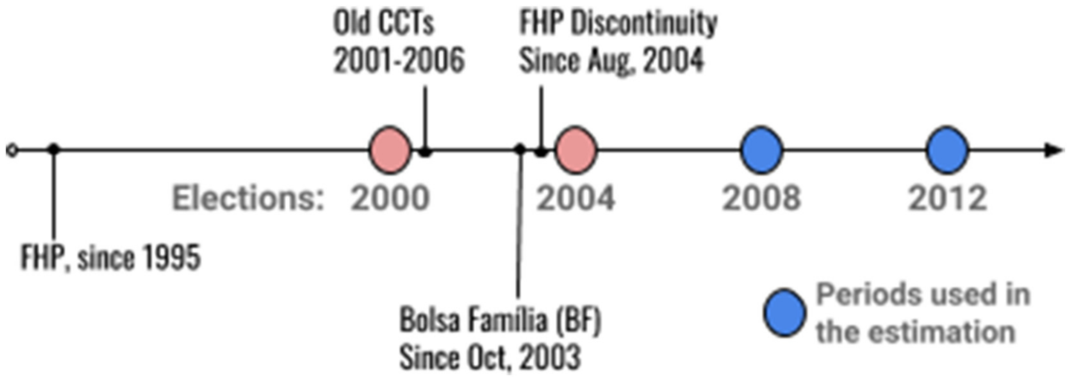

Fig. I. Timeline of events. The FHP started in 1995. Bolsa Família was created in Oct 2003, and the FHP funding discontinuity in Aug 2004. All elections happened in early October.

$[\theta \log(e) + \delta(y)(E - e) - \theta \log(E)] > \sigma_i + \lambda$, in which case the incumbent's vote share is:

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

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.

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

This article focuses on how the equilibrium quantities respond to a permanent change in income, which in this context comes from CCTs. The relevant comparative statics are summarized below, and all proofs are shown in the online appendix. Section 5.3 details how each these propositions can be tested using the data described in Section 5.2, and the empirical strategy in Section 5. Online Appendix I also discuss some welfare implications of CCTs in the context of this model.

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.

4. Identification of CCT effects on political outcomes

Bolsa Família was first implemented in late 2003 and has covered nearly all its targeted population since 2006. Nevertheless, there is a persistent and significant variation in program coverage across municipalities. In 2012, the average number of CCT benefits exceeded the number of poor families by 10 percentage points (pp), with a standard deviation of 16pp (Fig. II).[23]

This paper uses this coverage variation across municipalities to estimate the effects of CCT on local political variables, using a discontinuity in the funding for the Family Health Program (FHP) as an instrument for CCT coverage. The argument is straightforward: Health care professionals are one of the most important sources of information about government benefits in Brazil, including CCTs, due to the capillarity and penetration of the FHP program within poor communities. Given that BF requires eligible households to self-enroll, and for that they need information, differential FHP funds triggered variation in BF enrollment across municipalities by increasing the ability of health professionals to reach more households, more frequently.

This information channel was key to generating coverage differentials, especially in the early years of the BF program. While the smaller, older CCT programs had at most 6 million beneficiaries in 2002 (Zucco, 2013),[24] BF's global coverage target was 11.1 million families. This could only be achieved with a concerted effort to reach new households, in which health professionals played a significant role. Anecdotal evidence comes from a 2009 survey (AIBF - *Avaliao de Impacto do Bolsa Família*) including >10,000 CCT-eligible households. The survey shows that health agents asked households about their coverage status in 50% of their visits. In 12% of the households surveyed, the information about BF came first from a health professional.[25] What is more, another 9% of the respondents stated that they would prefer to speak to a health

[23] Benefits include *Bolsa Famlia*, *Bolsa Escola*, *Bolsa Alimentao* and *Carto Alimentao*. The number of poor families comes from the MDS estimates (see the Data and description of variables for details). The 2004 coverage in the plot is adjusted to account for 974,000 households receiving both *Bolsa Famlia* and *Bolsa Escola* at the same time, as disclosed to the press by MDS. The elimination of this double-counting increases the coverage gap. Accordingly, 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 Famlia, so the coverage is measured precisely.

[24] 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.

[25] 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.

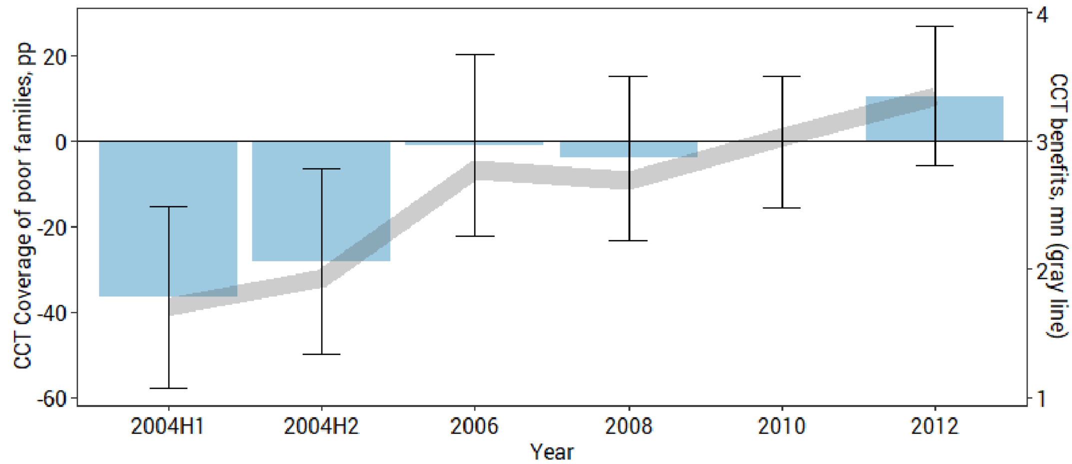

Fig. II. CCT benefits vs. number of poor families. The left y-axis shows the ratio between CCT benefits and the number of poor families in each municipality, in percentage points (the bar plot). The vertical lines show the standard deviation of the coverage gap across municipalities, also in percentage points. The right y-axis shows the number of CCT benefits paid each year (the gray line). All variables are calculated for municipalities in the potential sample of this paper, which excludes large metropolitan areas.

professional when it came to questions about BF, instead of other local officials.

Fig. II shows that BF reached nearly full coverage by the end of 2006. Nevertheless, the program continued to expand in 2006–2012, albeit at a lower rate (gray line in the same figure). This later expansion was a consequence of successive increases in the global number of benefits granted.[26] Given that the FHP funding discontinuity remained in place during this entire period, it contributed further to the uneven spread of information, and also the coverage gap. What is more, the fact that households seldom left the program in this period makes these coverage differentials very persistent. Although beneficiaries could lose the benefit if they did not biannually update their income, this rule was not properly enforced prior to 2009. After 2009, MDS allowed households that surpassed the income threshold to receive the benefit for two additional years.[27]

4.1. IV assumptions

The identification of treatment effects in every instrumental variable (IV) strategy relies on three major assumptions: (1) exogenous assignment of the instrument across subjects; (2) a strong first stage; and (3) the exclusion restriction. As for the first assumption, given that the FHP funding across municipalities is a discontinuous function of both municipal population and HDI, 'quasi-random' assignment of this variable can be achieved by employing a multivariate regression discontinuity design (MRDD). The quality of this assumption is then assessed with the help of the usual tests employed in RD designs, such as the balance of covariates at the discontinuity, or the robustness of estimates to bandwidth and kernel choice, and inclusion of covariates. This process is thoroughly explained in Section 5, and results are shown in Section 6.

In the specific case of RD designs, 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, and one of the population thresholds (30,564) is close to the 30,000 mark in this paper. The implications of this fact are thoroughly discussed in Section 6.1. In short, it seems unlikely that the FPM is driving the results found in this paper, given that the estimated political effects are observed neither for the 30,000 threshold before our identification strategy existed (electoral cycles of 2000 and 2004) nor for FPM discontinuities at neighbor population thresholds during the period under analysis (2008 and 2012).

The quality of the first stage assumption is easily assessed empirically. A strong first stage requires the instrument (FHP funding) to explain a significant part of the variation in the endogenous regressor (CCT coverage). Empirical evidence for the first stage's pattern of heterogeneity across the sample is shown in Fig. V, and widely discussed in Section 6. Table A.XI in the online appendix also shows that the extra FHP funds significantly increased the number of families visited by local health professionals in treated localities.

The exclusion restriction assumption requires that the estimates for political outcomes are not a direct effect of the instrument. Contrary to the first stage assumption, it cannot be directly assessed in the context of this identification strategy. This paper uses a combination of falsification tests to address this issue, which are detailed in Section 6.1. The first piece of evidence supporting the instrument's excludability comes from the heterogeneity in the results. The intuition is that, if the information channel is absent for some municipalities in the sample, the researcher can whether direct effects of FHP on politics exist, and therefore evaluate the validity of the excludability assumption. The MRDD allows me to systematically observe how the results vary along different values of population and HDI for municipalities (details in Section 5), and check whether the heterogeneity pattern in the first stage estimates matches the one in the reduced form ones.

To illustrate this, take a group of municipalities in which most poor households already had information about CCTs before 2004. In these locations, the mechanism described before was nonexistent, and the FHP discontinuity should have had no effect on CCT coverage (no first stage). It is easy to see that, in this sample, the reduced-form estimates should only pick up the direct effects of FHP funding on political outcomes. Accordingly, I find that significant effects on political outcomes are only found for parts of the sample that also present a strong first stage.

A similar test is performed with the subset of municipalities that had very high CCT coverage pre-2004. Given that the FHP funding had limited room to increase BF enrollment in these

[26] The first global coverage target for BF was 11.1 million families, effective between 2004 and 2009. It was changed in 2009, and again in 2011, to finally reach >13 million families.

[27] This is called the 'permanence rule'. It was created by the Portaria MDS No 617, from August 11, 2010.

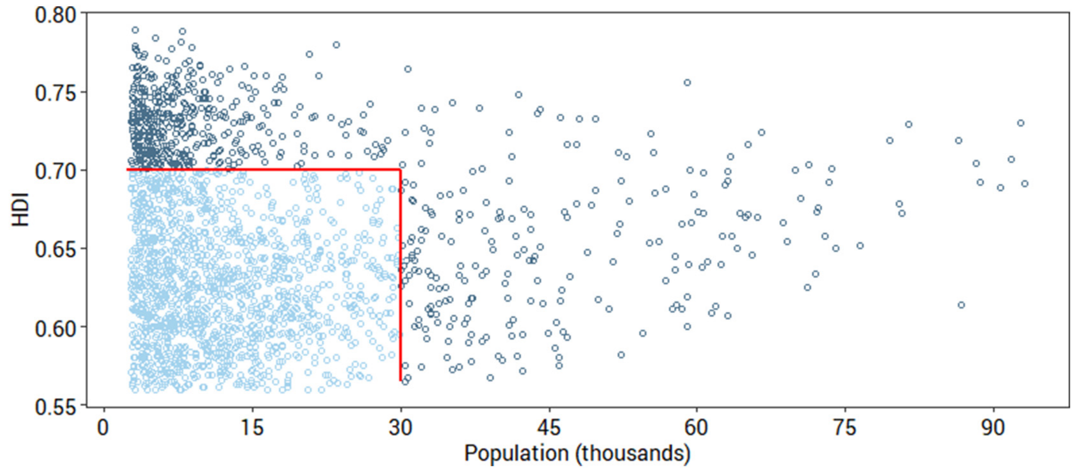

Fig. III. Potential sample and treatment frontier. The potential sample includes all municipalities within the central 95% percentile in population and HDI. The treatment frontier is the line. Light colored dots represent municipalities eligible to treatment.

places, any effects on political outcomes are more likely to be a direct consequence of the FHP program. Again, the multiple political effects are much weaker, or even non-existent, for this subsample.

The second class of falsification tests exploits the effects of the FHP program before the arrival of CCTs. For this, I use panel data on both the introduction and coverage levels of the FHP program during the four electoral cycles in 2000–2012.[28] Using municipality and time fixed effects, I regress the political outcomes on a variable measuring the coverage of FHP, and its interaction with a dummy indicating the existence of the CCT program. This regression allows us to observe the direct effects of the program on political outcomes pre-CCT, but more importantly, how this effect changes when CCT arrives.

As an illustration, consider the variable measuring the incumbent's vote share. In the period between 2000 and 2004, high FHP coverage is (weakly) associated with the incumbent having a higher vote share. This is not surprising given that the FHP is the best-rated government program run by municipalities in Brazil (IPEA, 2011). Post-2004, when CCT is already present, this coefficient falls significantly, in line with the results in this paper. Similarly, for other political outcomes, the FHP effects also significantly change after the introduction of CCT, and in the same direction of the effects observed in the core MRDD analysis in this paper.

Section 6.1 has a more detailed discussion on the construction of these tests and their results, and also comments on the main threats to the validity of the exclusion restriction.

5. Empirical strategy

This section presents this paper's empirical strategy, emphasizing how the use a multivariate regression discontinuity design (MRDD) contributes to a more effective and transparent identification of the political effects of CCTs. I also describe the data and the variables used in the estimation, and how the parameters estimated under this empirical design can be mapped to the theoretical predictions from Section 3.

5.1. The multivariate regression discontinuity design

Single score RDDs have been widely explored in recent 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 et al., 2011; Keele and Titiunik, 2015) or test scores (Jacob and Lefgren, 2004; Papay et al., 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 the case of this paper, a municipality m with population p_m, and HDI denoted by h_m, has the ATE defined over the frontier: $F = (p_m, h_m) : (p_m \leq 30, h_m = 0.7) \cup (p_m = 30, h_m \leq 0.7)$, with respective treatment cutoffs for population nd HDI at 30,000 and 0.7 (Fig. III).

The literature on identification and estimation of MRDD effects lacks consensus on a definitive strategy.[29] Researchers often approach 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 et al., 2013), or by collapsing them into a single score, which 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 approach, however, is more compelling when the variables are on the same scale as test scores or spatial coordinates. Another approach is to impose a parametric function over the two scores.[30] Dell (2010) provides an application of this strategy by using a cubic polynomial in a geographical MRDD.[31] In both these strategies, 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 (Fig. III).

Heterogeneity and the identification strategy. There are two very good reasons to explore how treatment effects vary along the frontier in the context of this paper. First, the two scores have a distinct nature, and under the former approaches the sub-populations being compared along the frontier might differ considerably, defeating the spirit of the RDD.[32] Second, the causal estimation of the political effects depends

[28] Unfortunately, the discontinuous FHP funding cannot be used for this analysis, as the discontinuity was only created in 2004 a few months after BF (Fig. I).

[29] Reardon and Robinson (2010) and Wong et al. (2013) review potential estimation strategies, and Papay et al. (2011) propose a framework to estimate the ATEs nonparametrically when there are multiple treatments. All the proposed methodologies focus on the average effects, without emphasis on their heterogeneity along the frontier.

[30] In this parametric approach, the ATE along the frontier can only be inferred if the polynomial on the scores is not interacted with the treatment dummy (spline). With 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.

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

[32] A municipality with 30,000 population and HDI below 0.60 might be compared to one with 3000 population and 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.

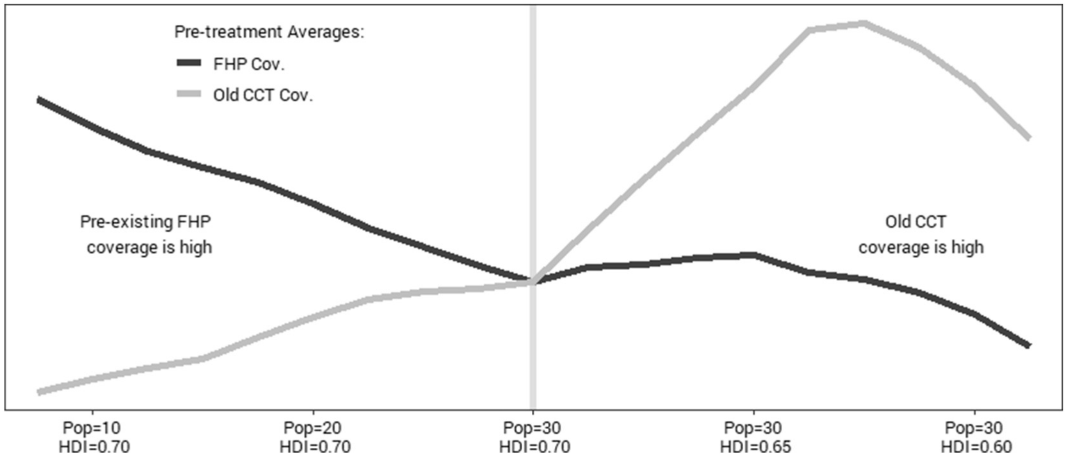

Fig. IV. Heterogeneity of pre-existing conditions along the treatment frontier. The lines plot the pre-treatment average of the variables for points along the treatment frontier. The left side reflects the horizontal part of the frontier where HDI is fixed at 0.7, and the population varies. The right side reflects the vertical part of the frontier where population is fixed at 30,000, and HDI varies. Variables are normalized by standard deviation for the purpose of the presentation.

on the existence of a strong instrument for CCT coverage, which is only found by the way of exploiting the heterogeneity in municipal characteristics along the population × HDI frontier. This is also the source of the main falsification test used to assess the exclusion restriction assumption in this paper.

Accordingly, the information channel that drives the effect of FHP on CCT coverage (Section 4) would only be observed in a municipality under two conditions: (1) the poor population did not already have adequate information about *Bolsa Família* before treatment; and (2) the 50% extra FHP funding significantly changed the penetration and capillarity of local health services, effectively increasing the spread of CCT information. I systematically identify municipalities that meet these conditions using the pre-treatment average of two variables along the frontier.

First, the penetration of old CCT programs. The higher the coverage of these programs, the more likely that the local poor population had ample information about BF beforehand. Most targeted households in these locations just had to be 'migrated' to BF, as opposed to newly enrolled. Second, the pre-existing level of FHP services. The better the FHP coverage pre-treatment, the lesser the impact of new, discontinuous FHP funding on the ability of doctors to see more patients, and spread the information about BF.

Fig. IV illustrates how this works. The profile of municipalities is depicted along the x-axis. In the extreme left, both treated and non-treated municipalities are small in population, but have relatively high economic development (HDI around 0.70). Not surprisingly, these locations had very high FHP coverage before treatment, on average. In the other extreme, all municipalities have around 30,000 population, but are underdeveloped with low HDIs. This group was the main target of old CCT programs, and had a large share of its poor population already enrolled in old CCTs before 2004. The combination of these two variables hence predicts that the FHP funding is a plausible instrument for new CCT coverage only for municipalities in the central region of the frontier, which is later confirmed by the empirical results in Section 6.

To explore this heterogeneity within the confines of the available data, this paper follows the general estimation approach in Zajonc (2012): the conditional ATE (CATE) is estimated for several points of the treatment frontier, and the average effect for any segment is derived by averaging these CATEs. Thus, the heterogeneity along the frontier becomes fully observable. For population p_m and HDI h_m, the CATE (τ_{Con}) at a point (p,h) in the frontier is given by Eq. (3):

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

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.[33]

First stage. Following the procedure for the similar, single-score RDD (Imbens and Lemieux, 2008; Imbens and Kalyanaraman, 2012), the CATE is estimated nonparametrically by local linear regression for any frontier point (p,h), using the two scores as dependentvariables. In municipality m at period t, CCT coverage is defined as C_{mt}, and the first stage estimation is given by Eq. (4).

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

where the treatment effect is denoted by α_1, and $\delta_m = 1[(p_m \le 30, h_m \le 0.7]$. The values of population and HDI centered around point (p,h) are denoted by (p_m^c, h_m^c). The regression includes state effects (γ_s), a period dummy (η_t), and a vector of municipal controls (θ_m).[34] This is usual in RDDs to reduce the sample variability (Lee and Lemieux, 2010). The local linear regression is weighted by the edge kernel. Online Appendix I shows that the results are robust to the kernel choice, the inclusion of controls, and state and period effects.[35]

I run this regression for a total of 19 frontier points, limiting the data on each regression to a bandwidth defined over the two score variables.[36] For a simpler interpretation of the bandwidths, I normalize the score variables to the same scale dividing them by their standard deviation. The average effect for any frontier segment (τ_{avg}) is estimated as

[33] 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.

[34] 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 also include the share of votes in the past election, and a dummy indicating if the candidate belongs to PT. 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. For the budget shares, the past share of the budget (1997–2000 tenure) is also included as a control.

[35] All kernels are two-dimensional, defined as the following product: $K(u_1,u_2) = [K(u_1) \cdot K(u_2)]$.

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

the average of CATEs for k points along the frontier, weighted by the joint density $\lambda(p_k,h_k)$ of each point (Eq. (5)).

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

Standard errors of the averaged coefficients are bootstrapped with a sample of 5000, and confidence intervals are calculated using the bias corrected and accelerated method (Efron, 1979). I present results estimated under optimal bandwidths, calculated with a data-driven plug-in algorithm for bandwidth selection in two dimensions, described in the online appendix.[37] For robustness, I also present results for constant bandwidths of 0.90 and 0.75 standard deviations.

IV estimation. For any political outcome P_{mt}, the second stage of the IV estimation for each frontier point is shown in Eq. (6)

$$\begin{aligned} P_{mt} = \beta_0 + \beta_1 \widehat{C_{mt}} + \beta_2 p_m^c + \beta_3 h_m^c + \beta_4 \delta_m p_m^c + \beta_5 \delta_m h_m^c + \eta_t + \gamma_s \\ + \theta_m + \epsilon_{mt} \end{aligned} \tag{6}$$

where the political effects are regressed on the predicted values of CCT coverage obtained in the first stage.[38] 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 ($\hat{\tau}_{IVavg}$) are again calculated using Eq. (5), replacing α_{1k} by β_{1k}.

Finally, Figs. A.IV through A.VII in the online appendix illustrate how the treatment effects for all political outcomes can be estimated under the alternative specifications explained before: the reduced single-score RDD, or the parametric approach over the two dimensional plane.[39] The construction of each plot is described in the online appendix.

5.2. Data and description of variables

Data on monthly municipal CCT coverage comes from the Ministry of Social Development (MDS). CCT coverage is measured as the percentage difference between the number of households receiving the benefit, and the number of poor households in each municipality.[40] In addition to *Bolsa Família*, I also include households receiving older CCT programs from the Ministry of Education (*Bolsa Escola*), Ministry of Health (*Bolsa Alimentao*) and the MDS (*Carto Alimentao*). However, given that 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 data on basic health transfers to municipalities, including the FHP. It also makes data available with respect to coverage by health teams, and outcomes measured within the scope of the public system. Annual budget allocation data has been obtained from the National Treasury database (FINBRA), which provides two different breakdowns of public expenses. First, in terms of capital investment, personnel expenses, and other. Second, it categorizes them by function (e.g. education and health).[41] Not all municipalities release the data every year, hence I only use the ones 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, used as a control.[42] 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 candidates.

The main empirical 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. In keeping with the model, the sample includes only municipalities where the mayor has reelection incentives.[43] Cases in which an eligible mayor did not run for reelection are not excluded from the main sample, as this decision is likely endogenous. However, for the estimation of the election outcomes, only municipalities where the incumbent is actually running can be used.[44]

Political parties are classified according to their level of clientelism following the Democratic Accountability and Linkages Project (DALP).[45] 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; and share of males. The GDP per capita is the average from 2000 to 2002. Again, in keeping with the model, 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 online appendix (Table A.IV).

5.3. Mapping the theory to the data

The model in Section 3 presents four predictions for the CCT effect on budget allocation, incumbency advantage, quality of politicians, and clientelism. They are tested using the cross-municipal variation in CCT coverage, instrumented by the FHP funding. All outcomes are measured for the 4-year electoral cycles of 2008 and 2012. For example, for the

[37] This algorithm is based on Zajonc (2012), with the following modifications: (1) Bandwidths are capped. Excessively wide bandwidths are a common problem in outputs of plug-in algorithms, so by capping the maximum bandwidth I effectively limit the amount of bias in the estimation; (2) the algorithm is expanded to estimate the bandwidth for different kernels; and (3) it allows the optimal bandwidths for the two scores to differ.

[38] The interaction between treatment δ_m and the score variables is also included in the second stage, given thatthe single instrument used in the estimation is the ATE for CCT coverage at each discontinuity point.

[39] The 2D graphs in Figure A.IV are similar to the ones presented in Dell (2010), and are 'three-dimensional analogues to standard two-dimensional RD plots.'

[40] The number of poor families per municipality is based on MDS estimates. MDS recalculated these estimates three times in the period of interest, 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.

[41] 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.

[42] 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 assume that health and sanitation expenses had the same ratio as in 2002–04, and I adjust the data accordingly.

[43] In some cases the previous 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.

[44] 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.

[45] 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.

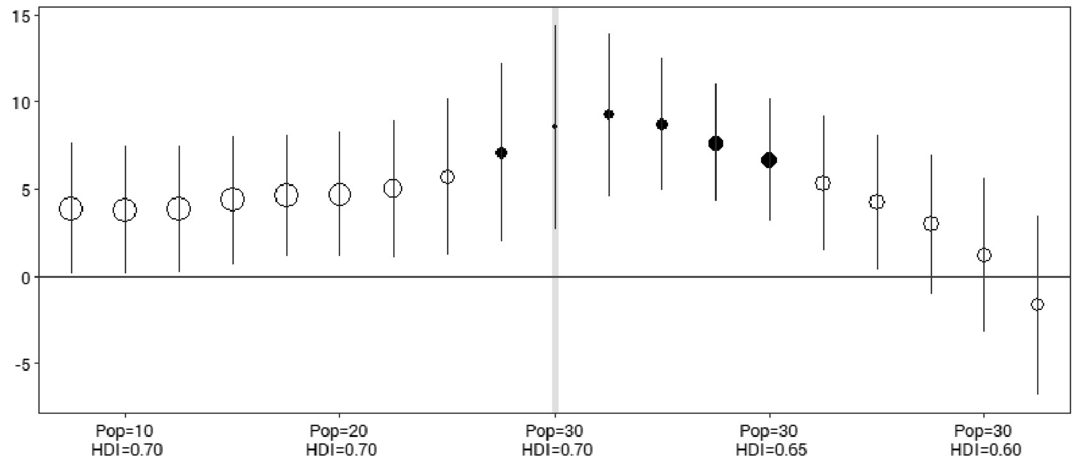

Fig. V. Effects on CCT coverage along the frontier. The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI fixed at 0.7 and population between 7500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. Coefficients are from a local linear regression (edge kernel) including year and state effects, and the controls listed in the text. The number of observations in each of the 19 bins is reflected in the size of dots. The dark dots represent the preferred frontier segment for which the instrument is strong.

2005–08 tenure, CCT coverage is recorded at the period's end, spending is measured in 2005–08, and election results in 2008.[46] The intuition is that during each 4-year tenure, mayors are constantly deciding on effort allocation, and constantly being affected by 'exogenous' CCT coverage in this process. At the end of each period, the election reflects the joint impact of CCT coverage level, its effects on voter's vulnerability, and past allocation of effort by incumbents.

Proposition 1 is tested using the 4-year budget allocation. The data categorizes spending on the basis of function, and the following six categories represent nearly 90% of all expenses: 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.[47]

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 (margin of victory and number of candidates in elections).

For Proposition 3, I proxy the incumbent's type using their education level, and the clientelism score of their parties. The underlining assumption here is that more educated politicians are also relatively more efficient at public good distribution (and less efficient at clientelism). 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.

As additional test of the theory is conducted using the election results for challengers. If we assume that voters could observe, at least imperfectly, the challenger's type before the election, CCTs should also make clientelistic challengers less competitive. This paper categorizes challengers according to their education and party, through four different dependent variables: (1) the share of less-educated challengers entering the race; (2) the share of less-educated challengers finishing in the top 2; (3) the share of challengers from clientelistic parties entering the race; and (4) the share of challengers from clientelistic parties finishing in the top 2.

[46] The same is done for the period 2009–12, so the two periods are pooled in the regressions. This means that the same municipality can appear at most two times in the main sample.

[47] Under this estimation strategy, the timing of CCT coverage increases and spending decisions might not perfectly coincide. A potentially more precise measure of spending could be taken only in the last 2 years of the mayoral tenure, which are closer to the election. Table A.III in the online appendix shows that under this specification, the results for pro-poor spending are actually stronger in magnitude and power. Keeping the 4-year measure in the main results is therefore a conservative approach.

Although Proposition 4 does not provide new insights on political outcomes, the alignment of the empirical results with the prediction provides support for the clientelism-driven 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 (online appendix) shows the main results for all municipalities.

Finally, the model prediction that interconnects all the others is that CCTs reduce clientelism. As it is the case in most of the related literature, the actual *quid-pro-quo* exchanges between patrons and voters is not directly measured. Accordingly, this paper's support for this narrative relies on three pillars: (1) there is extensive evidence in the literature that clientelistic relationships play a significant role in Brazilian municipal politics, which makes it natural to approach the model from this perspective; (2) the variation in the several political outcomes examined here suits the clientelism-driven mechanism, as predicted by the theoretical model, better than the most likely competing explanation. I discuss this at length in Section 6 with the results; and (3) the proxies for clientelism based on the performance of politicians that belong to parties traditionally identified with the practice, albeit imperfect, provide added evidence for this mechanism.

6. Results

First stage. The estimated MRDD effects on CCT coverage for 19 points along the frontier are shown in Fig. V. In line with the predictions from Section 5, the first stage results are fairly heterogeneous across municipal characteristics, and follow the pattern suggested by the information channel described before. The discontinuous FHP funding only generates an increase in CCT coverage around the central region of the treatment frontier, where neither the average old CCT coverage nor the pre-existing level of FHP service were high before treatment. In the extremes of the frontier, the instrument is either non-existent or weak. Thus, I define the six darker dots in the plot as the 'strong IV' segment in the sample,[48] and from now on all estimation results are reported for this 'preferred' segment of six frontier points. Table A.VII in the online appendix shows how the estimation results vary as we include more points to the left or right of this segment, using a sample with a weaker instrument. Overall, all coefficients remain with similar

[48] The six darker dots in the plot are only continuous frontier segment for which the average t-statistics is at least 3.2. Formally, the segment (S) is: $S = (p_m, h_m) : (p_m \geq 27.5, h_m = 0.7) \cup (p_m = 30, h_m \geq 0.65)$.

Table I
Regression discontinuity results.

Dependent variable	Mean (Pre-Treat.)	Coefficient (1)	(2)	(3)	(4)	(5)
Health funds, R$mn	1.752***	0.219***	0.245***	0.220***	0.249***	0.213***
CI	[1.63,1.86]	[0.14,0.29]	[0.13,0.30]	[0.15,0.34]	[0.18,0.31]	[0.13,0.29]
Obs. per bin		329	329	329	476	314
Bandwidths		(0.71,0.99)	(0.71,0.99)	(0.71,0.99)	(0.90,0.90)	(0.75,0.75)
CCT cov, p.p.	0.515	7.897***	5.915**	6.969**	8.439***	8.935***
CI	[−2.61,3.13]	[3.92,11.77]	[1.23,10.43]	[2.38,11.65]	[4.19,12.59]	[4.01,13.69]
Obs. per bin		612	612	612	476	314
Bandwidths		(1.00,1.00)	(1.00,1.00)	(1.00,1.00)	(0.90,0.90)	(0.75,0.75)
Bandwidth type	Optimal	Optimal	Optimal	Optimal	0.90	0.75
State and year effects	Yes	Yes	No	Yes	Yes	Yes
Controls	Yes	Yes	No	No	Yes	Yes

Health Funds effects are estimated 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 controls is described in the text.

magnitude and power under these different specifications. The average optimal bandwidths for each outcome variable in this preferred segment are also shown in the online appendix (Table A.I).

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

Table A.II in the online 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).

The remainder of the table shows the robustness of results to the choice of bandwidth, and the inclusion of controls and state and period effects. Robustness to the choice of kernel is shown in Table A.V (online appendix). The magnitude of the coefficients remains stable across specifications, only increasing as the bandwidth becomes narrower. This means that the paper's main specification is likely more conservative, i.e., any potential bias coming from widening the bandwidth, although seemingly small, would work in favor of the results.

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

Proposition 1. The main result of interest is the budget allocated to pro-poor spending. Table II shows that these services represent 52% of the budget, pre-treatment. In line with the theory, column (1) shows an increase of 3.7 pp (percentage points) in the share of budget allocated to these areas. From the IV regression (column 3), for a 10 pp increase in CCT coverage, the pro-poor spending increases by 4.4 pp. 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.[49] The positive effect for pro-poor spending was absent from previous electoral tenures, when this coefficient was negative and insignificant.

One potential concern here is that this positive effect could be a consequence of the mechanical health care budget increase generated by the FHP discontinuity. Nevertheless, this does not seem to be the case. While the annual increase in FHP funds isR$0.43mn for a municipality at the discontinuity, its budget for pro-poor public goods is nearly R $19mn/year. This means that the mechanical effect is responsible for, at most, a 0.55 pp increase in the share of pro-poor spending.[50] In contrast, the point estimate in Table II shows a pro-poor budget increase that is 7 times larger, with the lower end of the confidence interval still showing an increase 3 times larger than the maximum mechanical effect.

Table A.III in the online appendix shows the detailed coefficients for all six individual budget categories. No other category had significant coefficients,[51] which implies that the cost of re-allocating budget to pro-poor spending was divided across most other categories. The same Table also shows the effect on total budget, and on its breakdown by expense type (i.e. capital investment, personnel or others). This breakdown shows a significant decrease in capital investments, and an increase in expenses with personnel. This suggests that incumbents reallocate resources to redistributive spending from infrastructure investment. This is highly plausible, as education and health tend to be labor-intensive. While the potential implications of this budget shift in the profile of public employees is not the focus of this paper, it is discussed in the online appendix.

Proposition 2. The model predicts that CCTs will reduce the incumbent's support in reelection races. 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.

49 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 August 2004 the budget allocation for the year was already defined, and the candidates for the municipal election were campaigning.

50 This is the maximum mechanical effect. The FHP is jointly financed by states, municipalities and the federal government. As a response to an increase in federal financing, mayors could always reallocate some resources away from the program to other budget lines.

51 Transportation spending (2.7% of the total) was negative and significant. However, this result has to be treated with skepticism given that the negative effect pre-dates the treatment (it was already present in 1997–2004). Transportation spending in small municipalities is mostly infrastructure spending in road transport, as opposed to a public system of transportation.

Table II
Main results: political outcomes.

	Coefficient, [90% CI]				Avg band. (Pop. HDI) Obs. per bin
	Mean (Pre-Treat.)	RDD (1)	RDD past (2)	IV (3)	
Incumbent's vote share (%)	50.834	−7.712***	1.075	−0.822***	(0.99,0.96)
	[47.59,53.59]	[−12.42,-3.67]	[−3.03,5.94]	[−2.00,-0.30]	430
Margin of victory (p.p.)	16.425	−5.922**	−0.726	−0.627*	(1.00,1.00)
	[13.53,20.45]	[−11.24,-1.59]	[−4.13,3.43]	[−1.72,-0.09]	466
Candidates (number)	2.314	0.392***	0.118	0.041***	(1.00,0.98)
	[2.24,2.40]	[0.20,0.63]	[−0.08,0.32]	[0.02,0.10]	449
Pro-poor spending, (% share)	51.816	3.700***	−0.981	0.439**	(1.00,1.00)
	[49.67,52.51]	[1.63,6.71]	[−3.27,1.07]	[0.11,1.98]	350
Challenger's entry (share with HS)	0.113	−0.001	−0.020	0.000	(0.97,0.98)
	[0.08,0.17]	[−0.08,0.09]	[−0.11,0.07]	[−0.01,0.01]	785
Challenger's entry (share clientelistic)	0.375	−0.140**	−0.049	−0.014*	(0.97,0.95)
	[0.32,0.43]	[−0.23,-0.03]	[−0.15,0.05]	[−0.03,0.00]	761
Challenger is top 2 (share with HS)	0.140	−0.129**	0.051	−0.014**	(0.98,0.98)
	[0.09,0.20]	[−0.22,-0.04]	[−0.08,0.19]	[−0.04,0.00]	464
Challenger is top 2 (share clientelistic)	0.463	−0.118	−0.069	−0.012	(1.00,0.98)
	[0.39,0.54]	[−0.26,0.03]	[−0.22,0.07]	[−0.04,0.00]	469

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 are the average effect for the preferred frontier segment. Regressions include year and state effects and the list of controls described in the text. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment. Columns (1) and (2) show the reduced-form effects for the post- and pre-treatment periods, respectively; Column (3) shows the IV regression for the post-treatment period (2008, 2012).

From the IV regression, for a 10 pp increase in CCT coverage, there is an 8.2 pp vote loss for the incumbent. This result is robust to the choice of kernel, bandwidth, and the exclusion of municipal controls and fixed effects (Tables A.V and 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 1.1 pp.

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, but measure different dimensions of electoral competition. 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 6.3pp lower margin of victory. Again, estimates are also robust to different specifications; and there are no significant effects in past elections.

The overall direction shown by the results suits 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. Without treatment, the share of challengers with less than 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 10 pp increase in municipal CCT coverage, the share of less educated challengers that rank top 2 in the election is 14 pp lower.

The share of challengers from clientelistic parties that enter the race is 38%, whereas the share that rank top 2 in elections is 46%, on average before treatment. At the discontinuity, challengers from clientelistic parties are significantly less likely to enter the race in treated municipalities, but there is no difference in the share that rank top 2. Column (3) shows that, for a 10 pp increase in municipal CCT coverage, the share of challengers from clientelistic parties that run for mayor is 14 pp 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 finish in the top 2. Results are robust to the choice of kernel, bandwidth, and the exclusion of controls and effects. Also, no significant effects on those variables are observed in past elections.

This model prediction can also be tested by splitting the sample by either the party or the education level 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 model prediction, but some coefficients lose statistical significance due to small sample size. As expected, incumbents with less than a college degree lose more votes, and also shift relatively less effort towards pro-poor redistributive spending.[52] A similar pattern is observed when comparing the vote shares of incumbents from programmatic and clientelistic parties. Clientelistic incumbents suffer a much higher and significant vote loss.[53] Finally, the differences by party in the effects on pro-poor spending are less pronounced.

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.

Alternative interpretation: anti-clientelism or credit claiming by left parties? One potential alternative explanation for these effects

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

[53] 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.

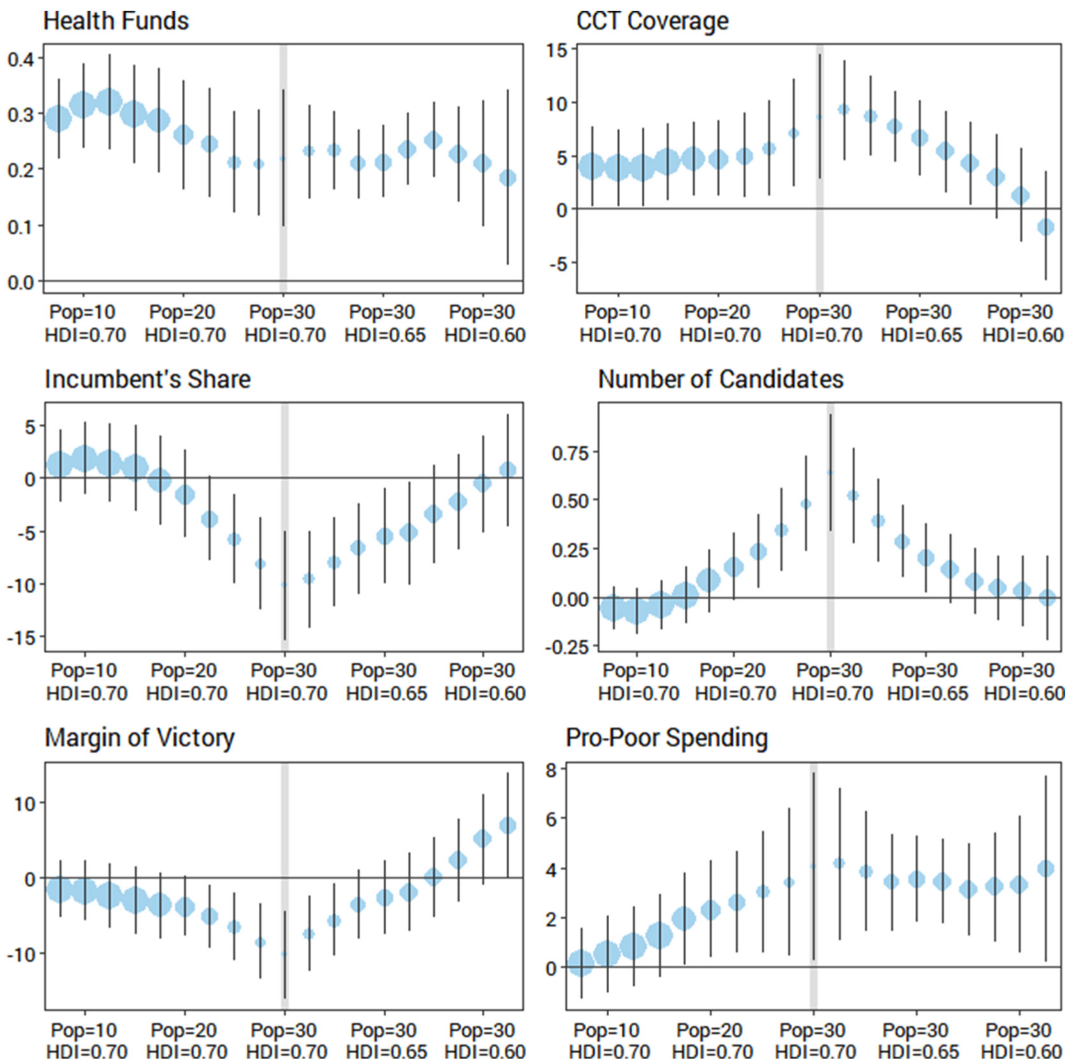

Fig. VI. Heterogeneity of the CATE (Continued). The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI fixed at 0.7 and population between 7500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. Coefficients are from a local linear regression (edge kernel) including year and state effects, and the controls listed in the text. The number of observations in each of the 19 bins is reflected in the size of dots.

would be that CCT shifts political support towards left parties, given that right-wing parties still held the majority of mayor positions in the period under analysis. Leftist President Lula (2003–2010) and his party (PT) are widely seen by the poor population as responsible for the successful spread of CCT benefits in Brazil (Zucco, 2013). Nevertheless, this alternative explanation does not fit the results. This paper tests party-specific political effects in two ways: (1) splitting the sample into incumbents from clientelistic and non-clientelistic parties, or left-wing and right-wing parties, and examining their electoral performance in reelection races (Table III);[54] and (2) using variables that measure the success of challengers in entering or being competitive in races against incumbents (see Fig. VII). While the former approach has the advantage of directly examining the performance on incumbents, the latter benefits from a larger sample size and more estimation power.

Table III shows the effects on the incumbent's vote share by party type. 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 other words, there is no evidence that CCT leads voters to increase support for leftist incumbents in municipal elections, as all mayors lose votes. In the online Appendix, I conduct a similar exercise to verify if there are potential credit claiming effects operating solely in favor of PT incumbent mayors, but again I find no evidence of such mechanism. On the contrary, other leftist mayors seem to lose slightly less votes than PT mayors with the arrival of CCTs. Fig. VII examines the performance of challengers by party. Again, I find no evidence that leftist challengers are more likely to enter races against incumbents as a result of increased CCT coverage, nor that they are more likely to be competitive after entering. The online appendix also shows similar results for challengers of PT, with no evidence of any credit claiming mechanism.

The lack of evidence for this alternative narrative does not come as a surprise in the Brazilian context. The country has candidate-driven local elections, with weak party identities, and politicians that constantly change allegiances. What is more, the local CCT program was carefully designed to promote the federal government brand and Lula's administration, in detriment of local politicians that only play a secondary

[54] In Brazil, the lines that separate clientelistic and non-clientelistic parties are similar to the ones separating left- and right-wing parties, but still distinct. 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. Given that PSDB is a very large party, the two different definitions result in very distinct party aggregations. As a matter of fact, even though PSDB was the main opposition party to PT during the period under analysis, its mayors perform in a way that is much more similar to PT than the mayors from clientelistic parties that belong to PT's federal coalition.

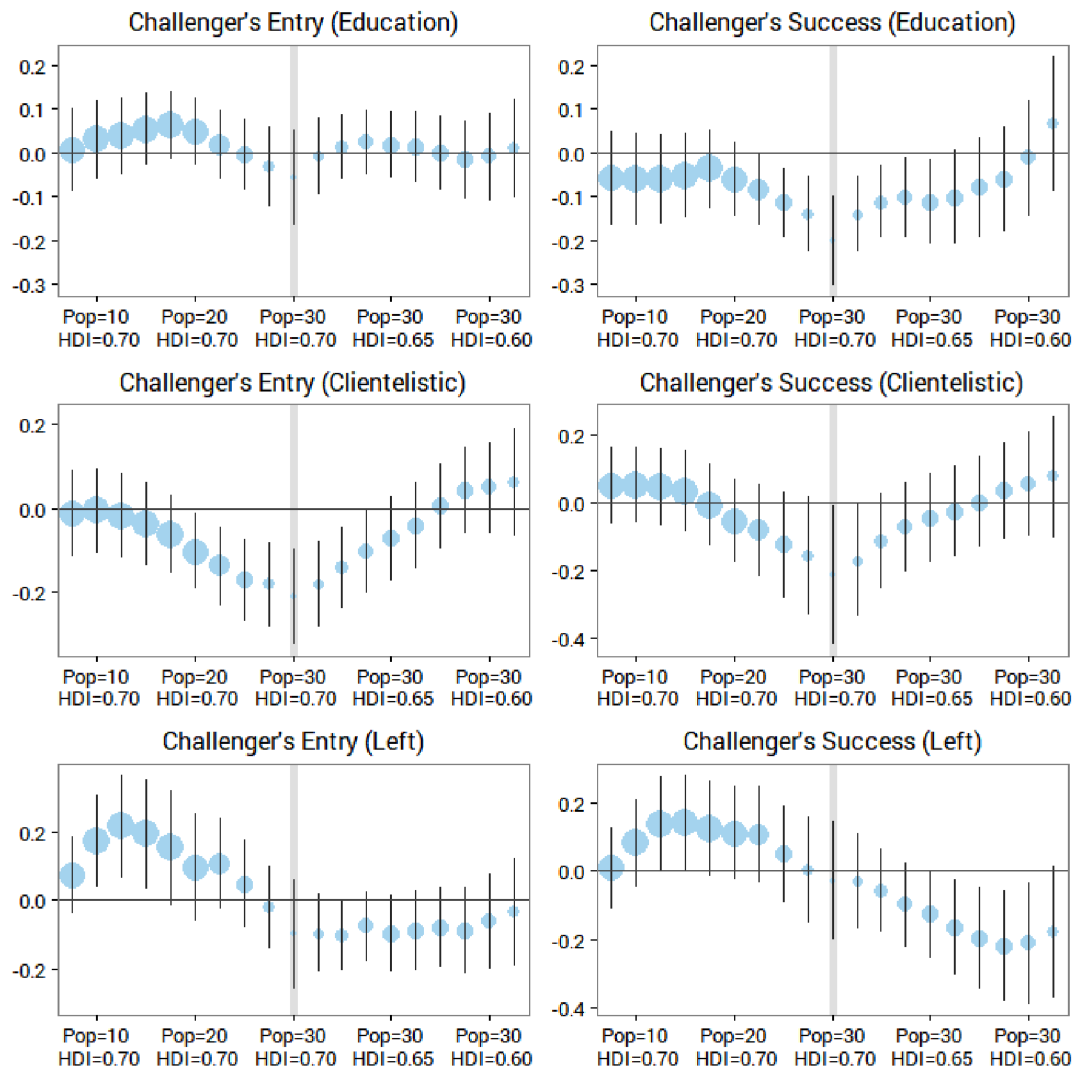

Fig. VII. Heterogeneity of the CATE. The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI fixed at 0.7 and population between 7500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. Coefficients are from a local linear regression (edge kernel) including year and state effects, and the controls listed in the text. The number of observations in each of the 19 bins is reflected in the size of dots.

administrative role. In this environment, voters do not associate CCTs with local administrations (Sugiyama and Hunter, 2013), and even PT mayors might find it difficult to claim credit for the program's arrival.

6.1. Assessing the identification strategy

This section discusses the empirical evidence supporting this paper's identification strategy, and the sources of its main potential violations. Given that the empirical assessment of both the first stage and the instrument's exogeneity were discussed previously, this section focuses on the threats coming from alternative policy discontinuities, and on the exclusion restriction.

Alternative policy discontinuities. The existence of other discontinuous assignments of policies around the same thresholds of population and/or HDI used in this paper could confound the results. The *Fundo de Participacao dos Municipios* (FPM) is an important source of budget resources for municipalities, and it distributed in a discontinuous form across several population thresholds,[55] one of which is 30,564. Nevertheless, I show four pieces of evidence that strongly suggest that the FPM is not causing the political effects observed in this paper.

First, the methodology of fund allocation in this design (FHP) differs from the FPM methodology. The population numbers that determines the eligibility to FHP funding were fixed in 2003, while the population for FPM purposes is re-estimated every year. The fact that the population of many municipalities crossed the relevant thresholds (30,564 and 30,000) between 2003 and 2012 forcibly reduces the possibility of confounding coming from this program, especially in the estimation under the edge kernel.[56] Second, there is no significant effect on the total budget of municipalities, estimated at the discontinuity (online appendix, Table A.III). This suggests that the FPM does not generate a funding gap at the 30,000 threshold. This is not surprising, given that the FPM funding premium is only high at the lower population thresholds.[57]

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

56 The edge kernel overweights observations closer to the discontinuity. Those are more likely to come from municipalities where the treatment status for the FHP and the FPM differ.

57 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 III
Heterogeneity in political outcomes.

	Incumbent's vote share		Pro-poor spending	
split by: incumbent's education	*High*	*Low*	*High*	*Low*
Coefficient	−4.433	−11.133**	7.067***	0.385
CI	[−10.58,0.84]	[−20.04,-2.99]	[4.09,14.41]	[−3.74,3.00]
Obs. per bin	226	205	136	122
split by: incumbent's party	*Programmatic*	*Clientelistic*	*Programmatic*	*Clientelistic*
Coefficient	−2.537	−7.877*	3.352**	2.658*
CI	[−11.75,4.59]	[−15.73,-0.38]	[1.08,7.50]	[0.03,6.28]
Obs. per bin	152	216	112	187
split by: incumbent's party	*Left*	*Right*	*Left*	*Right*
Coefficient	−7.007*	−7.512**	2.865	3.472**
CI	[−14.90,-0.18]	[−13.64,-2.16]	[−0.29,6.66]	[1.10,6.87]
Obs. per bin	130	300	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.

Third, the FPM system dates back to the 1980s. Any direct effects on political outcomes would have been necessarily observed in past elections (2000 and 2004), which is not the case as discussed before (Table II). Fourth, Table A.VII (online appendix) shows a falsification test. I estimate the reduced form coefficients for political outcomes in 2008 and 2012, setting the population cutoff for different FPM thresholds, at 23,773 and 37,356 (one threshold lower and one higher than 30,564). None of the variables had any statistically significant result in the same direction as the results in this paper. All-in, these tests suggest that the FPM is only a threat to identification under the following highly unlikely scenario: if it generates the political results for the period of 2008 and 2012, and not before; and only for the municipalities around the population threshold of 30,564, and not for lower or higher thresholds (23,773 and 37,356 population). Finally, I am not aware of any other policy discontinuity at either HDI of 0.7, or population of 30,000, in the same period.

The exclusion restriction. In its most restrictive form, this assumption requires that the FHP funding only affects political outcomes through CCT coverage. By construction, this assumption cannot be directly tested with the existing data. Nevertheless, even if the FHP has unobserved direct effects on politics, they are not a significant threat to the identification in this paper under two conditions: (1) if they push the political outcomes in the opposite direction of this paper's estimates; or (2) if their magnitude is small relative to the estimated MRDD coefficients. In the following paragraphs, I argue that these conditions are likely met in the case of the political variables examined in this paper, and also explain how the multiple falsification tests here provide support for this argument.

To illustrate this, consider the potential existence of FHP direct effects on incumbency advantage. Even if this is plausible, excludability can be shown if one provides evidence that this direct effect is likely positive, i.e., if higher FHP funding helps mayors. Given that this paper finds the exact opposite, this would suggest that our MRDD results are being underestimated. This is in fact the most plausible scenario in Brazil, given that the FHP program is the best-rated government policy run by municipalities in the country (IPEA, 2011). What is more, mayors can actually elect whether or not to join and help financing this program. Given that the vast majority of them do so, it is likely that the program is electorally attractive.

Under these circumstances, a failure in the exclusion restriction automatically implies that the observed loss of political support by incumbents is a direct consequence of the improvement in the quality of health care to poor households. I emphasize that this narrative does not seem plausible. Poor beneficiaries are likely to reward and not punish incumbents for better health care. What is more, wealthy beneficiaries have no reason to reduce their support for the mayor, given that the FHP funding did not cannibalize any other budget category.

Consider also pro-poor spending. The FHP direct effect is obviously positive here, as more FHP funds mechanically increase the share of budget spent in health services. In this case, I can neither show nor argue that the direct effects are opposite to our positive coefficients. However, I can provide evidence that their magnitude is small enough that still justifies the existence of positive and significant effects mediated by the CCT coverage. As discussed before, this mechanical effect could be responsible for, at most, a 0.55 pp increase in the share of pro-poor spending. In contrast, the MRDD estimate is this paper shows a pro-poor budget increase that is 7 times larger than this, with the lower end of the confidence interval still showing an increase 3 times larger. A failure in the exclusion restriction here would suggest that higher pro-poor spending is the result of a positive multiplier effect on the FHP funding. Although I cannot rule-out this explanation with the existing data, I do not see any clear mechanism that would trigger such propagation, apart from CCT coverage.

Falsification tests. I provide empirical support for these arguments with two types of falsification tests, which are applied on all political outcomes. The intuition is that the direct effects of FHP on politics can only be observed in a counterfactual world where the program has no effect on CCT coverage. Thus, these tests consist in reproducing this counterfactual world with both the variation in the first stage estimates across groups of municipalities, and the variation in FHP effects in the pre- and post-treatment periods.

The first two tests rely on the heterogeneity of the first stage results. As detailed in the MRDD estimates discussed in Sections 5 and 6, the first stage coefficients are fairly heterogeneous along the treatment frontier. For municipalities in both extremes (very low population or very low HDI), the instrument is weak or nonexistent. In these areas, any effects on political outcomes should be directly generated by the FHP program. Figs. VI and VII show, for all the paper's eight political outcomes, that in the absence of a significant first stage, there is also little change in political outcomes. In fact, for all variables of interest, their effect heterogeneity closely resembles the heterogeneity in first stage estimates.

To further investigate this, I compare two equal size ranges of the frontier (6 bins): the first is the preferred segment for which the instrument is statistically strong. The second is a segment with population in the range 7500–17,500, and HDI of 0.7, where the instrument is notably weak.[58] The coefficients for the weak-IV segment are shown in Table A.X in the online appendix. All variables that were significant in the preferred segment become statistically insignificant, suggesting again that direct effects of FHP do not play a significant role on political outcomes.

A second falsification test now compares two groups of municipalities: one with high and one with low CCT coverage already in 2004. The idea is that, even for the municipalities in the 'strong IV' range of the frontier, there is some heterogeneity in pre-existing CCT coverage. Municipalities that were better covered would have less room for CCT expansion based on the information channel, even though they have

[58] 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 and higher estimation power. 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 I.

been equally affected by the FHP funding. Again, if the FHP affects politics mainly through CCT coverage, we should see stronger political effects in municipalities with LOWER pre-existing CCT coverage. Table A.XIII (online appendix) shows this for all outcomes.

There is one potential source of excludability violation that these tests cannot assess: the existence of an unobservable characteristic of municipalities that generates the same heterogeneity pattern along the treatment frontier in the estimates for both the first stage results and the direct effects of FHP on politics. The existence of such potential unobservable is minimized with the inclusion of several covariates as controls in the MRDD regressions.

The second approach to falsification tests uses panel data on the implementation and coverage of the FHP program in the four electoral cycles in 2000–2012.[59] I regress the multiple political outcomes (y_{mt}) on a dummy that indicates the presence or coverage of FHP in municipality m at time t (FHP_{mt}). This variable is interacted with a dummy that assumes value of one in the periods after the creation of BF ($postBF_m$). I also include municipality dummies (γ_{Muni}) to identify the corresponding FHP effects using only variation in coverage within municipality. The intuition for this approach is that, even if direct FHP effects exist, they should significantly change post-BF, in the same direction of our MRDD estimates for the multiple political outcomes. The regression is shown in the equation below.

$$y_{mt} = \alpha_0 + \alpha_1 FHP_{mt} * postBF_m + \alpha_2 FHP_{mt} + postBF_m + \gamma_{Muni} + \mu_{mt} \quad (7)$$

where α_2 represents the FHP direct effect, and α_1 represents the change in this effect post-BF. Table A.IX (online appendix) shows the coefficients α_2 and α_1 for two different definitions of FHP presence: specification (A) defines FHP as a dummy indicating whether the program has been implemented or not in the municipality, and specification (B) defines FHP as a dummy indicating whether the program covers at least 50% of the targeted population in the municipality.

This panel regression suggests that some direct effects of FHP on the incumbent's vote share and pro-poor spending might exist, but are indeed as expected. Both the program's presence and coverage have a positive effect on the political support for the mayor in the period before CCTs. They also increase the share of pro-poor spending, due to the positive mechanical effects on the health care budget. The most interesting results, however, come from the change in these effects post-BF (2008 and 2012 electoral cycles). They both change in the same direction of our MRDD estimates: the effect on the support for the incumbent falls, and the effect on pro-poor spending increases even further. What is more, the same pattern is observed for all the other political outcomes, albeit with different levels of statistical confidence.[60]

Even though this exercise cannot rule out the existence of direct FHP effects, it suggests that they are either in opposition to the effects found in this paper (incumbent's vote share) or not strong enough to justify the variation in pro-poor spending. In any case, this evidence supports the existence of FHP effects on politics that are intermediated by CCT coverage. The main violation to this specific assessment would come from an unobservable event that, in the period post-BF, changed either the direction or the magnitude of all the FHP effects on politics in the same direction as the MRDD estimates in this paper. I am not aware of any significant change in either policy or the economy that would have been as relevant to these outcomes in the exact same period.

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 CCTs reduce incumbency advantage, increase both electoral competition and the quality of candidates, and weaken the support for clientelistic parties. They 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 significant policy implications. First, they suggest that policy spillovers matter. In addition to the political impacts of the CCT program, the paper also shows that a small differential in funding for a health program generated a much larger impact on the CCT allocation across municipalities, which was an 'unintended' but desirable consequence of the original program. Second, these findings are useful to inform other social policies. I believe that this paper's results could be generalized to any policy that, shielded from capture, permanently reduces the vulnerability of voters. In that case, they suggest that effective redistributive efforts can provide a path out of clientelism for developing democracies plagued by the practice. Third, the results also shed light on 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.

This paper also shows that the CCT program is positive for the poor 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 capital-intensive investment, for example, is beyond the identification strategy employed here. Finally, the magnitude of effects has to be treated carefully when applied to a more general context, for two reasons. First, it 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 paper provides solid evidence for the exclusion restriction, it does so without definitively ruling out the potential existence of such direct effects.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpubeco.2019.05.002.

References

Alston, Lee J., Mueller, Bernardo, 2006. Pork for policy: executive and legislative exchange in Brazil. J. Law Econ. Org. 22 (1), 87–114.
Alves, Denisard, Timmins, Chris, 2003. Social exclusion and the two-tiered healthcare system of Brazil. In: Behrman, J.R., Trujillo, A.G., Szekely, M. (Eds.), Who Is in and Who Is out: Social Exclusion in Latin America. Inter-American Development Bank, Washington, DC.
Ames, Barry, Smith, Amy Erica, 2010. Knowing left from right: ideological identification in Brazil, 2002–2006. J. Polit. Lat. Am. 2 (3), 3–38.
Anderson, Siwan, Francois, Patrick, Kotwal, Ashok, 2015. Clientelism in Indian villages. Am. Econ. Rev. 105 (6), 1780–1816.
Baez, Javier E., Camacho, Adriana, Conover, Emily, Zarate, Roman A., 2012. Conditional cash transfers, political participation, and voting behavior. Policy Research Working Papers.
Bobonis, Gustavo, Gertler, Paul, Gonzalez-Navarro, Marco, Nichter, Simeon, 2017. Vulnerability and Clientelism (Working Paper).
Boix, Carles, Stokes, Susan C., 2009. Political clientelism. The Oxford Handbook of Comparative Politics.
Brollo, Fernanda, Nannicini, Tommaso, Perotti, Roberto, Tabellini, Guido, 2013. The political resource curse. Am. Econ. Rev. 103 (5), 1759–1796.
Brusco, Valeria, Nazareno, Marcelo, Stokes, Carol, 2004. Vote buying in Argentina. Lat. Am. Res. Rev. 39 (2), 66–88.
Clark, Damon, Martorell, Paco, 2014. The signaling value of a high school diploma. J. Polit. Econ. 122 (2), 282–318.
Cruz, Cesi, Schneider, Christina J., 2017. Foreign aid and undeserved credit claiming. Am. J. Polit. Sci. 61 (2), 396–408.

[59] I use FHP presence and coverage for the panel analysis, as opposed to funding, because the discontinuity in payments was only created right after the arrival of CCTs, unfortunately.

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

Cruz, Cesi, Keefer, Philip, Labonne, Julien, 2018. Buying Informed Voters: New Effects of Information on Voters and Candidates (Working Paper).

De Janvry, Alain, Finan, Frederico, Sadoulet, Elisabeth, 2012. Local electoral incentives and decentralized program performance. Rev. Econ. Stat. 94 (3), 672–685.

De La O, Ana L., 2013. Do conditional cash transfers affect electoral behavior? Evidence from a randomized experiment in Mexico. Am. J. Polit. Sci. 57 (1), 1–14.

De La Oh, Ana L., 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.

Efron, Bradley, 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.

Ferraz, Claudio, Finan, Frederico, 2011. Electoral accountability and corruption: evidence from the audits of local governments. Am. Econ. Rev. 101 (4), 1274–1311.

Finan, Frederico, Schechter, Laura, 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 Dev. 40 (5), 1042–1053.

Fujiwara, Thomas, 2015. Voting technology, political responsiveness, and infant health: evidence from Brazil. Econometrica 83 (2), 423–464.

Fujiwara, Thomas, Wantchekon, Leonard, 2013. Can informed public deliberation overcome clientelism? Experimental evidence from Benin. Am. Econ. J. Appl. Econ. 5 (4), 241–255.

Gerber, Alan S., Kessler, Daniel P., Meredith, Marc, 2011. The persuasive effects of direct mail: a regression discontinuity based approach. J. Polit. 73, 140–155.

Hicken, Allen, 2011. Clientelism. Annu. Rev. Polit. Sci. 14 (1), 289–310.

IBOPE, 2005. Compra de Votos nas Eleicoes 2004. Ibope OpiniÃ£o.

Imbens, Guido W., Kalyanaraman, Karthik, 2012. Optimal bandwidth choice for the regression discontinuity estimator. Rev. Econ. Stud. 79 (3), 933–959.

Imbens, Guido W., Lemieux, Thomas, 2008. Regression discontinuity designs: a guide to practice. J. Econ. 142 (2), 615–635.

IPEA, 2011. SIPS - Sistema de Indicadores de Percepo Social (IPEA).

Jacob, Brian A., Lefgren, Lars, 2004. Remedial education and student achievement: a regression-discontinuity analysis. Rev. Econ. Stat. 86 (1), 226–244.

Keele, Luke J., Titiunik, Roco, 2015. Geographic boundaries as regression discontinuities. Polit. Anal. 23 (1), 127–155.

Labonne, Julien, 2013. The local electoral impacts of conditional cash transfers: evidence from a field experiment. J. Dev. Econ. 104 (0), 73–88.

Lee, David S., Lemieux, Thomas, 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355.

Lindert, Kathy, Linder, Anja, Hobbs, Jason, de la Briere, Benedicte, 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, Miguel, Edward, Vigorito, Andrea, 2011. Government transfers and political support. Am. Econ. J. Appl. Econ. 3 (3), 1–28.

Nichter, Simeon, Peress, Michael, 2017. Request fulfilling: when citizens demand clientelist benefits. Comp. Pol. Stud. 50 (8), 1086–1117.

Papay, John P., Willett, John B., Murnane, Richard J., 2011. Extending the regression-discontinuity approach to multiple assignment variables. J. Econ. 161 (2), 203–207.

Reardon, Sean F., Robinson, Joseph P., 2010. Regression Discontinuity Designs With Multiple Rating-Score Variables (Center for Education Policy Analysis Working Papers).

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

Sugiyama, Natasha Borges, Hunter, Wendy, 2013. Whither clientelism? Good governance and Brazil's Bolsa Familia Program. Comp. Polit. 46 (1), 43–62.

Vicente, Pedro C., Wantchekon, Leonard, 2009. Clientelism and vote buying: lessons from field experiments in African elections. Oxf. Rev. Econ. Policy 25 (2), 292–305.

Wong, Vivian C., Steiner, Peter M., Cook, Thomas D., 2013. Analyzing regression-discontinuity designs with multiple assignment variables a comparative study of four estimation methods. J. Educ. Behav. Stat. 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. Am. J. Polit. Sci. 47 (3).