A Mathematical Appendix

The equation to be maximized by incumbents is the following:

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

The first order condition is then:

$$\psi n \mu_x [\theta \frac{1}{e} - \delta(y)] - \kappa = 0 \tag{2}$$

We can solve for e as follows:

$$e = \frac{\theta}{\delta(y) + \mathrm{K}} \tag{3}$$

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

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

$$e_y = \frac{-\theta \delta_y}{(\delta(y) + \mathrm{K})^2} > 0 \tag{4}$$

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

$$\pi = \phi n \mu_x [\theta \log(e) - \theta \log(E) + \delta(y)(E - e)] + \phi(\frac{1}{2\phi} - \lambda) \tag{5}$$

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

$$\pi_y = \phi n \mu_x \delta_y [E - (e + \frac{\theta \mathrm{K}}{(\delta(y) + \mathrm{K})^2})] \tag{6}$$

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

Proof of proposition 3 First I show that clientelistic types allocate less effort to public goods when y increases. Starting with e_y from equation 4, the derivative of this term with respect to θ is shown below, giving us the desired result:

$$e_{y\theta} = \frac{-\delta_y}{(\delta(y) + \mathrm{K})^2} > 0 \tag{7}$$

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

$$\pi_y = \phi n \mu_x \delta_y [E - (\frac{\theta}{(\delta(y) + \mathrm{K})} + \frac{\theta \mathrm{K}}{(\delta(y) + \mathrm{K})^2})] \tag{8}$$

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

$$\pi_{y\theta} = -\phi n \mu_x \delta_y [\frac{\delta(y) + 2\mathrm{K}}{(\delta(y) + \mathrm{K})^2} > 0] \tag{9}$$

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

$$e_{yn} = e_{y\mathrm{K}} \mathrm{K}_n = \frac{-2(\delta(y) + \mathrm{K})}{(\delta(y) + \mathrm{K})^4} * -\frac{\mathrm{K}}{n} > 0 \tag{10}$$

We can do the same to π_y, shown in equation 8. We know that $\mathrm{K}_n = -\mathrm{K}_n/2 < 0$ when $\kappa > 0$. Now, equation 8 can be decomposed as follows:

$$\pi_y = \phi n \mu_x \delta_y [E - \frac{\theta \delta(y)}{(\delta(y) + \mathrm{K})^2}] - \frac{2\theta \kappa \phi \delta_y}{\psi(\delta(y) + \mathrm{K})^2} \tag{11}$$

Derivating with respect to n and reorganizing we have:

$$\pi_y = \phi \mu_x \delta_y [E - e\frac{\delta(y)}{(\delta(y) + \mathrm{K})} - 2\theta \mathrm{K} \frac{(2\mathrm{K} + \delta(y))}{(\delta(y) + \mathrm{K})^3}] < 0 \tag{12}$$

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

B CCT, Vote Buying and The Poor's Welfare

The current theoretical model examines the effect of an exogenous income shock to poor voters – brought out by CCT – on both the electoral behavior of these voters and the responses of incumbent mayors. In doing so, the model predictions do not directly address the welfare of poor voters. For example, are voters better off with more CCT, and consequently more public goods, even if that leads politicians to reduce vote buying? In this section I show how the current model could provide some insights on these welfare considerations.

The implementation of CCTs influences the utility of poor voters both directly and indirectly. The indirect effect is given by the parameter $\delta(y)$ in the second term of the poor's utility: $U^I(g,c) = \theta \log(e) + \delta(y)(E-e)$. In words, as CCTs make households marginally less attractive for vote buying, politicians put less effort in this strategy, and the contribution of the second term to the total utility falls.

As for the direct effects of CCTs, the intuition is simple: higher income should increase the voter's welfare. The current version of the model does not explicit include this. Given that CCT is being treated as exogenous, the inclusion would be inconsequential for the comparative statics of interest. Nevertheless, for the purpose of examining the welfare of poor voters, assume that their utility has an additional component $h(y)$, which is a concave function of income. Also, let us include a multiplier m on the term $\delta(y)$, which measures the voter's 'taste' for vote buying. The new utility is now: $U = \theta \log(e) + m\delta(y)(E-e) + h(y)$.

We are ultimately interested in the sign of the utility's derivate with respect to income, which is given below:

$$U_y = \frac{\partial U}{\partial y} + \frac{\partial U}{\partial e} e_y \tag{13}$$

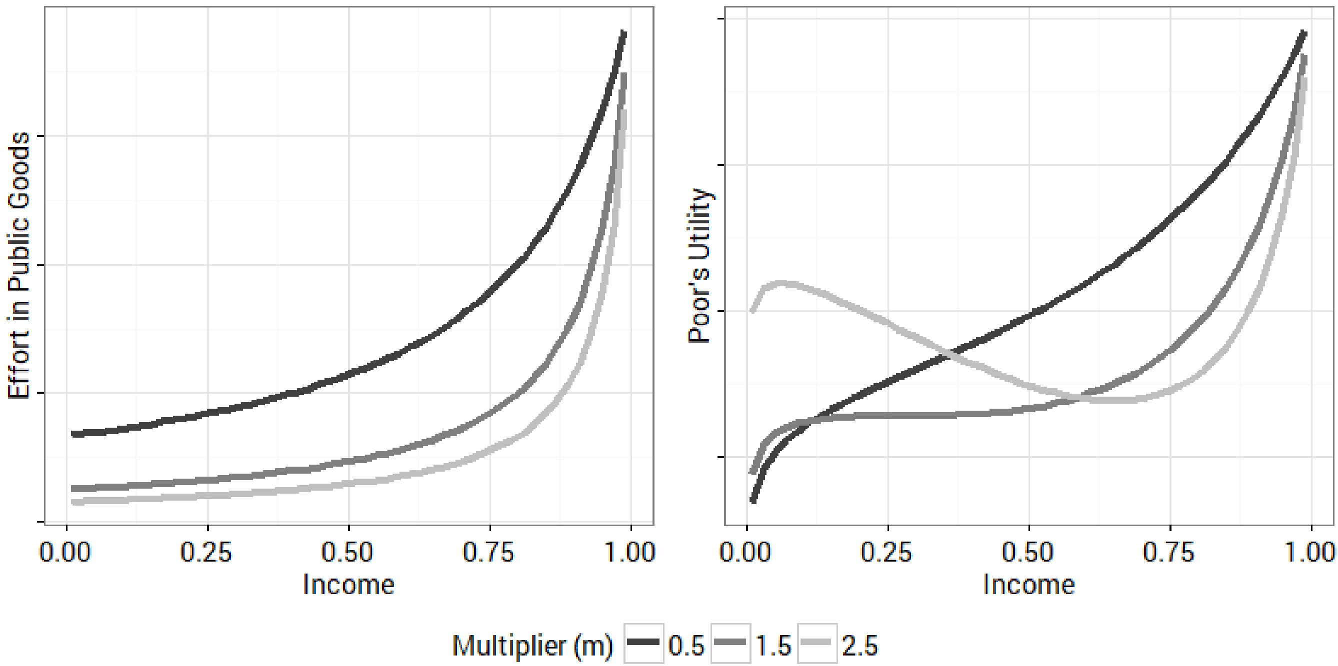
After derivating and rearranging, we arrive at the following equation:

$$U_y = m[\delta_y(E-e) - \delta(y)ee_y] + h_y + \frac{\theta}{e}e_y \tag{14}$$

The second and third term in this equation represent the welfare impact coming from direct increases in both income and the provision of public goods, respectively. By model assumptions, both these terms are positive (welfare increasing). The first term represents the utility loss coming from less vote buying, and it is always negative in equilibrium. It is easy to see that the net welfare changes depend on the magnitude of m, i.e., if voters have a high taste for vote buying, CCT might actually bring an utility loss.

Figure A.I below illustrates this mechanism by showing both the effort e and the poor's utility U in equilibrium, for different values of income y, and taste m. Under all parameter values, increases in income always lead politicians to increase effort in public goods, thereby reducing vote buying (left-side plot). As expected, income increases are always improve welfare if m is low enough. However, as the taste for vote buying increases, CCTs might be welfare reducing when the initial income y is low enough (right-side plot).

Figure A.I: Welfare Under Different Model Parameters



The other model parameters are: $\psi = 5$, $n = 0.4$, $\mu_x = 0.05$, $\kappa = 0.01$, and $\theta = 0.9$. I also assume $\delta(y) = 1 - y$, and $h(y) = 0.1\log(y)$.

C Effects on Hiring

The empirical evidence indicates that the budget shift to redistributive services of health and education happens in tandem with an increase in the budget share spent with personnel. I briefly examine the available hiring data[61] to provide insight on how this budget reallocation can affect public services in the long term. The estimation results are shown in Table A.XII.

In Brazil, public servants are hired through a competitive, meritocratic process, which includes a written exam. Employees hired through this process have job security guaranteed by the legislation. Mayors are allowed to bypass the process in two ways. First, the creation of "political" positions, usually reserved for high level executive jobs in the administration. These are often allocated to

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

allies, or used as political exchange. Second, the hiring of temporary employees for a specific project and a limited tenure.

Table A.VII shows the effects, at the discontinuity, on both the total employment and the share of the permanent work force hired in political jobs. These effects are shown for both the level and rate of change in these variables. The only noteworthy effect is a 8.3pp higher change in the share of temporary hiring in 2008-12. In the data, temporary hiring is generally positively correlated with budget changes.[62] This indicates that it might be a procedure used by mayors to conduct budget changes quickly, without committing to long term labor costs.

The education level of the labor force provides another evidence that this type of hiring is consonant with the budget reallocation narrative in this study. The share of temporary workers with less than a high school degree is 18.0pp lower for treated municipalities. This is in line with the expectation that a shift to health and education would increase the demand for high-skilled jobs (e.g. doctors, health agents, teachers), while the fall in capital investment should reduce the demand for low-skilled jobs (construction workers).

While these results are consistent with the budget shift, their persistence over the long term might have implications for public good distribution that are still unknown, and beyond the scope of this study. If temporary jobs are not replaced by permanent positions[63], thus outliving the budget changes, there are at least two potential issues for future discussion here. First, there is a potential trade-off between quality and effort in public service. Where the lack of a formal selection process for temporary jobs might negatively impact the quality of public service, the absence of job security might work as a mechanism to extract a higher effort. Second, the persistence of this type of hiring could be evidence that mayors are using a different avenue to conduct clientelistic exchanges with the wealthier population. Because there is no significant effect observed in 2008-12 in the political hires, the evidence for this argument seems limited now, but it might worth examining in the future.

D Bandwidth Selection

Bandwidth selection is a significant component of RDDs, as the bandwidth is usually the tool used to control the trade-off between bias and efficiency in the estimation. Although there are plenty of approaches to select the bandwidth in one-score RD designs,[64] there is still sparse literature discussing similar methodologies for the MRDD. In fact, most of the literature applying some form

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

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

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

of MRDD does not discuss the issue at all.[65] I will use a plug-in algorithm for bandwidth selection building on the work of Zajonc (2012). In this section I briefly discuss the main practical challenges in implementing the procedure in the multivariate context. I describe the algorithm on that paper and how it tackles those problems, and I propose alterations. All technical notation in this section is taken from that paper to facilitate reference.

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

The main difficulties for the two-score case are described here. First, the plug-in method does not have a closed form solution for more than one bandwidth. Zajonc (2012) uses the same bandwidth for both scores for a feasible solution. Second, bandwidths cut the kernel differently along the frontier, so the shape of the treated and non-treated subsets is endogenous to the bandwidth selection near the origin.[66] Here, the solution was to calculate the bandwidth only for points far from the origin. Finally, unreliable bandwidth values can arise due to the assumptions involved in the calculation of the MSE parameters. (Imbens and Kalyanaraman, 2012) correct this problem with a regularization term. Zajonc (2012) calculates the optimal bandwidth for various points and uses the minimum value for under-smoothing.

I propose a different approach for tackling these three problems. First, I allow the bandwidths to be different for each score variable. Thus, instead of using an expression for the single optimal bandwidth, I will minimize the MSE expression numerically for different pairs of bandwidths. This provides efficiency gains in nearly all cases, as the two-dimensional bandwidth outperforms the unique one in terms of MSE. Furthermore, I use an elliptical bandwidth instead of a rectangular one, to ensure that the points within a lower distance from the cut-off are used.[67]

Second, if the sample is not balanced along the frontier, the optimal bandwidth calculated away

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

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

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

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

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

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

1. Using the entire sample, calculate the standard deviations for population σ_p and HDI σ_h.

2. Select a pilot bandwidth for each variable using Scott's rule (e.g.$\hat{h} = \sigma_p n^{-\frac{1}{6}}$) and limit the sample to those bandwidths.

3. Calculate the conditional variance $\hat{v}(p, h)$ and density $\hat{f}(p, h)$.

4. Apply again a rule-of-thumb bandwidth to create a subsample for the estimation of the second derivatives, which are calculated using a second degree local polynomial regression on both sides of the discontinuity. Here, although I keep the rule-of-thumb bandwidth originally proposed, I add a cap at of 1.65 in order to avoid using the extreme points in the estimation of the second derivatives, i.e. 5% of the sample on each side. Estimate the second derivatives for both sides.

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

$$MSE = (Bias_1 + Bias_2 + Bias_3)^2 + 2 * Variance \tag{15}$$

where,

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

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

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

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

Conditional bias:

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

Conditional variance:

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

where

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

As for notation, $H^{\frac{1}{2}}$ is a bandwidth matrix assumed to be diagonal as $diag([\ h_p\sigma_p\ \ h_i\sigma_i\ ])$, M^0 and M^1 are the hessian matrices for the second degree polynomial estimated using pilot bandwidths for the non-treated and treated subsamples, respectively. $K(u)$ is the kernel,

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

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

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

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

E Additional Tables

Table A.I: Optimal Bandwidths

Kernel	Edge	Epanechnikov	Normal	Uniform
Health Funds	(0.71,0.99)	(0.64,0.97)	(0.53,0.94)	(0.51,0.90)
CCT Coverage	(1.00,1.00)	(1.00,0.98)	(1.00,0.94)	(1.00,0.92)
Incumbent's Vote Share	(0.99,0.96)	(0.99,0.93)	(0.99,0.90)	(0.99,0.89)
Number of candidates	(1.00,0.98)	(0.99,0.96)	(0.99,0.93)	(0.99,0.92)
Margin of Victory	(1.00,1.00)	(1.00,1.00)	(1.00,0.99)	(1.00,0.99)
Challenger's entry (share without HS)	(0.97,0.98)	(0.93,0.96)	(0.88,0.93)	(0.89,0.93)
Challenger's entry (share Clientelistic)	(0.97,0.95)	(0.95,0.94)	(0.91,0.91)	(0.90,0.90)
Challenger is top 2 (share without HS)	(0.98,0.98)	(0.95,0.96)	(0.89,0.93)	(0.89,0.92)
Challenger is top 2 (share Clientelistic)	(1.00,0.98)	(0.99,0.95)	(0.97,0.92)	(0.96,0.90)
Total Budget (log)	(0.94,1.00)	(0.85,0.99)	(0.71,0.96)	(0.68,0.95)
Capital Investment	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)
Personnel Spending	(1.00,1.00)	(1.00,1.00)	(1.00,0.96)	(0.99,0.95)
Other Spending	(1.00,1.00)	(1.00,0.99)	(0.99,0.95)	(0.99,0.94)
Pro-poor spending	(1.00,1.00)	(0.98,0.99)	(0.97,0.95)	(0.96,0.94)
Education	(1.00,1.00)	(0.99,1.00)	(0.98,1.00)	(0.97,1.00)
Health	(1.00,0.96)	(1.00,0.94)	(0.99,0.92)	(0.99,0.91)
Administration	(1.00,1.00)	(1.00,1.00)	(0.98,0.90)	(0.98,0.84)
Urbanization	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)	(1.00,1.00)
Social Security	(0.98,1.00)	(0.96,1.00)	(0.93,0.96)	(0.93,0.95)
Transportation	(0.99,0.98)	(0.99,0.97)	(0.99,0.93)	(0.99,0.93)

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

Table A.II: Balance of Fixed and Pre-determined Variables

	Mean (Pre-Treat.)	Coefficient [90% CI]			Opt. Band (Pop,HDI)
Bandwidth	*Optimal*	*Optimal*	*0.90*	*0.75*	
Longitude	-40.590	-0.158	-0.214	-0.278	(1.00,0.97)
(degrees)	[-41.19,-40.04]	[-0.64,0.26]	[-0.73,0.24]	[-0.87,0.25]	
Latitude	-13.161	-0.064	-0.039	-0.039	(0.98,1.00)
(degrees)	[-14.01,-12.32]	[-0.67,0.50]	[-0.71,0.60]	[-0.86,0.75]	
Inequality	21.418	2.314	2.256	2.498	(1.00,1.00)
top 10%/ bottom 40%	[20.44,22.68]	[-0.02,4.83]	[-0.26,5.06]	[-0.54,5.81]	
Age Profile	39.605	-0.441	-0.362	-0.292	(1.00,1.00)
share 20-50	[39.30,39.91]	[-1.05,0.15]	[-1.06,0.26]	[-1.15,0.43]	
per capita GDP[a]	0.818	0.261	0.280	0.314	(0.96,0.99)
R$‘000	[0.66,0.99]	[-0.12,0.62]	[-0.13,0.66]	[-0.18,0.77]	
Area[a]	63.012	1.134	1.062	-0.296	(1.00,1.00)
km2	[60.48,65.67]	[-3.88,5.55]	[-4.43,5.65]	[-6.36,5.03]	
Urban Pop.	49.902	0.024	0.032	0.095	(1.00,1.00)
% share	[49.70,50.11]	[-0.31,0.53]	[-0.35,0.61]	[-0.40,0.80]	
Gender	58.557	-0.193	0.019	-0.577	(1.00,1.00)
% share of male	[52.91,64.36]	[-12.34,11.03]	[-13.67,12.39]	[-15.82,14.84]	
FHP Teams	39.874	0.746	0.697	0.778	(1.00,1.00)
% coverage	[38.30,41.21]	[-1.97,2.97]	[-2.31,3.17]	[-2.72,3.82]	
Poverty Rate	0.003	0.066	0.048	0.023	(0.99,0.99)
% poor	[0.00,0.00]	[-0.16,0.30]	[-0.20,0.31]	[-0.27,0.35]	
CCT Coverage	-14.335	1.202	1.661	3.164	(0.99,1.00)
% over target	[-19.06,-8.94]	[-6.51,9.60]	[-6.66,10.98]	[-6.53,14.13]	
Old CCT Benefits	18.567	-0.849	-1.031	-1.354	(1.00,0.99)
% pop.	[17.25,20.06]	[-3.18,1.33]	[-3.50,1.30]	[-4.14,1.31]	

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

Table A.III: Main Results: Other Budget Categories

	Coefficient, [90% CI]				Avg Band.
	Mean	RDD	RDD past	IV	(Pop. HDI)
	(Pre-Treat.)	(1)	(2)	(3)	Obs. per bin
Total budget	149.135	-0.042	0.085	-0.003	(0.94,1.00)
(RSmn)	[137.78,160.16]	[-0.13,0.03]	[-0.02,0.20]	[-0.01,0.00]	325
Capital investment	10.558	-1.726*	0.773	-0.150*	(1.00,1.00)
(% share)	[9.70,11.73]	[-3.34,-0.22]	[-1.18,3.32]	[-0.35,-0.02]	352
Expenses w. personnel	46.577	2.852**	0.201	0.248**	(1.00,1.00)
(% share)	[44.82,47.97]	[0.66,5.20]	[-1.65,2.21]	[0.07,0.60]	352
Other	42.750	-1.149	-0.812	-0.104	(1.00,1.00)
(% share)	[41.59,43.91]	[-3.05,0.82]	[-3.08,1.35]	[-0.34,0.05]	351
Health	22.225	2.174***	0.153	0.268**	(1.00,0.96)
(% share)	[21.10,23.39]	[1.05,3.73]	[-1.76,1.81]	[0.07,1.28]	329
Education	29.644	1.842**	-1.202	0.234	(1.00,1.00)
(% share)	[28.33,30.90]	[0.34,3.91]	[-2.56,0.09]	[-0.02,1.04]	350
Urbanization	9.750	-0.121	2.047*	-0.031	(1.00,1.00)
(% share)	[9.10,10.40]	[-1.76,1.24]	[0.13,4.48]	[-0.50,0.19]	353
Administration	14.359	-1.146	0.301	-0.127	(1.00,1.00)
(% share)	[13.42,15.40]	[-3.16,0.96]	[-1.94,2.83]	[-0.67,0.25]	353
Social Security	6.251	-0.061	-0.445	-0.003	(0.98,1.00)
(% share)	[5.59,7.10]	[-1.50,1.12]	[-1.63,0.60]	[-0.23,0.18]	347
Transportation	2.711	-0.794**	-1.089**	-0.095	(0.99,0.98)
(% share)	[2.28,3.21]	[-1.34,-0.19]	[-2.03,-0.19]	[-0.29,0.05]	344
Pro-poor (2Y bfr. election)	52.146	4.195***	-1.939	0.532**	(1.00,1.00)
(% share)	[50.47,53.45]	[2.25,7.11]	[-4.29,0.06]	[0.20,1.80]	407

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

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

	Coefficient, [90% CI]				Avg Band.
	Mean	RDD	RDD past	IV	(Pop. HDI)
	(Pre-Treat.)	(1)	(2)	(3)	Obs. per bin
Incumbent's vote share	50.292	-6.147*	1.680	-0.672*	(0.99,0.97)
(%)	[47.59,53.59]	[-11.06,-1.06]	[-2.21,6.35]	[-1.66,-0.11]	534
Margin of victory	16.870	-3.482	-1.463	-0.385	(1.00,1.00)
(p.p.)	[14.48,20.31]	[-9.01,3.10]	[-4.79,2.54]	[-1.27,0.32]	570
Candidates	2.359	0.368***	0.112	0.040***	(0.99,0.98)
(number)	[2.29,2.44]	[0.18,0.61]	[-0.07,0.32]	[0.02,0.10]	551
Pro-poor spending,	52.277	2.483**	-1.834	0.321	(0.99,0.99)
(% share)	[50.34,53.06]	[0.43,5.34]	[-4.11,0.28]	[-0.02,1.95]	458
Challenger's entry	0.134	-0.030	-0.030	-0.003	(0.97,0.98)
(share with HS)	[0.10,0.20]	[-0.11,0.05]	[-0.12,0.06]	[-0.02,0.01]	985
Challenger's entry	0.386	-0.121*	-0.021	-0.013*	(0.97,0.96)
(share Clientelistic)	[0.34,0.43]	[-0.21,-0.02]	[-0.12,0.09]	[-0.03,0.00]	960
Challenger is top 2	0.144	-0.139***	0.047	-0.015***	(0.98,0.99)
(share with HS)	[0.10,0.19]	[-0.22,-0.06]	[-0.09,0.18]	[-0.03,-0.01]	576
Challenger is top 2	0.489	-0.108	-0.026	-0.012	(1.00,0.98)
(share Clientelistic)	[0.43,0.55]	[-0.25,0.04]	[-0.17,0.13]	[-0.04,0.00]	583

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

Table A.V: RDD: Robustness to Kernel Choice

Dependent Variable	(1)	(2)	(3)
Health Funds	0.219***	0.215***	0.198***
(R$mn)	[0.14,0.29]	[0.14,0.29]	[0.10,0.30]
Obs. per bin	329	269	180
CCT Coverage	7.897***	7.645***	7.292***
(p.p. over target)	[3.92,11.77]	[3.71,11.41]	[3.31,10.93]
Obs. per bin	612	590	544
Incumbent's vote share	-7.712***	-7.360***	-7.042***
(%)	[-12.42,-3.67]	[-11.82,-3.46]	[-11.20,-3.20]
Obs. per bin	430	414	385
Margin of victory	-5.922**	-5.335**	-4.699**
(p.p.)	[-11.24,-1.59]	[-10.14,-1.33]	[-9.14,-0.99]
Obs. per bin	466	466	464
Candidates	0.392***	0.369***	0.338***
(number)	[0.20,0.63]	[0.18,0.59]	[0.15,0.55]
Obs. per bin	449	432	406
Pro-poor spending,	3.699***	3.495***	3.439***
(% share)	[1.63,6.71]	[1.54,6.14]	[1.52,5.81]
Obs. per bin	350	346	319
Challenger's entry	-0.001	-0.008	-0.010
(share with HS)	[-0.08,0.09]	[-0.08,0.08]	[-0.09,0.08]
Obs. per bin	785	741	676
Challenger's entry	-0.140**	-0.139**	-0.130**
(share Clientelistic)	[-0.23,-0.03]	[-0.23,-0.03]	[-0.22,-0.02]
Obs. per bin	761	730	675
Challenger is top 2	-0.129**	-0.126**	-0.121**
(share with HS)	[-0.22,-0.04]	[-0.21,-0.04]	[-0.21,-0.03]
Obs. per bin	464	441	389
Challenger is top 2	-0.118	-0.111	-0.102
(share Clientelistic)	[-0.26,0.03]	[-0.25,0.04]	[-0.24,0.04]
Obs. per bin	469	449	417
Kernel	Epanech.	Gaussian	Uniform

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

Table A.VI: RDD: Other Robustness Tests

Dependent Variable	(1)	(2)	(3)	(4)
Incumbent's vote share	-7.043**	-7.582***	-7.883***	-8.040**
(%)	[-11.56,-2.29]	[-12.11,-3.11]	[-12.96,-3.51]	[-13.96,-2.29]
Obs. per bin	430	430	364	241
Margin of victory	-4.819**	-5.640**	-6.494**	-7.011**
(p.p.)	[-9.39,-0.75]	[-10.98,-0.96]	[-12.49,-1.69]	[-14.54,-1.21]
Obs. per bin	466	466	364	241
Candidates	0.328***	0.345***	0.424***	0.463***
(number)	[0.11,0.58]	[0.15,0.58]	[0.22,0.68]	[0.22,0.77]
Obs. per bin	449	449	364	241
Pro-poor spending	3.553**	2.938**	3.875***	4.244**
(% share)	[1.15,5.90]	[0.48,5.82]	[1.52,7.39]	[1.06,8.61]
Obs. per bin	350	350	275	179
Challenger's entry	-0.015	-0.002	-0.004	0.006
(share without HS)	[-0.08,0.05]	[-0.08,0.08]	[-0.08,0.09]	[-0.08,0.12]
Obs. per bin	785	785	664	455
Challenger's entry	-0.118**	-0.128**	-0.148**	-0.144*
(share Clientelistic)	[-0.21,-0.02]	[-0.22,-0.02]	[-0.25,-0.03]	[-0.26,-0.02]
Obs. per bin	761	761	664	455
Challenger is top 2	-0.131***	-0.130**	-0.145**	-0.163**
(share without HS)	[-0.20,-0.06]	[-0.21,-0.05]	[-0.24,-0.05]	[-0.27,-0.05]
Obs. per bin	464	464	382	255
Challenger is top 2	-0.112	-0.121	-0.135	-0.139
(share Clientelistic)	[-0.26,0.04]	[-0.27,0.04]	[-0.29,0.02]	[-0.34,0.04]
Obs. per bin	469	469	382	255
State and Year Effects	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Bandwidth	Optimal	Optimal	0.90	0.75

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

Table A.VII: RDD: Other Frontier Segments

Dependent Variable	(1)	(2)	(3)	(4)
Health Funds	0.219***	0.218***	0.221***	0.227***
(R$mn)	[0.14,0.29]	[0.14,0.29]	[0.15,0.29]	[0.16,0.29]
Obs. per bin	329	330	359	383
CCT Coverage	7.897***	7.590***	7.441***	6.648***
(p.p. over target)	[3.92,11.77]	[3.56,11.40]	[3.60,11.06]	[3.04,10.07]
Obs. per bin	612	625	646	699
Incumbent's vote share	-7.712***	-7.436***	-7.272***	-6.318***
(%)	[-12.42,-3.67]	[-12.10,-3.48]	[-11.86,-3.21]	[-10.50,-2.52]
Obs. per bin	430	432	444	458
Margin of victory	-5.922**	-6.020**	-5.239**	-4.783**
(p.p.)	[-11.24,-1.59]	[-11.26,-1.80]	[-10.43,-1.09]	[-9.60,-0.96]
Obs. per bin	466	472	492	524
Candidates	0.392***	0.386***	0.348***	0.305***
(number)	[0.20,0.63]	[0.20,0.62]	[0.17,0.57]	[0.14,0.50]
Pro-poor spending,	3.699***	3.598***	3.655***	3.412***
(% share)	[1.63,6.71]	[1.57,6.62]	[1.78,6.42]	[1.71,5.92]
Obs. per bin	350	350	365	378
Challenger's entry	-0.001	-0.002	0.001	0.002
(share with HS)	[-0.08,0.09]	[-0.08,0.09]	[-0.07,0.09]	[-0.07,0.08]
Obs. per bin	785	796	811	854
Challenger's entry	-0.140**	-0.144**	-0.123*	-0.114**
(share Clientelistic)	[-0.23,-0.03]	[-0.24,-0.04]	[-0.21,-0.02]	[-0.20,-0.02]
Obs. per bin	761	771	776	811
Challenger is top 2	-0.129**	-0.127**	-0.124**	-0.114**
(share with HS)	[-0.22,-0.04]	[-0.21,-0.04]	[-0.21,-0.04]	[-0.20,-0.04]
Obs. per bin	464	479	484	531
Challenger is top 2	-0.118	-0.119	-0.102	-0.090
(share Clientelistic)	[-0.26,0.03]	[-0.26,0.03]	[-0.23,0.04]	[-0.21,0.04]
Obs. per bin	469	484	487	533
Number of bins in the frontier	6	7	7	10
Population range	27.5-30	25-30	30	20-30
HDI range	0.7-0.65	0.7-0.65	0.7-0.6375	0.7-0.625

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

Table A.VIII: Outcomes at Different FPM Thresholds

Dependent Variable	(1)	(2)
CCT Coverage	-2.998	-4.806
(pp over target)	[-6.84,0.99]	[-10.47,1.10]
Obs. per bin	1053	354
Incumbent's vote share	1.992	1.261
(%)	[-2.31,5.84]	[-6.65,7.95]
Obs. per bin	739	246
Margin of victory	4.039*	0.716
(p.p.)	[0.18,8.03]	[-11.24,9.44]
Obs. per bin	802	270
Candidates	-0.164	0.053
(number)	[-0.35,0.01]	[-0.21,0.31]
Obs. per bin	768	256
Pro-poor spending,	0.483	-0.491
(% share)	[-1.47,2.03]	[-3.77,2.99]
Obs. per bin	591	203
Challenger's entry	0.020	-0.135*
(share with HS)	[-0.07,0.11]	[-0.33,0.00]
Obs. per bin	1248	493
Challenger's entry	0.124*	0.161
(share Clientelistic)	[0.01,0.23]	[-0.08,0.41]
Obs. per bin	1196	476
Challenger is top 2	0.007	-0.158
(share with HS)	[-0.10,0.11]	[-0.40,0.03]
Obs. per bin	789	268
Challenger is top 2	0.125	0.053
(share Clientelistic)	[-0.03,0.27]	[-0.24,0.35]
Obs. per bin	795	268
Pop. cutoff	23,772	37,356

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

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

	Specification (A)		Specification (B)	
	FHP	FHP x Post-BF	FHP	FHP x Post-BF
Incumbent's vote share	3.736***	-2.750*	3.452***	-2.486
(%)	[0.882]	[1.560]	[0.920]	[2.324]
Obs	(7356)	(7356)	(7356)	(7356)
Pro-poor spending	0.944***	0.098	0.704***	1.323**
(% share)	[0.236]	[0.386]	[0.238]	[0.572]
Obs	(7124)	(7124)	(7124)	(7124)
Challenger's entry	-0.029	-0.043	-0.032*	-0.080*
(share with HS)	[0.019]	[0.029]	[0.019]	[0.043]
Obs	(11851)	(11851)	(11851)	(11851)
Challenger's entry	-0.031	-0.043	-0.034	-0.106**
(share Clientelistic)	[0.022]	[0.033]	[0.021]	[0.047]
Obs	(11851)	(11851)	(11851)	(11851)
Challenger is top 2	-3.120	-2.426	-1.779	-7.203
(share with HS)	[2.666]	[4.162]	[2.673]	[6.235]
Obs	(7729)	(7729)	(7729)	(7729)
Challenger is top 2	-0.868	1.922	-1.705	-8.074
(share Clientelistic)	[3.011]	[5.003]	[2.985]	[7.105]
Obs	(7729)	(7729)	(7729)	(7729)

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

Table A.X: Results for the Weak IV segment

	Coefficient, [90% CI]				Avg Band.
	Mean	RDD	RDD past	IV	(Pop. HDI)
	(Pre-Treat.)	(1)	(2)	(3)	Obs. per bin
Incumbent's vote share	49.019	0.779	0.184	0.144	695
(%)	[46.23,51.92]	[-2.31,4.27]	[-3.53,3.60]	[-0.89,1.72]	(0.71,0.93)
Margin of victory	16.823	-2.665	-1.967	-0.517	945
(p.p.)	[14.15,20.22]	[-6.68,1.23]	[-5.39,0.84]	[-3.97,0.33]	(1.00,0.87)
Candidates	2.308	0.001	0.110	-0.001	792
(number)	[2.22,2.42]	[-0.12,0.12]	[-0.03,0.27]	[-0.10,0.06]	(0.78,1.00)
Pro-poor spending,	49.149	1.033	-1.777*	-0.511*	614
(% share)	[47.97,50.18]	[-0.34,2.61]	[-3.22,-0.26]	[-130.66,-0.08]	(0.79,0.78)
Challenger's entry	0.243	0.040	-0.006	0.007	1420
(share with HS)	[0.17,0.30]	[-0.04,0.12]	[-0.09,0.07]	[-0.01,0.04]	(0.90,1.00)
Challenger's entry	0.400	-0.038	-0.138***	-0.006	1292
(share Clientelistic)	[0.33,0.47]	[-0.13,0.05]	[-0.21,-0.05]	[-0.04,0.02]	(0.84,1.00)
Challenger is top 2	0.306	-0.055	-0.042	-0.011	1104
(share with HS)	[0.24,0.38]	[-0.14,0.04]	[-0.14,0.06]	[-0.06,0.01]	(1.00,1.00)
Challenger is top 2	0.462	0.024	-0.232***	0.006	1106
(share Clientelistic)	[0.39,0.54]	[-0.09,0.13]	[-0.31,-0.13]	[-0.02,0.06]	(1.00,1.00)

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. Regressions include year and state effects, and the controls listed in the text. The first column corresponds to the predicted values for a municipality at the discontinuity before treatment. Column (1) shows the RDD effects of the discontinuity in health funding on political outcomes for the post-treatment period (2008,2012); Column (2) shows the same effects as (1), but for the pre-treatment period (2000,2004); Column (3) shows the IV regression for the post-treatment period (2008,2012).

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

	Mean	Coefficient, [90% CI]	
	(Pre-Treat.)	2008-12	2000-04
Number of Visits[a]	86.54	0.11***	0.03
('000 per year, 4-year avg.)	[83.46,89.52]	[0.04,0.17]	[-0.11,0.17]
Obs per bin		476	542
Children below 2y[a]	0.38	0.10**	0.06
('000 in any given month)	[0.36,0.39]	[0.03,0.18]	[-0.07,0.19]
Obs per bin		476	542
Number of Babies Born[a]	0.33	0.10**	0.09
('000 per year, 4-year avg.)	[0.32,0.34]	[0.02,0.19]	[-0.03,0.21]
Obs per bin		476	542
Visits per Family	11.41	0.96*	-0.61
(per year)	[11.10,11.77]	[0.13,1.79]	[-1.81,0.40]
Obs per bin		476	542
Mortality Rate	10.79	1.09	1.98
(children less than 11m)	[9.48,12.03]	[-0.96,3.19]	[-1.33,5.50]
Obs per bin		470	529
Pre-Natals	66.27	0.69	3.58
(% of pregnancies)	[63.69,69.90]	[-4.31,7.20]	e [-2.91,11.09]
Obs per bin		423	408
Children < 2y Vaccinated	95.71	0.04	1.36
(% of children)	[94.69,96.26]	[-0.91,1.09]	[-1.09,3.87]
Obs per bin		476	542

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

Table A.XII: Hiring Outcomes

	Mean (Pre-Treat.)	Coefficient [90% CI]
Total Employment	1.187	-0.04
('000)	[1.12,1.27]	[-0.14,0.05]
Obs per bin		475
Share of Political Employment	0.09	-0.05
(pct)	[0.08,0.11]	[-0.29,0.16]
Obs per bin		476
Share of Temporary Employment	0.19	0.17
(pct)	[0.15,0.23]	[-0.17,0.56]
Obs per bin		431
Total Employment (chg.)	28.06	0.66
(p.p.)	[22.49,39.18]	[-10.12,9.89]
Obs per bin		474
Share of Political Employment (chg.)	16.68	-5.65
(p.p.)	[-0.23,2.56]	[-23.08,2.15]
Obs per bin		471
Share of Temporary Employment (chg.)	2.11	8.19***
(p.p.)	[-1.32,4.99]	[4.08,13.72]
Obs per bin		464
Share of Less Educated Employees	28.06	-4.27
(pct of total)	[25.62,31.01]	[-9.54,0.96]
Obs per bin		421
Share of Less Educated Employees	30.81	-17.20***
(pct of temporary)	[25.23,42.50]	[-34.13,-7.47]
Obs per bin	0	421
Formal hiring process in 2007-11	0.76	-0.06
(Yes=1)	[0.63,0.82]	[-0.23,0.11]
Obs per bin		476

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

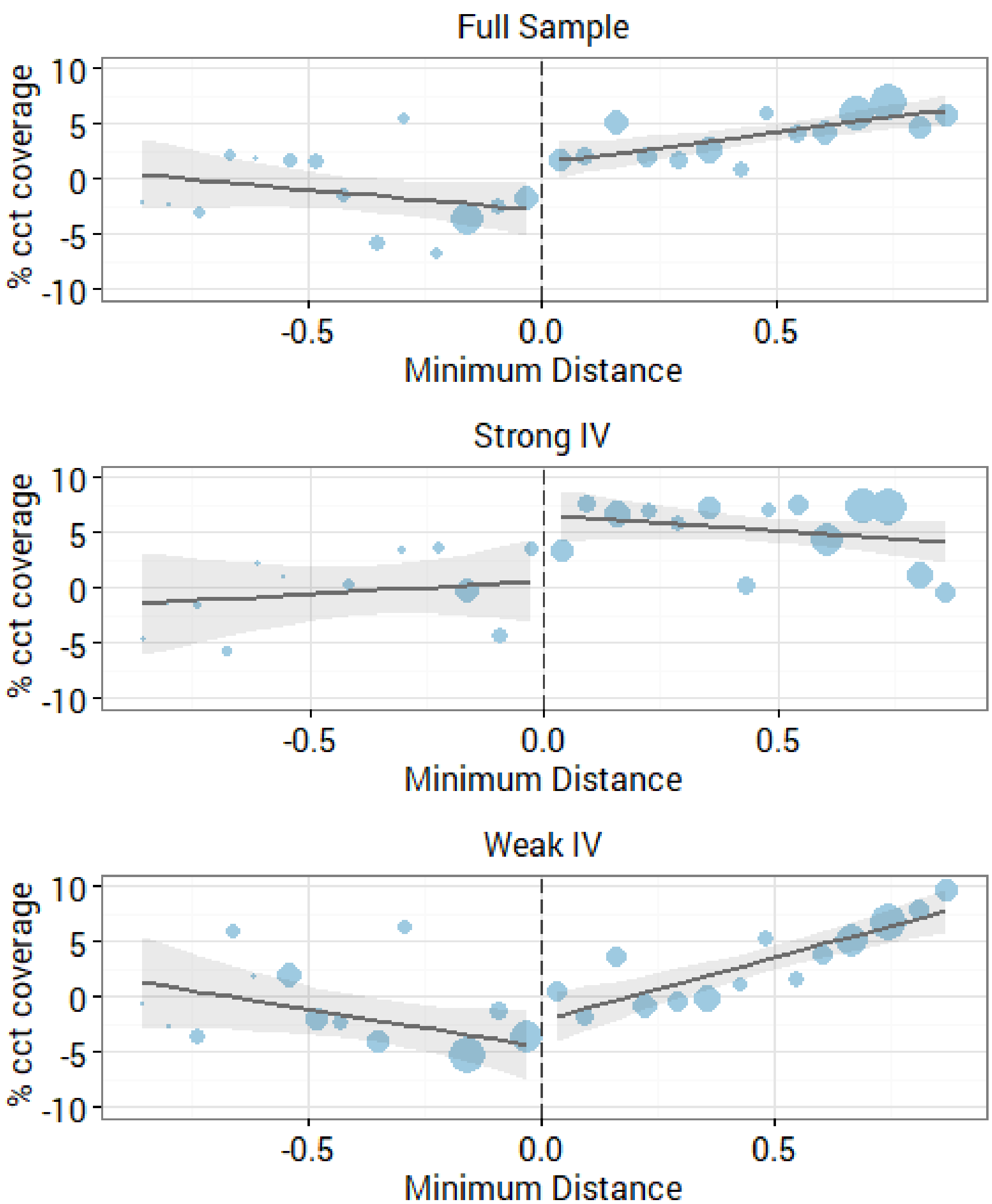
Table A.XIII: Outcomes at Different Pre-Existing CCT Coverage

Dependent Variable	(1)	(2)
Pre-existing CCT Cov.	High	Low
CCT Coverage	6.442**	12.105***
(p.p. over target)	[1.45,11.91]	[5.19,18.58]
Obs. per bin	367	254
Incumbent's vote share	-3.247	-11.892**
(%)	[-9.27,2.50]	[-20.89,-4.11]
Obs. per bin	251	184
Margin of victory	-3.992	-4.853
(p.p.)	[-9.28,0.62]	[-14.90,2.83]
Obs. per bin	274	199
Candidates	0.269*	0.504***
(number)	[0.00,0.57]	[0.21,0.93]
Obs. per bin	263	192
Pro-poor spending,	2.855**	5.119*
(% share)	[0.77,5.83]	[0.89,11.72]
Obs. per bin	208	146
Challenger's entry	0.091	-0.101
(share with HS)	[-0.01,0.20]	[-0.24,0.02]
Obs. per bin	452	342
Challenger's entry	-0.042	-0.218
(share Clientelistic)	[-0.21,0.11]	[-0.38,0.00]
Obs. per bin	434	335
Challenger is top 2	-0.031	-0.214**
(share with HS)	[-0.14,0.08]	[-0.40,-0.07]
Obs. per bin	272	198
Challenger is top 2	-0.051	-0.165
(share Clientelistic)	[-0.25,0.16]	[-0.39,0.04]
Obs. per bin	274	200

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

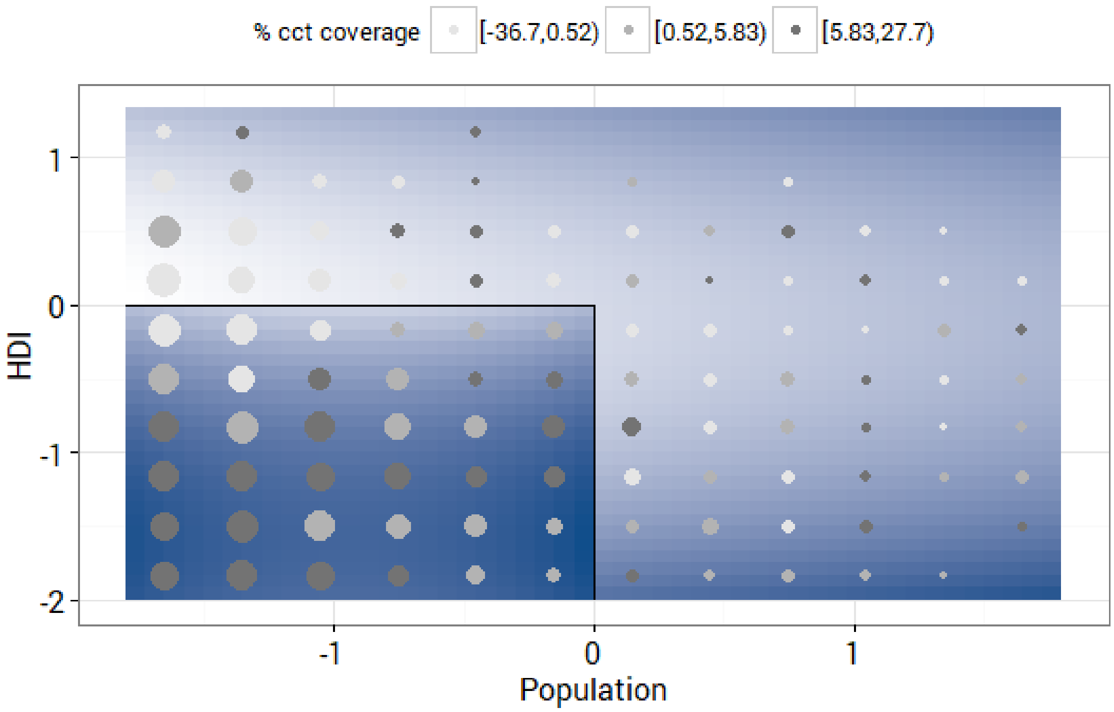
F Additional Figures

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



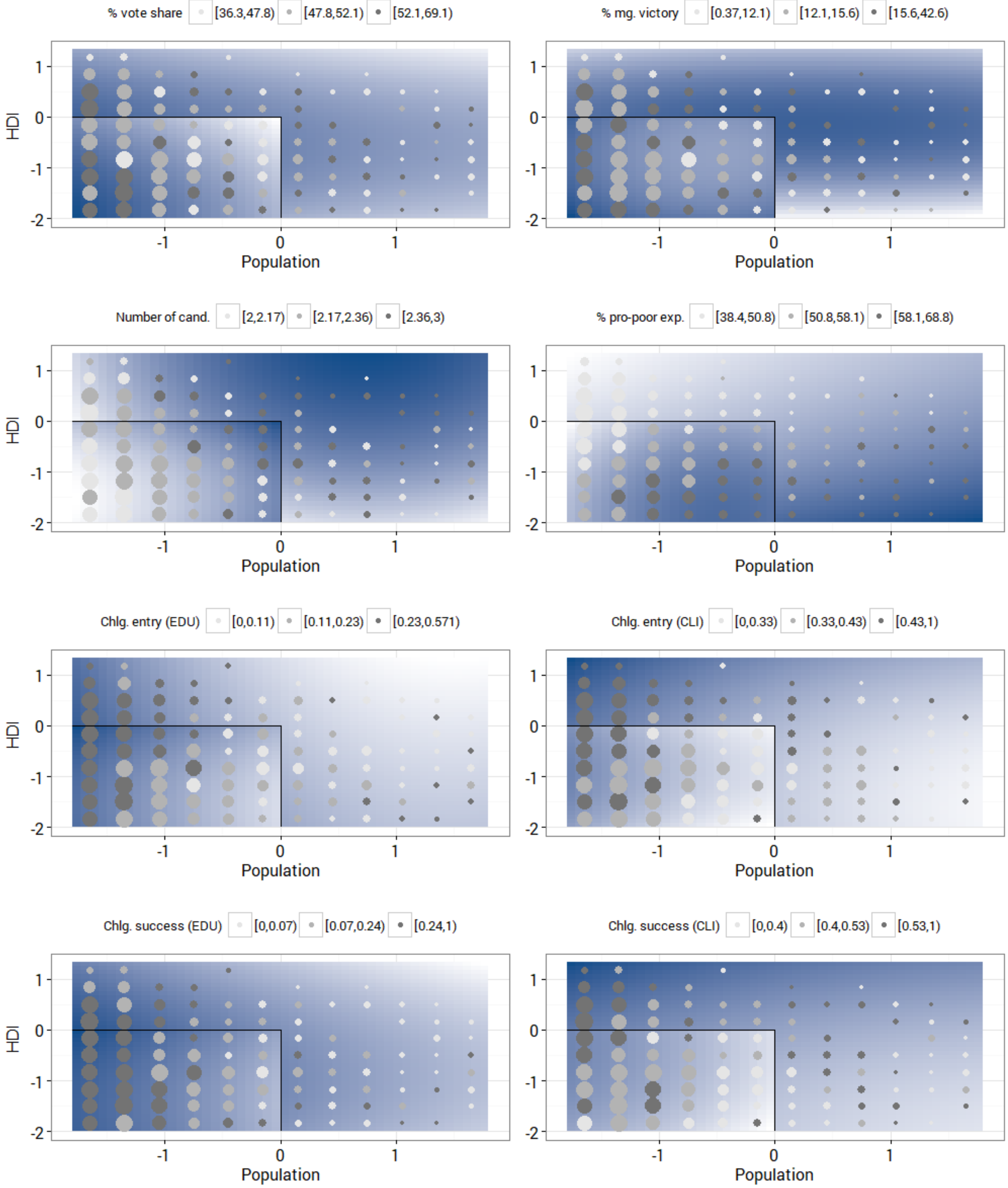
The vertical line represents the treatment frontier (Pop= 30,000 and HDI = 0.7). The points are the average CCT coverage for each one of the bins. The x-axis shows the minimum distance of each observation to the treatment frontier, measured by the lowest value of population and HDI (population and HDI are normalized to a common scale by their standard deviations). The lines are fitted by local linear regression on the unbined data. The grey shades are the 90% confidence level. The full sample includes the entire treatment frontier and all observations within 0.9 distance units from treatment (bandwidth). The Strong IV sample includes all observations with population above 15,000, and HDI above 0.58. The Weak IV sample is the complement set of the Strong IV one (all samples are estimated employing a 0.9 bandwidth).

Figure A.III: CCT Coverage on the Population x HDI plane



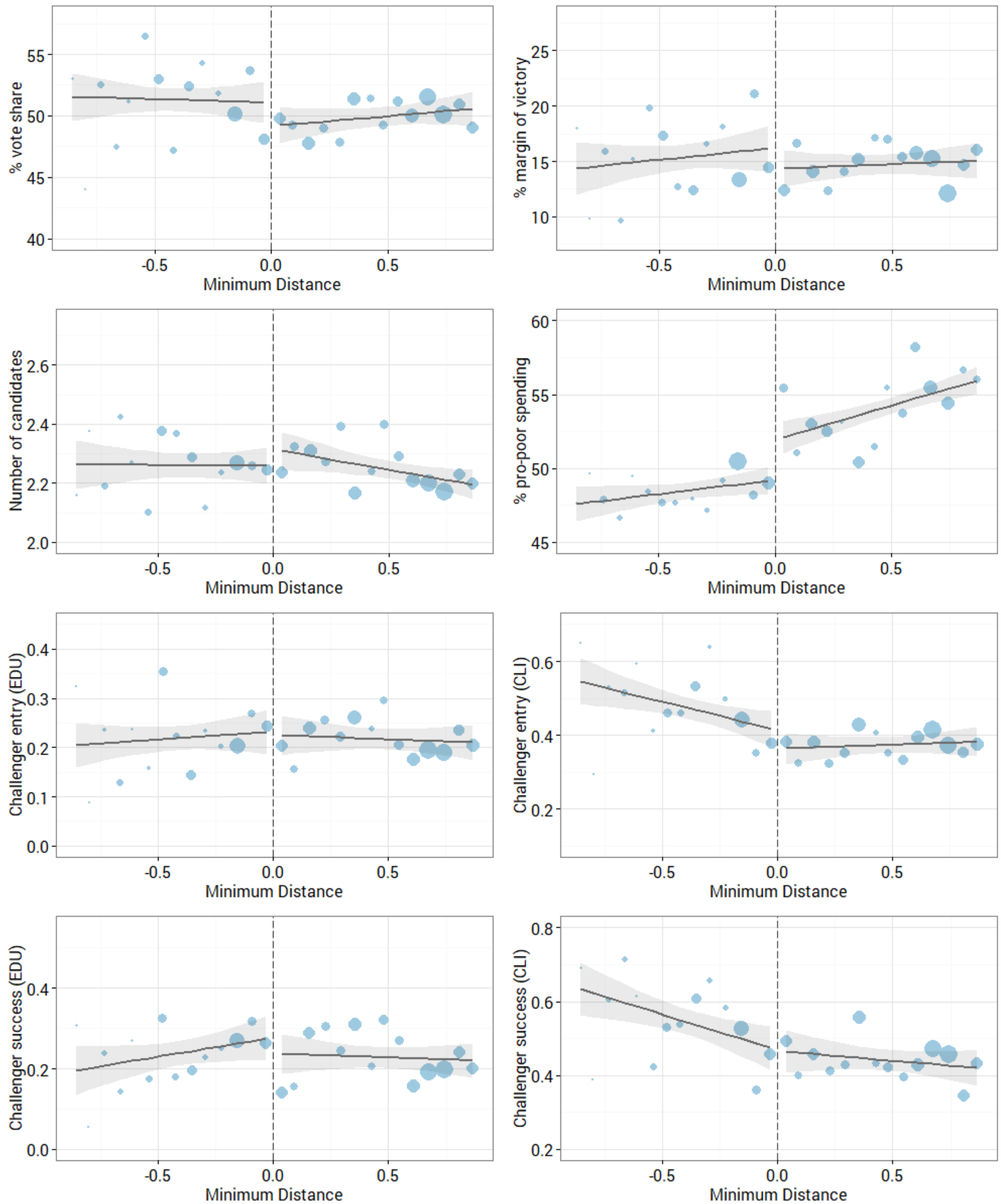
The bottom left quadrant (Pop <= 0, HDI <= 0) are the treated observations. The surface color represents the quadratic fit of the outcome variable on finely spaced grid of normalized population and HDI, for treated and non-treated areas. The size of each point represents the number of observations in each space on the grid, and their color reflects the average value of the outcome variable for those observations. Darker color represents higher outcome values for both the surface and the points.

Figure A.IV: Political Outcomes on the Population x HDI plane



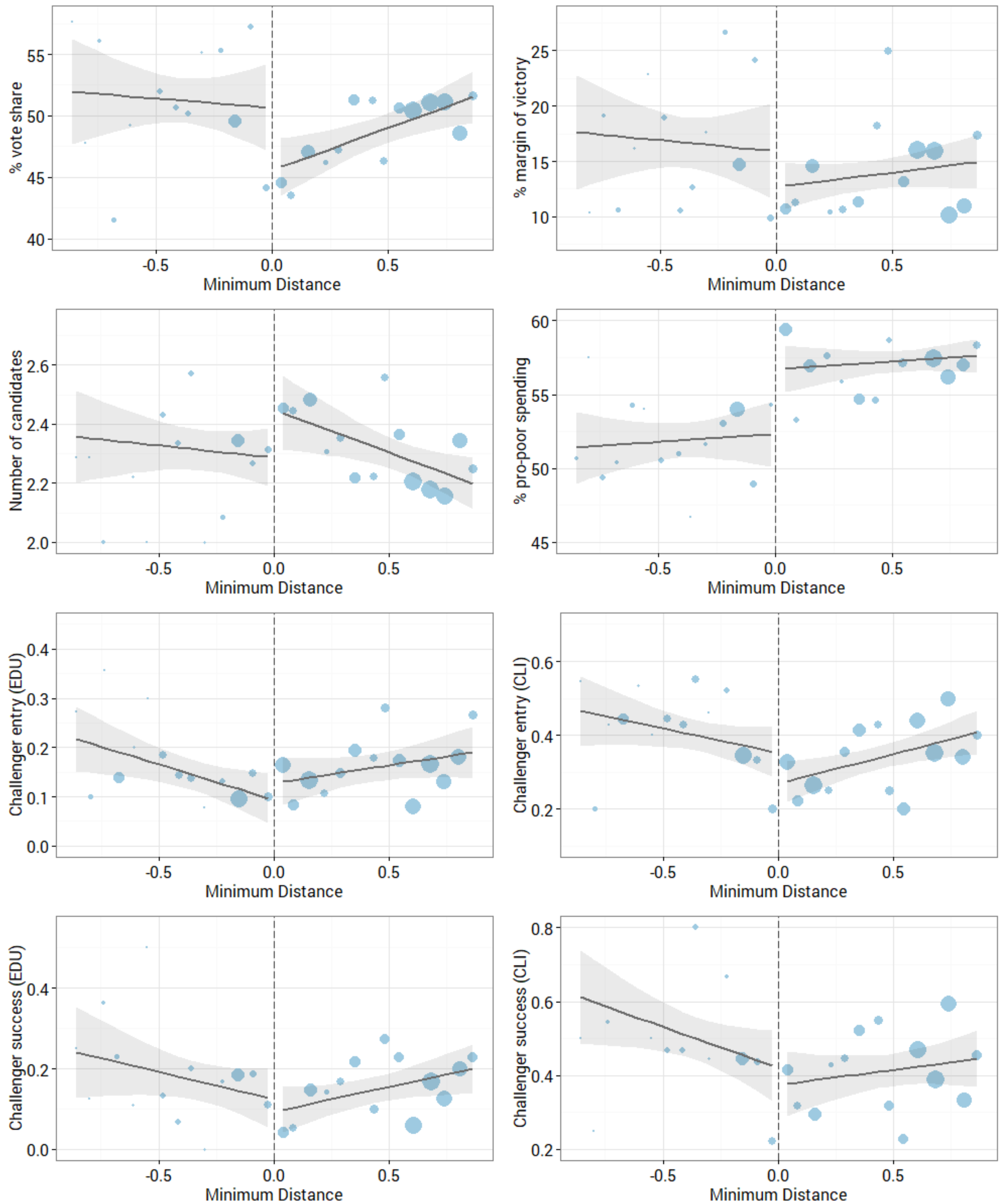
The bottom left quadrant (Pop <= 0, HDI <= 0) are the treated observations. The surface color represents the quadratic fit of the outcome variable on finely spaced grid of normalized population and HDI, for treated and non-treated areas. The size of each point represents the number of observations in each space on the grid, and their color reflects the average value of the outcome variable for those observations. Darker color represents higher outcome values for both the surface and the points.

Figure A.V: Political Outcomes: One-dimension RDD (Full Sample)



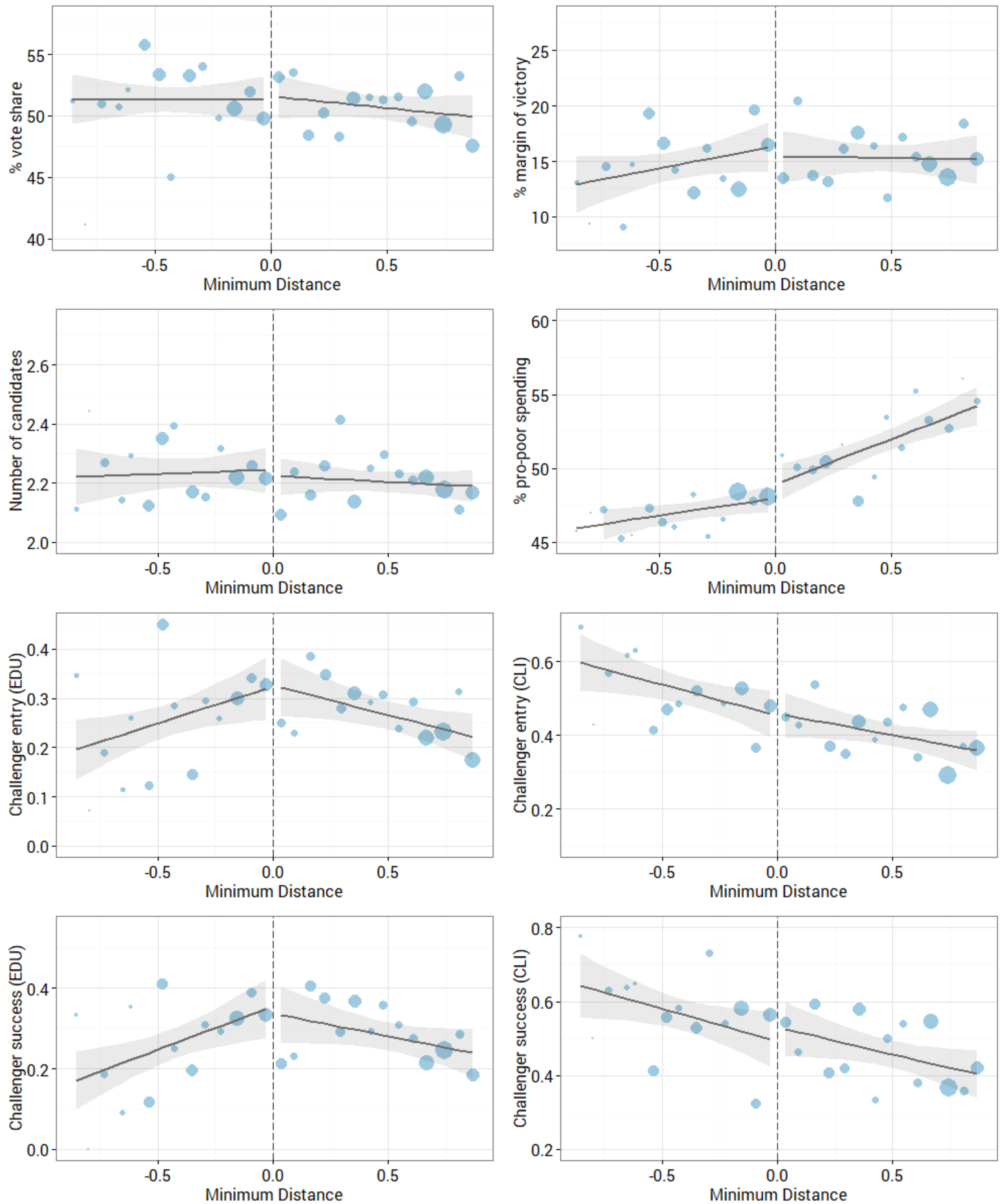
The vertical line represents the treatment frontier. The points are the average outcome for each bin. The x-axis shows the minimum distance of observations to the frontier, measured by the minimum value of normalized population and HDI. The lines are fitted by local linear regression on the unbined data. The grey shades are the 90% confidence level. Estimates include all observations within 0.9 distance units from treatment (bandwidth).

Figure A.VI: Political Outcomes: One-dimension RDD (Strong IV)



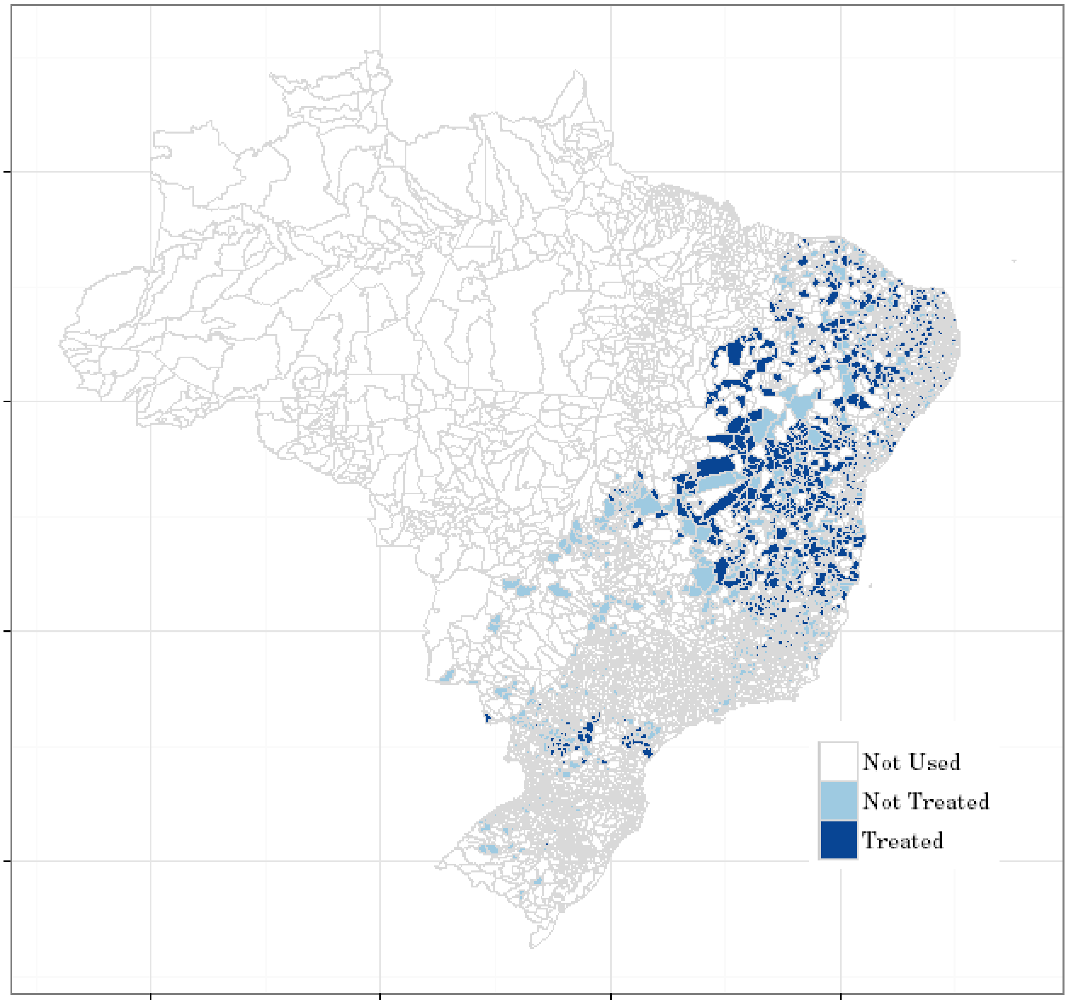
The vertical line represents the treatment frontier. The points are the average outcome for each bin. The x-axis shows the minimum distance of observations to the frontier, measured by the minimum value of normalized population and HDI. The lines are fitted by local linear regression on the unbined data. The grey shades are the 90% confidence level. The sample includes all observations with population above 15,000, and HDI above 0.58. Bandwidth is 0.9.

Figure A.VII: Political Outcomes: One-dimension RDD (Weak IV)



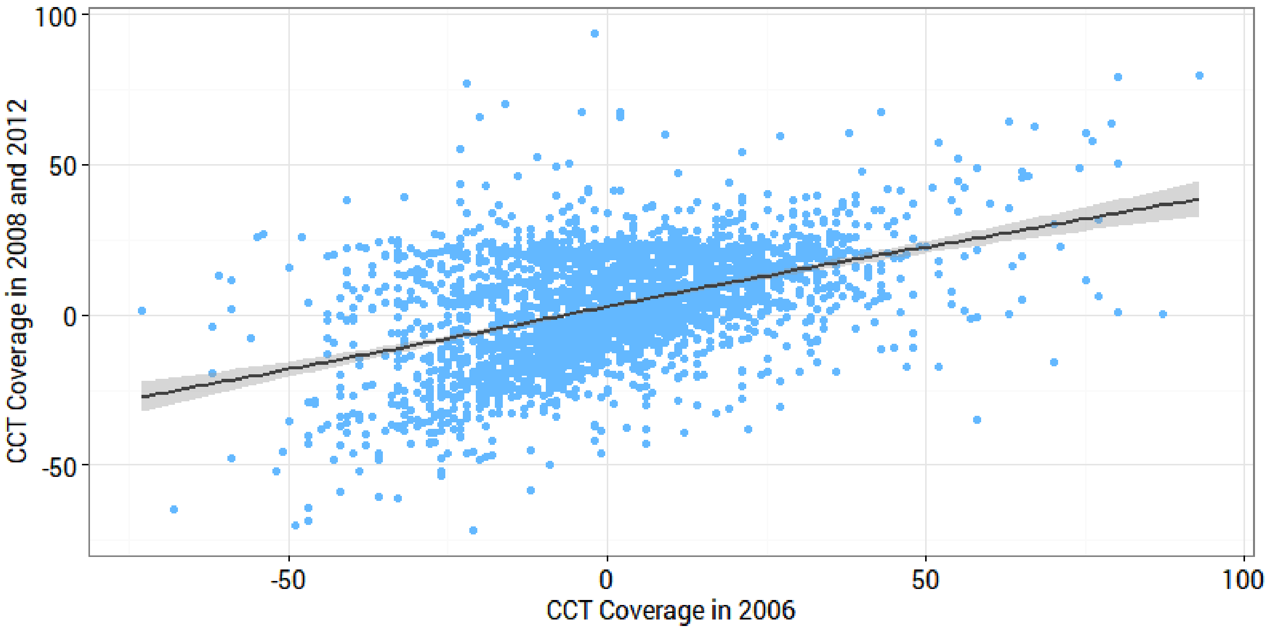
The vertical line represents the treatment frontier. The points are the average outcome for each bin. The x-axis shows the minimum distance of observations to the frontier, measured by the minimum value of normalized population and HDI. The lines are fitted by local linear regression on the unbined data. The grey shades are the 90% confidence level. The sample includes the complement set of the Strong IV sample. Bandwidth is 0.9.

Figure A.VIII: Map of Municipalities in the Sample



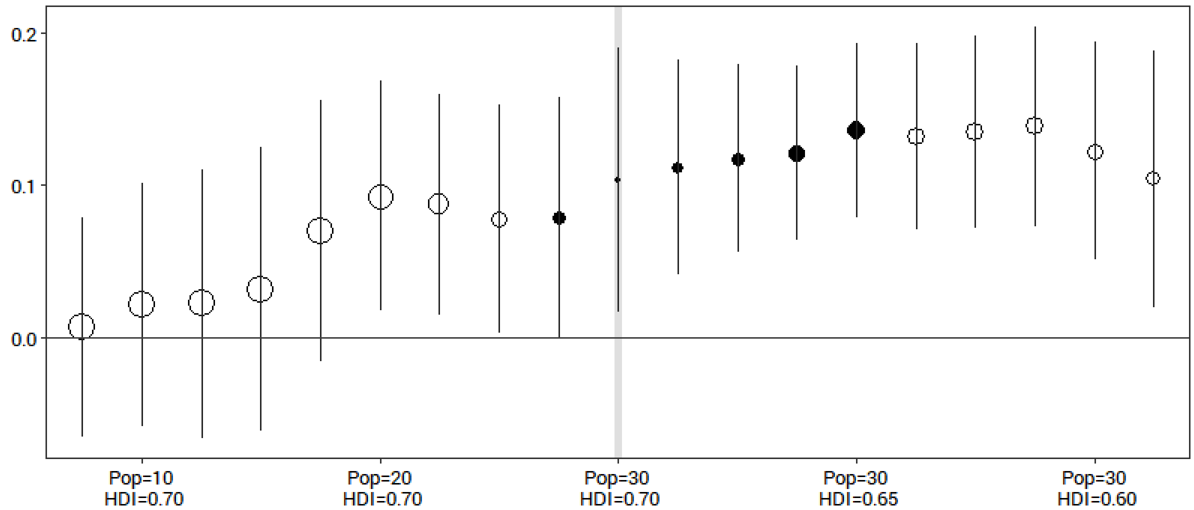
Municipalities in white are either outside of the bandwidth used for the RDD, located in the legal Amazon (high-left side) or have a share of poor population below 25% (most of the southern municipalities). The map shows a total of 1,441 colored municipalities.

Figure A.IX: Correlation between CCT Coverage in 2006 and 2008, 2012



The line represents a liner fit.

Figure A.X: Coefficient for Health Team Visits Along the Frontier



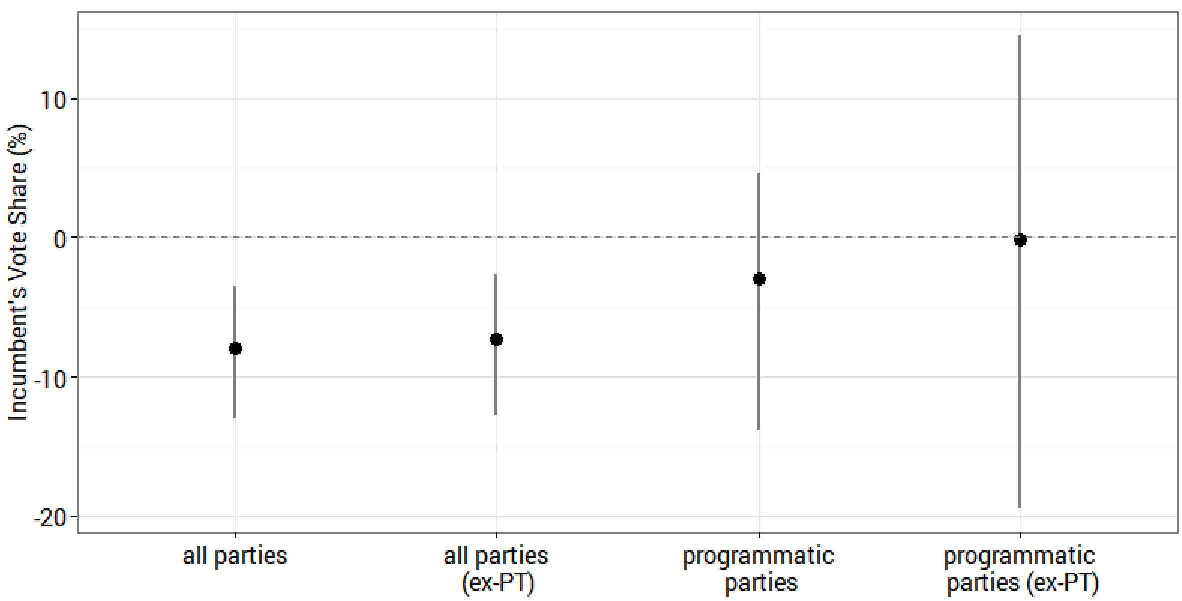
The y-axis shows the conditional ATEs along the treatment frontier, in logs. The left side has HDI fixed at 0.7 and population between 7,500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. Coefficients are from a local linear regression (edge kernel) including year and state effects, and the controls listed in the text. The number of observations in each of the 19 bins are reflected in the size of dots. The dark dots represent the preferred frontier segment for which the instrument is strong.

G PT's Electoral Performance

In this Section, I focus on the estimation results that reflect the electoral performance of the Labor Party (PT). There is a good reason for these specific tests, given that PT was the party that implemented the main CCT program in Brazil (*Bolsa Família* - BF). Accordingly, Zucco (2013) shows that higher BF coverage is associated with more votes for the party in presidential elections, as PT benefited from claiming credit over the program's creation. If this electoral goodwill in fact trickles down to partisan mayors, these politicians should be less affected by the negative 'income-effects' of CCT estimated by this paper.

The first test in shown below in Figure A.XI. Given that it is not possible to obtain reliable estimates of the CCT effects on the incumbent's vote share with a sample restricted to PT mayors only, I take an alternative approach: the plot shows this effect for the entire sample (as in Table 2), and for a sample of non-clientelistic parties only (as in Table 3). It also shows the same effect for these two subsamples, but now excluding PT incumbents from each one.

Figure A.XI: Incumbent's Vote Share (ATE)

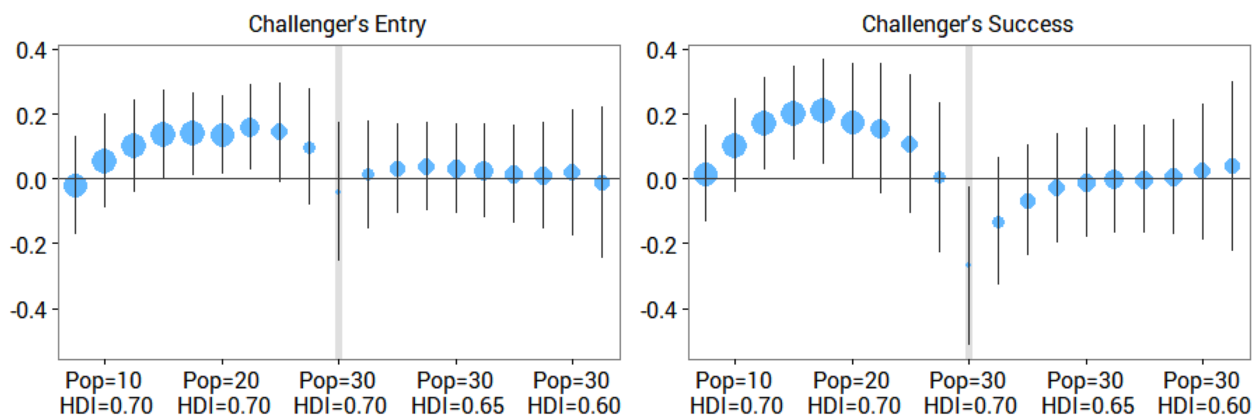


These are the coefficients when the outcome variable is the incumbent's vote share. 90% confidence intervals are clustered by municipality. The coefficients represent the average effect for the preferred frontier segment (6 bins), with a uniform bandwidth of 0.9 standard deviations of population and HDI. Regressions include year and state effects, and the list of controls described in the text.

The plot shows little evidence that PT mayors benefit from an extension of the 'credit-claiming' goodwill that voters attribute to party's presidential candidates. In fact, it seems that the PT incumbents lose slightly more votes than their counterparts from other parties with the arrival of CCTs. The plot shows that, in both cases, excluding PT mayors from the sample maintains the coefficients fairly stable. If anything, an ex-PT sample actually shows slightly higher point estimates, although not statistically different.

The second test examines the number of PT candidates that enter and/or are successful in races against incumbents (similar to the last four coefficients in Table 2). Given that PT is a non-clientelistic party, and this paper already established that clientelistic challengers are less likely to enter races against incumbents, I restrict the sample to challengers of non-clientelistic parties. The outcome variable is then a dummy that assumes value 1 if the challenger belongs to PT. Figure A.XII illustrates the results, and shows that the performance of PT candidates is no different than the performance of challengers from other non-clientelistic parties, both in terms of entry and success in races. This result is additional evidence that local PT politicians are not able to benefit from extended credit-claiming over BF coverage.

Figure A.XII: Entry and Success of PT Challengers



The y-axis shows the conditional ATEs along the treatment frontier. The left side has HDI fixed at 0.7 and population between 7,500 and 30,000. The right side has population fixed at 30,000 and HDI between 0.7 and 0.5875. 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 are reflected in the size of dots.

H Campaign Spending

In this Section, I examine the profile of expenditures in mayoral campaigns, in the context of this study. These results provide complementary evidence for the mechanism proposed in the paper, more specifically, that the effect of CCT on political outcomes happens through a reduction in the efficacy of clientelism.

The theoretical model predicts that CCTs, by reducing the vulnerability of poor voters, also reduce their marginal utility coming from clientelistic exchanges. One of the main problems faced by research on clientelism is that the *quid-pro-quo* between patron and client is rarely, if ever, observed and measured. Thus, most of this paper's evidence for this narrative comes the variation of other observable variables that are related to clientelism in light of the proposed mechanism.

In this spirit, the model also implies that the 'price of a vote' increases in the presence of CCTs, which would lead mayoral candidates to reach less voters with clientelism when the same amount of campaign resources is available. Again, we observe neither these payments nor the actual number of voters being offered clientelistic exchanges. However, we could observe an imperfect proxy, based on the declared campaign expenses by mayoral candidates, which are available for 2004, 2008, and 2012 from TSE. Mayoral candidates in Brazil typically hire '*cabos eleitorais*' to do street-level campaigning during the election period. These 'brokers' are responsible for canvassing support from voters, and in many cases, are the middleman intermediating vote buying transactions. In the data, campaign expenses with *cabos eleitorais* are reported under the line of personnel. What is more, these personnel expenses typically constitute several very small payments to voters (as opposed to companies), which in itself could configure direct vote buying (we obviously never observe the *quid-pro-quo*, that is why they are not illegal).

Using this data, we attempt to identify two pattens: (1) is CCT leading mayoral candidates to reduce the number of voters being paid by their campaigns? and (2) Is the median value of these payments going up when CCT coverage is higher? In Table A.XIV, we show the behavior of a few variables in between electoral periods: the share of campaign expenses with personnel, the mean and median payments, and the number of payees. The variables are measured in their first difference, i.e., as the change in their value in 2008 vs. 2004, or 2012 vs. 2008.[68]

I emphasize that this analysis is not part of the main paper due to the poor quality of the data and low statistical power. It suffers from both very small samples (due to poor data reporting by municipal campaigns, especially in 2004 and 2008), and extreme values, which can be seen when we estimate the effects without the log specification. Its main role is to merely illustrate how other political outcomes are potentially responding the the same mechanism proposed here. Accordingly,

[68] When measured in levels, there are no statistically significant results for these variables. These results are available upon request.

I estimate the effects for two sub-samples, one with all municipalities in which the incumbent could potentially run, and one in which she actually did. We estimate the change in the outcomes using both the log of their ratio and their percentage change.[69]

The Table shows that the share of resources spent with personnel does not change over time with the presence of CCT, which is convenient for the rest of the analysis. Thus, for about the same amount of resources, we see that the median or mean payment to voters seem to increase with higher CCT coverage, while the number of voters being paid decreases, which is consistent with our theory. Although the sign of the coefficients in consistent throughout, only a few of the specifications are statistically significant, most likely due to the small sample. Again, this Table provides more an illustration of other potential effects that conform to the theory.

Finally, I show the plots for the coefficients along the treatment frontier for the log specification. Because this is the one with less statistical power, the coefficients are not statistically significant. Nevertheless, we can clearly see that the heterogeneity in their magnitudes follows the same pattern as the first stage estimation in this paper, providing additional evidence that CCT might play a role in this variation.

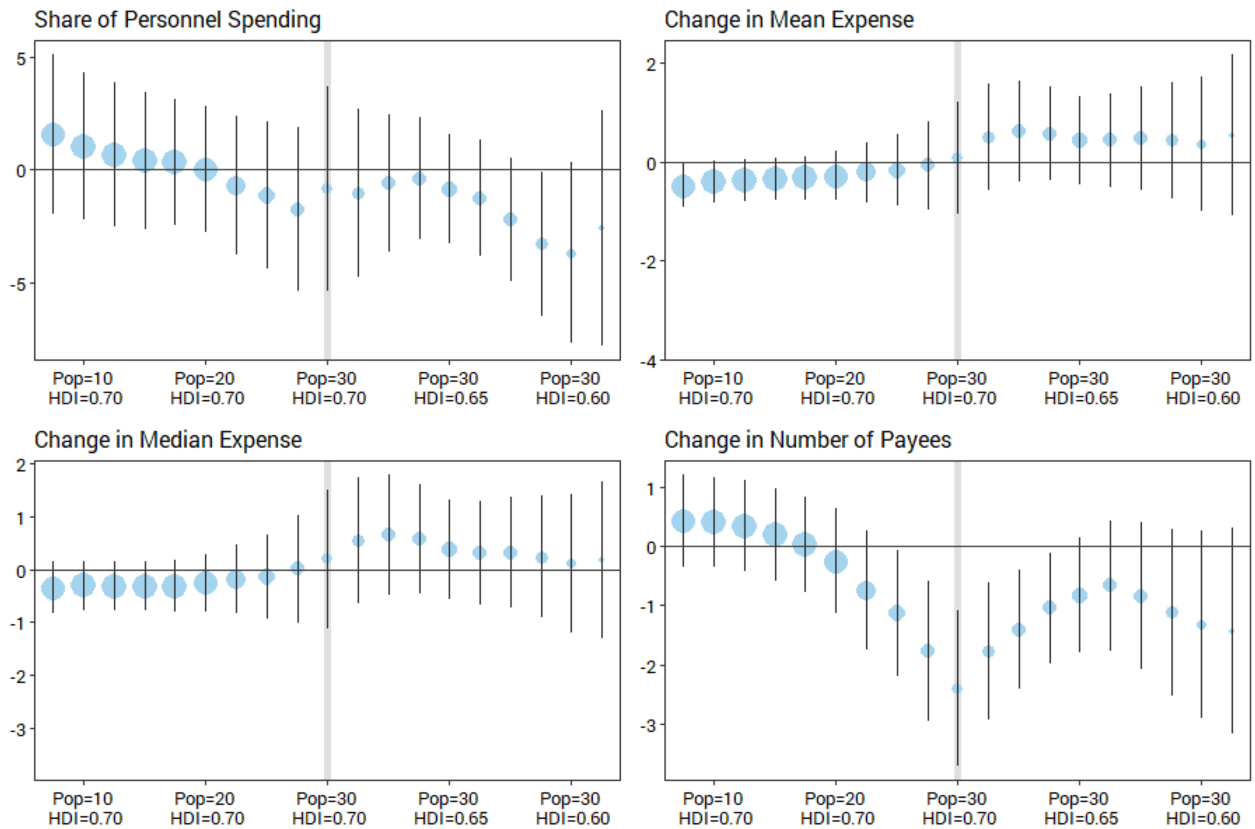
[69]the log specification is as follows: $\frac{\ln Outcome08}{\ln Outcome04}$. It helps to reduce the influence of extreme values in the estimation.

Table A.XIV: Profile of Campaign Expenses by Mayors

	Incumbents **COULD** Run		Incumbents **ACTUALLY** Run	
Dependent Variable	Mean (Pre-Treat.)	Coefficient [90% CI]	Mean (Pre-Treat.)	Coefficient [90% CI]
Share of Personnel Spending	5.354	-0.88	5.56	-0.87
(Change, p.p.)	[4.13,6.76]	[-3.93,2.51]	[4.19,7.26]	[-4.29,3.52]
Obs per bin		339		263
Change in logs				
Change in Mean Expense	0.36	0.38	0.37	0.72
(log of ratio)	[0.23,0.56]	[-0.53,1.57]	[0.21,0.64]	[-0.49,2.08]
Obs per bin		174		130
Change in Median Expense	0.30	0.41	0.32	0.63
(log of ratio)	[0.19,0.44]	[-0.59,1.72]	[0.18,0.53]	[-0.70,2.13]
Obs per bin		174		130
Change in Number of Payees	3.66	-1.49*	4.19	-2.14**
(log of ratio)	[2.18,5.99]	[-2.51,-0.18]	[2.31,7.34]	[-3.45,-0.53]
Obs per bin		174		130
Percentage Change				
Change in Mean Expense	18.54	173.68	26.24	252.46*
(pct. change)	[-14.85,72.54]	[-3.41,452.04]	[-13.77,98.45]	[12.13,562.37]
Obs per bin		174		130
Change in Median Expense	3.46	227.96*	7.47	296.87**
(pct. change)	[-23.01,38.47]	[12.53,527.81]	[-25.29,62.77]	[32.96,652.78]
Obs per bin		174		130
Change in Number of Payees	1204.81	-908.98	2015.88	-3245.65***
(pct. change)	[790,1808]	[-1955,201]	[1067,4422]	[-9695,-1204]
Obs per bin		170		127

Significant at: 99% ***, 95% **, 90% *. Confidence intervals are clustered by municipality and shown in square brackets. Bandwidth is one standard deviation 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, and the controls listed in the text. The Pre-Treat. mean corresponds to the predicted values for a municipality at the discontinuity before treatment.

Figure A.XIII: Heterogeneity of the CATE 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 7,500 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 are reflected in the size of dots.