Political Extremism and Election System: A Study of Confounding of the Results of Single Round versus Runoff Elections under Plurality Rule

Xuan Huynh
andrewjchang1509@gmail.com

Adam Rohde
adamrohde@g.ucla.edu

Andrew Shapiro
alshap1010@yahoo.com

Brandon Thoma
brandon.f.thoma@gmail.com

July 17, 2020

Abstract

The conclusions of Bordignon, Nannicini, and Tabellini (2016) with respect to the effects of runoff elections on political participation and policy volatility provide informative results that positively add to the discussion of the benefits of different electoral systems on societal political behavior and economic stability. The sensitivity of the authors' results survive the most common robustness challenges under a Regression Discontinuity Design, yet unobserved confounding ultimately remains possible but unable to be fully explored due to lack of Italian-specific knowledge. Additional studies are ultimately necessary to validate the results for a more recent Italian electorate, and to study the ability to generalize such results to countries outside of Italy.

1 Introduction

Political extremism arises in any society that fosters a political structure. Moderate policy opinions held by a majority of the population will oftentimes have corresponding radical opinions held by a minority of society. Yet despite a general lack of support within society for political extremes, these political views often find their way into enacted policy. This begs the question: how do extreme policies get implemented over a moderate ones, especially in a society whose structure intrinsically favors compromise? In Moderating Political Extremism: Single Round versus Runoff Elections under Plurality Rule (Bordignon et al., 2016), the authors investigate if certain electoral practices are more susceptible to having extremist voters unduly influence the political process in local governance; more specifically, if a single ballot voting system leads to more policy volatility than a runoff system, and if there is an effect of these systems on the number of candidates running for office.

The authors produce significant results, indicating that a single ballot election has both greater policy volatility and fewer candidates running for election than that of a runoff election, yet their work omits discussion of potentially important confounders that could invalidate these results. Our goal in this paper is to comment on the findings of Bordignon et al. (2016) by running tests of the sensitivity of their results to potential confounders, allowing us to potentially reconsider the validity of their conclusions that possible causal relationships exist between election type and the number of candidates in an election as well as subsequent policy volatility in local governance.

2 Background

2.1 Intuition behind Election Systems

Simple intuition concerning these two different election systems drives the analysis in Bordignon et al. (2016). With a polarized electorate and large moderate political parties, runoff systems should allow for moderate platforms to be implemented that do not need to acquiesce to extremist viewpoints. Under a single round election with genuine voting, extremist candidates can threaten a moderate candidate's defeat by refusing to strike an alliance; under a runoff election, this threat is not credible. Hence we expect intuitively that runoff elections will have more candidates while reducing the influence of extremist parties when compared with a single round election system.

2.2 Theoretical Mechanisms

The authors rely crucially on the results of Bordignon, Nannicini, and Tabellini (2015) who study a model that is geared towards the Italian context. This study predicted more candidates and policy moderation under a runoff system than under a single ballot system. Additionally, the results hinged on the assumptions that the Italian electorate is polarized in the sense that moderate voters hold views are more similar to their extremist counterparts than to opposing moderate voters, and that moderate groups are larger than the extremist groups. These two assumptions are validated by Bordignon et al. (2015) for the Italian context. While Bordignon et al. (2016) goes deeper into the related literature concerning runoff and plurality electoral systems and the effects on political outcomes, for brevity we will leave it to the reader to explore other literature that influenced the authors' empirical strategy and approach.

2.3 Electoral Mechanisms

The focus of Bordignon et al. (2016) is on mayoral elections in Italian municipalities between 1993 and 2007. During this period, municipalities with populations under 15,000 had single ballot elections for mayor. Each mayoral candidate had an associated list of city council candidates. The candidate that receives the most votes becomes mayor, and their corresponding list fills two-thirds of the city council seats, with the remaining seats filled proportionally by votes to losing city council lists. For populations above 15,000, parties present council candidates who would endorse mayoral candidates. In the initial voting round, voters cast both a mayoral vote and party list vote. If a mayoral candidate receives over half the vote, that candidate is elected, otherwise the two candidates with the most votes participate in a runoff election that does not include the party lists. The candidate that wins this runoff round becomes mayor. Losing party lists may re-endorse one of the runoff candidates, with permission. The majority of council seats are awarded to lists that endorse the winning mayoral candidate.

2.4 Policy Volatility in Italian Municipalities

Policy volatility at the municipal level is studied through a main policy tool available to local cities: the business property tax. This tax was introduced in 1993 to fund vital city functions and Italian municipalities were given discretion in determining the rate. Naturally, the tax rate is influenced by party identity, with right-leaning governments favoring lower business taxes. Bordignon et al. (2016) focuses on studying both the intertemporal variance (across legislative terms) and cross-section variance (comparisons across similar populations in the same year) in order to quantify volatility concerning the business property tax.

2.5 Data

To prevent the estimates from being affected by observations far away from the 15,000 threshold, where other policies may overlap with the election system, the authors restricted the sample to municipalities with 10,000 to 20,000 inhabitants. The final sample consists data of 2,027 mayoral terms during 1993-2000 period across 661 towns, which is 10% of all Italian municipalities. The data are from three sources: ANCI (*Associazione Nazionale Comuni Italiani*) for population, geographic and demographic features; The Statistical Office of the Italian Ministry of Internal Affairs for political variables; the Italian Ministry of Internal Affairs for municipal business property tax rate. Population size is available for two different census years: 1991 and 2001. The geographic and demographic features include per capita income, per capita transfer, age, household size, location (north, central or south of Italy, altitude), and area. In addition, the dataset consists of the following political variables: number of mayoral candidates, number of council lists (total number of lists in an election, where each candidate can be supported by more than one list), labor participation rate, term duration (days in office of mayoral term), term limit policy (whether a candidate can run for more than one term), year of election, and business property tax rate. To carry out falsification test, the authors also considered data of the years preceding the reform, meaning from 1985 and onward.

3 Methodology

3.1 Identification under a Regression Discontinuity Design

Since the data naturally has a discontinuity at a population of P_c = 15,000, we require some identification assumptions to employ a regression discontinuity design (RDD). First, there must be no other institutional differences in the elections across the population threshold of 15,000. As discussed in the subsection Data, this is handled by restricting the sample to municipality populations between 10,000 and 20,000, so the effects aren't impacted by other institutional changes such as differences in mayoral pay or the number of city council seats. Second, there should not be manipulative sorting across the population threshold so specific units do or do not receive treatment. This is satisfied through empirical analysis, which will be further discussed in the subsection Manipulative Sorting and McCrary Test. Finally, to identify local average treatment effect, we assume potential outcomes are continuous across the population threshold. Defined for a given municipality i, $Y_i(1)$ is the potential outcome under runoff elections, $Y_i(0)$ is the potential outcome under single round elections, P_i is the population from the most recent census, and Y_i is the observed outcome. Under the above assumptions and notation, the local average treatment effect is $E[Y_i(1) - Y_i(0)|P_i = P_c] = \lim_{P_i \to P_c^+} Y_i - \lim_{P_i \to P_c^-} Y_i$, or the difference between the observed outcomes as populations decrease towards 15,000 and as populations increase towards 15,000.

3.2 Econometric Strategy for Estimating Treatment Effect

The first method used to estimate the discontinuity at P_c is spline polynomial approximations of orders 2, 3, and 4, which includes municipalities with populations between 10,000 and 20,000. Using the dummy variable $D_i = 1$ when $P_i \geq P_c$, 0 otherwise, and the normalized population $P_i^* = P_i - P_c$, the model equation is

$$Y_i = \sum_{k=0}^{p} (\delta_k P_i^{*k}) + D_i \sum_{k=0}^{p} (\gamma_k P_i^{*k}) + \epsilon_i,$$

where p is the polynomial order, γ_0 is the effect across P_c, and the local average treatment effect is estimated as $\hat{\gamma}_0$.

The second method is a local linear regression over observations within a distance h of P_c. The model equation is

$$Y_i = \delta_0 + \delta_1 P_i^* + D_i(\gamma_0 + \gamma_1 P_i^*) + \epsilon_i,$$

where the local average treatment effect is estimated as $\hat{\gamma}_0$. The model is run for bandwidths of $h = 1000, h/2$, and $2h$.

4 Results/Analysis

4.1 Manipulative Sorting and McCrary Test

Prior to further analysis, it is important to check for manipulative sorting across our population threshold. The major concern is that individual cities "manipulatively sort" themselves to be either above or below 15,000 population in order to have their desired electoral system. The initial step is to conduct the McCrary Test for our data and visually inspect the results. The McCrary Test checks if there is a significant cluster of data points just above or below the threshold, but it is important to note that the presence of a cluster alone will not invalidate the results if the cluster is not a result of manipulative sorting. Since the data utilizes census information, we visually inspect the McCrary test results for the 1991 population data and the 2001 population data. The results initially seem worrisome; Figure 1 shows that there is a significant cluster of data points on one side of the threshold, and Figure 2 seems to have an almost significant cluster along the threshold. Fortunately, the new electoral system took place in 1993 and the initial census data is from 1991, so population manipulation for election purposes likely would not occur preemptively for policy that did not begin for two years. Thus the cluster is likely random or the result of something unrelated to the electoral process that was not yet in place.

So while the population was not manipulatively sorted in 1991, this does not mean that 2001 might not have had some manipulative sorting. However, since the 2001 population is dependent on the 1991 population, the McCrary test of the 2001 population data may not be very effective at probing manipulative sorting. To account for this, we look at density plots of the 1991 and 2001 population distributions, and look at the change in density from 1991 to 2001.

We see in Figure 3 that the density plots for the 1991 and 2001 municipal populations are fairly similar. The 1991 data has a greater portion of municipalities between 10,000 and about 12,000, while the 2001 data has more between 12,000 and 20,000, but there are no major differences. Additionally, looking at Figure 4, not only is the difference in density for all population sizes quite small, but the 95% confidence interval (the two gray lines around the blue line) includes 0 for all population sizes, further discrediting the existence of manipulative sorting (at least in one direction).

Now, there is additionally the possibility of manipulative sorting in both directions; that is, some municipality populations are influenced to increase above 15,000 while others are influenced to decrease below 15,000. Of the 661 towns included in the data, only 33 increased and 9 decreased passed the threshold after the 2001 census. While this bidirectional manipulative sorting possibility cannot be ruled out, it is difficult to attribute this mild amount of population change across the threshold to sorting as opposed to population births/deaths or population relocation. We see that more towns increased in population, which parallels the trend of Italy's population increasing between 1991 to 2001. Even in the worst case that it is manipulative sorting, the effect is minute, with fewer than a tenth of municipalities changing; some of these municipalities would have naturally changed independently of sorting. Overall, there is not much evidence of manipulative sorting, and it is more likely that sorting is close to effectively random across the threshold, satisfying an important assumption for a RDD.

4.2 Analysis of Number of candidates

To address the question of how the runoff system differs from the more common single ballot plurality rule, the authors considered the following outcomes: number of candidates, number of parties, number of opposition parties, the ratio of number of parties over number of candidates, and number of mayor's parties. Out of these five outcomes, the first three give significant results. In this section, we will analyze the estimated effect of the runoff system on the number of candidates. The hypothesis that the runoff system will increase the number of candidates has been studied in many papers and yielded conflicting results as the authors note in their paper. In the case of Italy, Bordignon et al. (2016) found supporting evidence for this hypothesis.

This hypothesis is based on the idea that under single round elections, there is greater pressure for moderate candidates to compromise with neighboring extremist candidates and form a coalition, since one more vote than another candidate means winning the election. For example, under sincere voting, if there is a left-wing candidate and a moderate-extremist right-wing coalition, the coalition will likely get votes from both moderates and right-extremists. In contrast, under a runoff system, since there are more moderate than extremist voters, the authors claim that moderate candidates frequently can pass to the second round without forming a coalition. Additionally, with less pressure to compromise with more extreme views, moderates that are closer to the center than extreme may be encouraged to run for office, increasing the number of moderate candidates and parties participating.

4.2.1 Estimation Results and Robustness Values

We first try fitting a lowess model to the entire data and two separate lowess models for data above and below the population threshold. As we move from just below to just above the population threshold in Figure 5, the line appears to be quite steep which implies that there is a jump. This trend is more apparent in Figure 6.

In Table 1 of the paper, Bordignon et al. (2016) study the effect of a runoff system on the number of candidates using six specifications mentioned in the section Methodology. Two models, one of which also controls for covariates, are run for each specification. The following covariates introduced in the section Data are considered in this table: geographic and demographic features, transfer and income per capita, participation rate, age/elderly index, duration of mayoral term, and whether one can run more than one term. The results of these twelve models are replicated and presented in Table 1. All twelve estimated effects are positive and significant at $\alpha = 0.05$. The increase ranges from approximately 1 for *Spline Fourth without Covariates* model to approximately 1.78 for the *LLR(h/2) with Covariates* model. To assess the confounding effect of the alliance restriction, the authors also estimated the effect of a runoff system on the average number of lists supporting each mayoral candidate. They find no significant difference from zero, and hence conclude that alliance restriction has no impact on their interested estimates. After including the covariates in the models, in a consistent fashion, the estimates increase while the standard errors decrease.

`sensemakr` is run on all twelve models of the number of candidates outcome to investigate their robustness against unobserved confounders. Even though `sensemakr` cannot be run with cluster robust standard errors, for these twelve particular models the normal standard errors as well as the t-statistics are relatively similar, in the sense that the difference is not large enough to overturn the statistical significance of the results. The most important outputs of `sensemakr` are the robustness values and partial R^2's. The first one is $R^2_{Y\sim D|\mathbf{X}}$ (the proportion of variation in the outcome uniquely explained by the treatment), which also is how strongly associated with the treatment an extreme confounder needs to be to eliminate the estimated effect if it explains all the residual variance of the outcome. The other two are $RV_{q=1}$ (how strong the equal association with treatment and outcome a confounder must have in order to reduce the estimated effect by $(100q)\%$), and $RV_{q=1,\alpha=0.05}$ (the partial R^2 value of both with the treatment and with the outcome to make the adjusted $1-\alpha$ confidence interval include $(1-q)|\hat{\tau}_{res}|$). For these twelve models, a confounder must explain approximately 5% to 20% of the residual variance in treatment and outcome to entirely eliminate the treatment effect. To make the results insignificant at $\alpha = 0.05$, this amount ranges from 1.2% to roughly 10%. As we move from the treatment-only models to the ones with covariates, all three robust values increase.

We will focus on analyzing the sensitivity of the RDD results from the paper. In reference to Table 1, for the *LLR(h) with Covariates* model, the estimated effect is 1.331 with standard error of 0.396 and t-statistic of 3.366. This translates to $RV_{q=1}$ of 15.4%, implying that a confounder that explains 15.4% of the residual variance in the electoral system and the number of candidates would eliminate the estimated effect. The effect would lose its statistical significance at $\alpha = 0.05$ if a confounder explains 6% of the residual variances. In an extreme case, if there exists a confounder that explains all of the residual variance in the outcome, it would need to explain 2.7% of the electoral system in order to explain away the estimated effect.

In their paper, the RDD model is fit with different bandwidths to check the sensitivity to changes in the window size around the population threshold. The chosen bandwidths are 1000 (h), 500, and 2000. We examined the same model with multiple bandwidths, and report the results in Table 1. It can be seen that as the window size changes, the estimated effect, in both the baseline and with covariates models, remains statistically significant at $\alpha = 0.05$. However, when we try fitting the model with smaller bandwidths, although the estimate does not change sign, the standard error increases. Ultimately, it was found that for bandwidths that are smaller than 500, the results of RDD becomes statistically insignificant. However, this might due to having much fewer observations as the window decreases in size. Using the `rdrobust` package, the optimal bandwidth 1414.631 is obtained. The RDD with this bandwidth yields statistically significant results. This implies that the estimated effect of a runoff system on number of candidates is relatively robust against the choice of bandwidth.

4.2.2 Covariate Balance and Benchmarking

We now investigate the balance of the covariates at the population threshold. The covariate balancing for the inter-temporal models looks at whether the cities just below and just above the 15,000 population cutoff are similar in their characteristics as captured by the observed covariates. The results are shown in Table 2. It appears that, using the t-statistic at $\alpha = 0.05$, we cannot reject the balance hypothesis of any covariate.

Next, we will analyze the sensitivity of the *LLR(h=1000)* estimated effect using the plots in Figure 7 from `sensemakr`. The contours show adjusted estimated effect of runoff system on the number of candidate at levels of hypothesized confounders parameterized by the strength of relationship to the treatment and the outcome. The bounds ("30 x alt_max","30 x active_pop") show the worst confounding that can exist, were we to assume that confounding is 30 times as bad as the geographic altitude or labor participation rate in terms of the residual variance of the treatment and outcome they explain. A confounder that is 30 times as bad as geographic altitude can bring the effect down to its half (0.699) while the effect is almost eliminated if the confounder is 30 times as bad as the labor participation rate. For a larger bound of 50, the effect will be close to zero for geographic altitude, and change its sign if labor participation rate is considered instead. It can also be seen from these plots that even if a confounder can explain as much as 40% of the residual variance from the outcome, it still needs to explain about 10% residual variance in the treatment in order to eliminate the estimated effect. These plots show the two strongest benchmarks: `alt_max` and `active_pop`. Other benchmarks are also included in Figure 8. Overall, the effect of the runoff system on the number of candidates found in the paper seems to be quite robust, relative to these covariates. We will discuss unobserved confounding in more detail below.

4.2.3 Placebo Thresholds

Lastly, as a part of assessing the sensitivity of their results, the authors compute 1,000 placebo estimates at false thresholds for the outcome. Figure 9 reports the cumulative distribution function of the 1,000 placebo point estimates, using a spline third-order polynomial model, normalized with respect to the baseline estimates. 100 stands for a placebo point estimate that is equal to the true baseline estimate at 15000. Thus, we expect to see most of the normalized estimates stay near zero. Only 1.6% of the placebo estimates stay below -100 for number of candidates.

4.3 Analysis of Policy Volatility

In addition to exploring the question of whether runoff elections lead to more candidates, Bordignon et al. (2016) explores whether runoff elections lead to less volatility in policy. To investigate this question with the data available the authors look at two forms of policy volatility: inter-temporal and cross-sectional volatility. The authors focus on volatility in the business property tax rate in a RDD.

Inter-temporal volatility is captured by calculating the variance in business property tax rate across terms in the each city. The authors average annual tax rates for each term, excluding election years to avoid mayoral overlap and any electoral cycle effects. The variance of these per-term average tax rates is then calculated for each city. Thus, the inter-temporal volatility models have one observation per city. For both the inter-temporal models, standard errors are clustered on cities. Robust standard errors are used.

Cross-sectional volatility is captured by calculating the variance in business property tax rate within *groups* of cities of similar population size. Specifically, the cities are sectioned off into groups based on 100-person intervals. Again, the authors use per-term average tax rates for each city as the input to the calculation. The authors calculate the variance of these per-term city-level average tax rates for each term-group. Then they take the average of these variances within each *group*, across terms. This yields the cross-sectional average variance across cities of similar size and terms. Thus, the cross-sectional models have one observation per *group*. Note that for the cross-sectional models, the observations are weighted by the number of cities in each group. Robust standard errors are used.

4.3.1 Estimation Results and Robustness Values

We first look at simple scatter plots with lowess fit lines using cities with all levels of population to gain intuition into the data Figures 10 and 11 plot the intertemporal variance scatters with lowess fit first for the entire range and then separately for the treatments and the controls. Figures 12 and 13 do the same for cross-sectional variance. We see that, for both outcomes, there is a steep downward trend in the lowess that is fit on all the data; there is a discontinuity between the fits on either side of the threshold when we fit separately. This indicates that an RDD might be able to reveal the treatment effect.

Bordignon et al. (2016) ran 24 models related to policy volatility, including both a linear regression on either side of the cutoff and different order polynomial splines. These results are replicated and expanded upon in Tables 3 and 4. Across all models, we find a negative discontinuity in policy volatility above the 15,000 population cutoff. This is true for both measures of volatility. Many of the results are statistically significant. As Bordignon et al. (2016) describes, their baseline inter-temporal estimate "corresponds to a decrease of about 61 percent in the variance of the tax rate just above the threshold." The authors see similar results for cross-sectional variance. With respect to these results, they state that, "in a neighborhood of the threshold, the runoff system decreases the variance of the property tax by about 71 percent." The results are directionally stable across different polynomial orders of spline models (all of which use the entire population range), different bandwidths for linear models, and running with and without covariates. We expand their three bandwidths for the linear models by running intermediate bandwidths in increments of 250 people. We also run `rdrobust` and run the linear models with this bandwidth. These results are broadly consistent with those from Bordignon et al. (2016).

We also run `sensemakr` on these models and include the sensitivity statistics in Tables 3 and 4. The robustness values for the entire treatment effect are generally in the double digits (as low as 8% and maxing out around 30%) for intertemporal variance. The robustness values for significance for intertemporal variance are in the single digits but increase somewhat just below the optimal bandwidth. The robustness values for the entire treatment effect are also generally in the double digits (as low as 10%, most below 40%, and maxing out around 98% (for a model with very few degrees of freedom)) for cross-sectional variance. The robustness values for significance for cross-sectional variance are in the single digits or are zero (due to the estimates not being statistically significant). Running the models with covariates doesn't change this pattern much. It is important to note that for the models with the most narrow bandwidths, we still see fairly high robustness values (specifically, around 30%). As mentioned above, the intertemporal models are run with robust standard errors clustered on city. The cross-sectional models are run with robust standard errors. To run `sensemakr` we rerun the models without robust standard errors; the standard errors are generally comparable. These robustness values should be kept in mind when reading subsection Unobserved Confounders.

4.3.2 Covariate Balance and Benchmarking

We will now discuss covariate balance and benchmarking. Note that the covariates used in the inter-temporal models are averages of the covariates across electoral terms for each city. For the cross-sectional models, covariates are averaged at the group level. It is also important to note that the covariates in the cross-sectional model are limited to regional dummies, area, and altitude.

The main question in a RDD for covariates is: "Do covariates jump at the threshold?" We check this with placebo outcome tests. That is, running the baseline model (bandwidth of 100 people) with each of the covariates replacing the outcomes. The results can be seen in Tables 5 and 6, where we see no statistically significant jumps. Note that we only ran placebo outcome tests on the covariates for which they were sensible, that is, excluding variables such as regional dummies.

We also run `sensemakr` on our baseline model with covariates. We see that fairly large multiples of covariate effects are required to eliminate the treatment effect. But these are lower for significance. Specifically, 5x for the total effect and 3x for significance for intertemporal variance and more than 7x for the total effect and 2x for significance for cross-sectional variance. This can be seen in Figures 14, 15, 16, and 17. Note that 7x the effect of altitude is actually enough to explain 100% of cross-sectional business tax rate variance; plotting further multiples was not possible. We will discuss these benchmarks as they relate to unobserved covariates below.

4.3.3 Placebo Thresholds

As mentioned above, Bordignon et al. (2016) run their models on 1,000 placebo thresholds to see whether the results might be from random chance rather than the treatment effect. They also do this for intertemporal and cross-sectional business tax variance. They do not see more than 5% of these placebo thresholds with larger effects than the 15,000 threshold. See Figure 9 and the above discussion for more detail. Thus, we can conclude that the jumps that we see at the 15,000 threshold are not due to random chance.

4.4 Falsification Test

One concern regarding the credibility of the effect of the political voting system on the number of candidates and political volatility is if the running variable, the population of the municipalities, has a significant effect on the outcomes independently of treatment when it increases past 15,000. A falsification test on the same towns during a pre-treatment period is used to challenge this possibility. Since a parliamentary system was in effect prior to 1993 , the test is conducted only on the number of electoral lists. The results are reported in Table 7. The authors claimed that no significant discontinuity is detected, i.e., before 1993, the number of lists was the same among the municipalities. However, because it is impossible to run falsification test on other outcomes of interest, the estimated effects on those outcomes may be unreliable.

4.5 Difference in Difference

Bordignon et al. (2016) also use a difference-in-difference model to estimate the effect of runoff system on number of candidates, number of parties, number of opposition parties, the ratio of number of parties over number of candidates, and number of mayor's parties. The model is

$$Y_{it} = \alpha_i + \beta_t + \gamma_0 D_{it} + x'_{it}\rho + \epsilon_{it}$$

where α_i and β_t are city and year-of-election fixed effect, respectively, and x_{it} is a vector of time varying covariates. In this model, the identifying variation comes from municipalities that crossed the population threshold between 1991 and 2001. The underlying parallel trend assumption is checked with the falsification test on data of pre-1993.

The results are reported in Table 8. These results are quantitatively similar to the RDD estimated effect and were replicated. The difference-and-difference model is only an extra tool to further confirm the effect found with the main methods mentioned in the section Methodology. Since it is not a main method in their paper, we limit our analysis for this model at replicating and observing the results.

4.6 Unobserved Confounders

While Bordignon et al. (2016) accounts for a variety of observed covariates in the analysis of election system on policy volatility and candidate participation in the electoral process, unobserved confounders may still be present, leading to biased results. As mentioned in the subsection Covariate Balance and Benchmarking, the question of unobserved confounding in a RDD is whether there exists an unobserved covariate that will change across the population threshold of 15,000 and affect the outcomes. Such variables are difficult to identify given that they must make a differential impact on outcomes very close to either side of the threshold. While for larger bandwidths and across wider differences in population many such confounders could be hypothesized, the task becomes more difficult as you approach the threshold of 15,000. Thus, our discussion will comprise of exploring the mechanisms behind some potential unobserved confounders while noting that these become less plausible as the window around the threshold narrows. Further, we note that perhaps the most plausible unobserved confounders are those that might better be called additional treatments. That is, other governmental policies that have an effect on the outcomes that are also triggered near the threshold. Such confounders might have a substantial impact on outcome, much more than observed covariates, and hence could eliminate the effects discussed above, despite our benchmark multiples being large. However, without intimate knowledge of the Italian political system, we are not in a position to thoroughly evaluate the presence of such policies. We, therefore, focus on hypothesizing about the flavor of such policy-oriented potential unobserved confounders that might be present below, without knowing whether such policies actually exist in the present context.

For the outcome of ”Number of Candidates” running for mayor, electoral lists, and opposition lists, a few specific potential unobserved covariates can be suggested:

1. Political Party Funding: Political parties naturally seek to fund candidates at the municipal government level to expand their policy influence. Yet such funding often is directed to more prominent towns (with respect to population size or race competitiveness for example) to provide the best return on political investment. It is reasonable to believe that such dramatic differences in funding could exist across some population threshold for city size, affecting either the number of candidates that can be supported by a party at the local level or how much monetary support each candidate will receive.

2. Presence of Organized Crime: The term *mafia* originated with regards to crime syndicates in Sicily, with the mafia playing a crucial role in the progression of Italian history. Criminal organizations such as the mafia historically have had deep roots throughout all of Italy, with multiple groups vying for control over lucrative illegal industries or kickback schemes. With funding not unlimited, such organizations would need to pick cities in which to project their presence, favoring either larger cities with perhaps a greater likelihood of evading detection by authorities or smaller towns with less likelihood of competition. Their presence could affect the safety and security of candidates for political office, potentially altering the number of mayoral candidates running for office in a municipality.

3. Local Immigration Policy: With Italy's participation in the founding of the European Union, and subsequent increased internal migration across its borders from migrants looking for work, the Italian people have become more polarized over immigration policy. Though we have little familiarity with Italian immigration laws, it is not unreasonable to think that municipal governments have some influence over local immigration policy, such as through issuing new housing permits and registration of new residents. Thus smaller cities might implement policies that, for example, disincentivize immigration into the town to protect the ”town”character”, whereas larger cities may encourage immigration so as to provide additional workers to support the local economy. As such policies may be polarizing for the electorate, they may affect enthusiasm by individuals to run for municipal political office in order to have a say in local immigration policy.

For the outcome of ”Policy Volatility” concerning the business property tax, a few specific unobserved covariates can be suggested:

1. Federal Government Municipal Funding: The Italian federal government likely plays a major role in funding municipal governments in addition to local taxation initiatives. For a variety of reasons such funding potentially varies across towns based on their population size. With successive federal governments potentially having different ideas regarding

such municipal funding, the stability of such municipal funding is questionable. As such, the business property tax is one such method for municipalities to raise revenue without relying on federal government funding, and hence the presence of such funding could affect the business property tax rate or the change in such rate over time depending on the population size of the municipality and its funding needs.

2. Business Registration Policies: The ease of starting a local business, while likely receiving some guidance from the federal government, oftentimes requires strong input from the municipal government with respect to zoning, new business fees, and local regulations. The ease of starting a business likely affects the number of businesses present in a community, which in turn may affect the business property tax rate of those business. For example, cities with fewer businesses may need to tax those businesses at a higher level to maintain business tax revenue, whereas municipalities with large numbers of businesses may be able to lower such a rate without affecting revenue as much. However, changing economic times often affects the number of businesses present in area which in turn can affect the stability of local policies regarding business taxation and property tax.

As mentioned above, the potential unobserved confounders just discussed are of the flavor of the sort of policy-oriented covariates that might effect the outcome and could have a discontinuous change near the threshold of 15,000. We are, however, uncertain of whether such confounding exists in reality. We urge caution either in assuming that no such policy exists or, conversely, in assuming that such confounding is likely, without extensive knowledge of a variety of aspects of Italian culture and politics. However, the evidence presented in this paper suggests that, in the absence of such unobserved confounders or in the event that such confounding is not strong, the observed treatment effects are quite robust to standard RDD robustness analysis.

4.7 Internal and External Validity

With regards to the validity of our results, we will discuss both the internal and external validity.

For internal validity, we already sufficiently covered the issues of unobserved confounding and manipulative sorting, finding that we are confident that neither issue affects the strength of our conclusions. With regards to the pre-treatment period falsification test, as discussed previously, the validity of estimated effect of the runoff system depends on the assumption that the population size alone does not have significant effect on the outcomes discussed in the paper. Remarkably, the authors could only implement falsification test on the number of electoral lists. As a result, their conclusion about the other outcomes can be less robust, or even invalid in the worst scenario. Disregarding this, we focus on a final threat to internal validity: attached voters. If voters who voted for a losing candidate in the first election round in a runoff system subsequently fail to participate in the runoff round then there would be no difference between the single ballot and runoff systems. This hunch is tested in Figure 18, if the second round drop in turnout equals to the number of people who voted for the candidate who lost in the first round, then most of the points would be on the 45 degree line. In fact, most of them stay below the line, i.e., most people who voted for an excluded candidate in the first round still participated in the subsequent round. This implies that there is no issue of attached voters. Hence we are confident that this RDD has high internal validity, assuming the absence of unobserved confounding.

Unfortunately high internal validity comes at the expense of poor external validity. Chiefly, two issues arise for the RDD with regards to external validity. First, the adoption of a runoff system in Italian towns is solely determined by its population size. For other countries where electoral rules are chosen, say, by referendum, the population size has no affect on the treatment but may have a direct effect on political outcomes, and in the worse case, the effect can be large enough to eliminate the effect of electoral rules. Secondly, the RDD displays weakness when generalizing the results over time and across countries. The data is primarily from the 1990s when the electoral system of municipalities in Italy transitioned. While the time period is not necessarily the issue, we need to be careful when considering the applicability of our results to modern political outcomes. Additionally, different countries, with their unique histories, politics, population behavior, and government structure mean that the specific RDD and threshold cannot be generalized outside of Italy; the results from this design may also not necessarily be applicable to other countries.

In the end, external validity of the RDD is sacrificed for high internal validity. The ability to generalize our results across time and country are suspect.

5 Conclusion

To summarize, the authors' conclusions that a runoff election system results in increased numbers of political candidates (mayoral, candidate and opposition lists) and decreased policy volatility (business property tax) survive the most common RDD robustness analyses. While unobserved confounding is ultimately possible, without in-depth knowledge about Italian policy, government, and population characteristics we are unable to confidently assess whether or not such confounding exists and is significant. However, the internal validity of the RDD is high at the expense of low external validity. We encourage readers of this paper to replicate our results and potentially explore alternate areas of interest, some avenues of which are discussed subsequently.

6 Future Analysis

While there are many avenues for future analysis, one are of key interest is that, over the last decade, Italian politics has been upended by the rise of the anti-establishment Five Star Movement founded by Beppe Grillo and the right-wing League led by Matteo Salvini. The rise of these particular political forces has altered the traditional Italian political landscape dominated by center-left and center-right political parties, leading to the formation of new governing coalitions at both the national and municipal level. An analysis that serves to validate or refute the authors' original results as applied to this new political environment is warranted, and may be more relevant to the continuing discussion of the effects of electoral systems on political engagement and policy volatility at the local level. However, without the existence of similar circumstances to allow for a study of the original question using a similar RDD as in the initial study, new methods will need to be utilized to probe the questions of interest.

7 References

Bordignon, M., Nannicini, T., & Tabellini, G. (2015). Policy moderation in single round vs runoff elections: A theoretical analysis. *Unpublished*.

Bordignon, M., Nannicini, T., & Tabellini, G. (2016, August). Moderating political extremism: Single round versus runoff elections under plurality rule. *American Economic Review*, *106*(8), 2349-70. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.20131024 doi: 10.1257/aer.20131024

8 Appendix

Table 1

Summary of Models for Number of Candidates Outcome

					Sensemakr Results				
Outcome: *Number of Candidates*	Est.	S.E.	t-value	Obs.	S.E.	t-value	$R^2_{Y\sim D\|\mathbf{X}}$	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$
Estimation without covariates									
Spline third	1.103	0.382	2.887	2027	0.351	3.14	0.5%	6.7%	2.6%
Spline fourth	1.098	0.487	2.254	2027	0.442	2.482	0.3%	5.4%	1.2%
Spline second	1.532	0.302	5.072	2027	0.263	5.816	1.6%	12.1%	8.2%
LLR(2000)	1.335	0.293	4.56	761	0.295	4.519	2.6%	15.1%	8.9%
LLR(1750)	1.175	0.318	3.701	650	0.323	3.634	2%	13.3%	6.4%
LLR(1500)	1.074	0.341	3.146	530	0.364	2.953	1.6%	12.1%	4.2%
LLR(1250)	0.889	0.401	2.218	441	0.401	2.219	1.1%	10.1%	1.2%
LLR(1000)	1.3	0.408	3.185	364	0.433	3	2.4%	14.6%	5.3%
LLR(750)	1.326	0.491	2.700	256	0.323	3.634	2%	13.3%	6.4%
LLR(500)	1.731	0.676	2.561	175	0.652	2.655	4%	18.3%	5%
LLR(250)	1.524	1.090	1.398	101	0.942	1.619	2.6%	15.1%	0%
LLR(200)	1.772	1.377	1.287	78	0.364	2.953	1.6%	12.1%	4.2%
LLR(150)	2.614	1.494	1.749	65	1.145	2.284	7.9%	25.3%	3.4%
LLR(100)	3.462	1.702	2.035	42	1.489	2.326	12.5%	31.3%	4.3%
LLR(50)	7.894	1.355	5.825	19	1.256	6.284	72.5%	77.3%	63.4%
LLR(optimal = 1414)	1.064	0.352	3.025	519	0.366	2.91	1.6%	12%	4.1%
Estimation with covariates									
Spline third	1.22	0.375	3.25	2027	0.346	3.524	0.6%	7.6%	3.4%
Spline fourth	1.147	0.472	2.428	2027	0.436	2.633	0.3%	5.7%	1.5%
Spline second	1.598	0.297	5.386	2027	0.259	6.162	1.9%	12.8%	8.9%
LLR(2000)	1.418	0.287	4.943	761	0.291	4.867	3.1%	16.3%	10.1%
LLR(1750)	1.242	0.308	4.030	650	0.319	3.892	2.3%	14.3%	7.4%
LLR(1500)	1.077	0.338	3.190	530	0.359	2.996	1.7%	12.4%	4.4%
LLR(1250)	0.901	0.391	2.303	441	0.393	2.29	1.2%	10.5%	1.6%
LLR(1000)	1.331	0.396	3.366	364	0.426	3.124	2.7%	15.4%	6%
LLR(750)	1.358	0.485	2.804	256	0.506	2.686	2.9%	15.9%	4.5%
LLR(500)	1.779	0.677	2.627	175	0.637	2.792	4.6%	19.8%	6.2%
LLR(250)	1.243	1.125	1.104	101	0.935	1.329	2%	13.3%	0%
LLR(200)	1.972	1.430	1.379	78	1.031	1.914	5.5%	21.4%	0%
LLR(150)	2.704	1.921	1.408	65	1.197	2.26	9.3%	27.3%	3.2%
LLR(100)	4.073	1.738	2.343	42	1.591	2.56	19.5%	38.6%	8.6%
LLR(50)	7.810	0.994	7.859	19	1.18	6.618	89.8%	90.6%	70%
LLR(optimal = 1414)	1.105	0.343	3.218	519	0.361	3.059	1.8%	12.7%	4.7%

Table 2

Number of Candidates Balance Table

Covariate	Est.	SE	t	p-value
area	-4.27	19.65	-0.22	0.83
alt_max	-40.60	149.55	-0.27	0.79
end_rev_transf_pc	-15.93	73.56	-0.22	0.823
income_pc	-283.58	484.72	-0.59	0.56
elderly_index	0.02	0.08	0.31	0.76
active_pop	-0.00	0.01	-0.51	0.61
family_size	0.01	0.04	0.25	0.80
duration	20.26	81.63	0.25	0.80

Table 3

Summary of Models for Intertemporal Variance of Business Property Tax

Outcome: *Intertemporal Variance of business property tax*					Sensemakr Results				
Specification	Est.	S.E.	t-value	Obs.	S.E.	t-value	$R^2_{Y\sim D\mid\mathbf{X}}$	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$
Estimation without covariates									
Spline Third	-0.455	0.182	-2.502	575	0.153	-2.981	1.5%	11.8%	4.2%
Spline Fourth	-0.647	0.24	-2.7	575	0.187	-3.461	2.1%	13.5%	6.1%
Spline Second	-0.238	0.14	-1.707	575	0.118	-2.016	0.7%	8.1%	0.2%
LLR (2000)	-0.378	0.16	-2.362	236	0.137	-2.769	3.2%	16.6%	5.1%
LLR (1750)	-0.455	0.174	-2.617	199	0.147	-3.1	4.7%	19.9%	7.7%
LLR (1500)	-0.423	0.2	-2.115	164	0.154	-2.75	4.5%	19.5%	5.9%
LLR (1250)	-0.636	0.219	-2.91	134	0.156	-4.07	11.3%	29.9%	16.7%
LLR (1000)	-0.651	0.255	-2.553	118	0.177	-3.672	10.6%	29%	14.6%
LLR (750)	-0.603	0.303	-1.994	89	0.189	-3.2	10.7%	29.2%	12.2%
LLR (500)	-0.697	0.389	-1.79	59	0.241	-2.889	13.2%	32.1%	11%
LLR (250)	-1.032	0.72	-1.434	39	0.429	-2.404	14.2%	33.2%	5.6%
LLR (Optimal = 1398)	-0.438	0.207	-2.121	158	0.156	-2.811	4.9%	20.2%	6.5%
Estimation with covariates									
Spline Third	-0.45	0.17	-2.642	575	0.153	-2.942	1.5%	11.7%	4.1%
Spline Fourth	-0.614	0.224	-2.734	575	0.188	-3.256	1.9%	12.9%	5.3%
Spline Second	-0.237	0.132	-1.792	575	0.118	-2.004	0.7%	8.1%	0.2%
LLR (2000)	-0.377	0.14	-2.688	236	0.135	-2.795	3.4%	17.1%	5.4%
LLR (1750)	-0.425	0.145	-2.936	199	0.141	-3.012	4.7%	19.9%	7.3%
LLR (1500)	-0.434	0.171	-2.542	164	0.149	-2.919	5.4%	21.2%	7.4%
LLR (1250)	-0.552	0.181	-3.05	134	0.147	-3.748	10.6%	29%	14.9%
LLR (1000)	-0.563	0.211	-2.667	118	0.167	-3.379	10%	28.2%	12.8%
LLR (750)	-0.462	0.221	-2.097	89	0.169	-2.733	9.2%	27.1%	8.1%
LLR (500)	-0.167	0.167	-0.998	59	0.221	-0.753	1.3%	10.7%	0%
LLR (250)	-0.596	0.385	-1.548	39	0.33	-1.808	12%	30.7%	0%
LLR (Optimal = 1398)	-0.403	0.166	-2.425	158	0.149	-2.705	4.9%	20.2%	5.9%

Figure 1

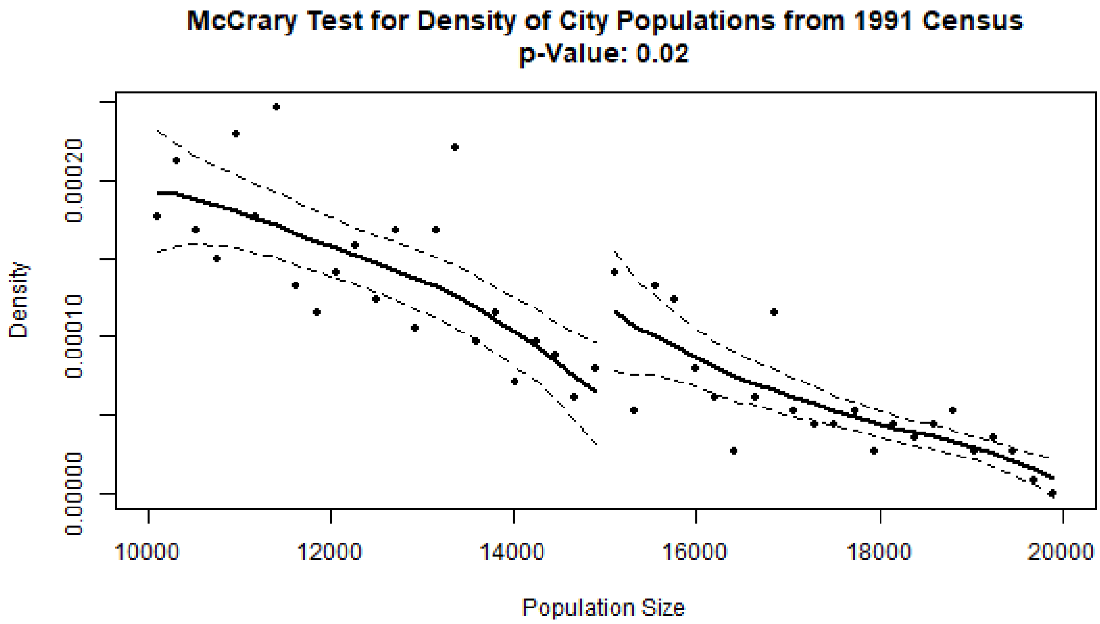


Table 4

Summary of Models for Cross-Sectional Variance of Business Property Tax

Outcome: *Cross-Sectional Variance of business property tax*					Sensemakr Results				
Specification	Est.	S.E.	t-value	Obs.	S.E.	t-value	$R^2_{Y\sim D\|\mathbf{X}}$	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$
Estimation without covariates									
Spline Third	-0.659	0.258	-2.552	92	0.281	-2.343	6.1%	22.5%	3.7%
Spline Fourth	-0.937	0.294	-3.187	92	0.391	-2.394	6.5%	23.2%	4.2%
Spline Second	-0.313	0.201	-1.557	92	0.199	-1.573	2.8%	15.6%	0%
LLR (2000)	-0.443	0.203	-2.185	37	0.189	-2.346	14.3%	33.3%	4.7%
LLR (1750)	-0.524	0.225	-2.333	32	0.215	-2.438	17.5%	36.7%	6.4%
LLR (1500)	-0.648	0.229	-2.835	28	0.233	-2.775	24.3%	42.8%	12.6%
LLR (1250)	-0.671	0.239	-2.801	24	0.274	-2.451	23.1%	41.8%	6.6%
LLR (1000)	-0.694	0.256	-2.71	19	0.364	-1.904	19.5%	38.5%	0%
LLR (750)	-0.155	0.541	-0.287	13	0.492	-0.315	1.1%	10%	0%
LLR (500)	-0.364	0.59	-0.617	9	0.44	-0.827	12%	30.8%	0%
LLR (300)	-1.684	0.088	-19.092	5	0.219	-7.681	98.3%	98.4%	NA%
LLR (Optimal = 1218)	-0.666	0.241	-2.766	23	0.284	-2.345	22.4%	41.2%	4.2%
Estimation with covariates									
Spline Third	-0.627	0.276	-2.273	92	0.282	-2.224	5.8%	22%	2.4%
Spline Fourth	-0.856	0.306	-2.801	92	0.385	-2.225	6%	22.2%	2.5%
Spline Second	-0.352	0.199	-1.767	92	0.198	-1.779	3.7%	17.8%	0%
LLR (2000)	-0.371	0.184	-2.013	37	0.189	-1.966	11.8%	30.5%	0%
LLR (1750)	-0.402	0.178	-2.263	32	0.222	-1.811	12%	30.8%	0%
LLR (1500)	-0.517	0.182	-2.843	28	0.254	-2.034	17.1%	36.3%	0%
LLR (1250)	-0.576	0.203	-2.84	24	0.279	-2.062	21%	40%	0%
LLR (1000)	-0.736	0.274	-2.691	19	0.361	-2.037	27.4%	45.4%	0%
LLR (750)	-0.324	0.564	-0.574	13	0.624	-0.519	5.1%	20.7%	0%
LLR (500)	-0.832	0.278	-2.989	9	0.319	-2.612	87.2%	88.5%	NA%
LLR (Optimal = 1218)	-0.578	0.211	-2.734	23	0.292	-1.979	20.7%	39.7%	0%

Table 5

Intertemporal Variance of Business Property Tax Balance Table

Covariate	Est.	SE	t	p-value
area	6.94	23.60	0.29	0.77
alt_max	9.50	168.26	0.06	0.95
end_rev_transf_pc	3.83	57.25	0.07	0.95
income_pc	452.91	702.09	0.65	0.52
elderly_index	-0.11	0.08	-1.26	0.21
active_pop	-0.02	0.01	-1.64	0.10
family_size	-0.06	0.05	-1.16	0.25
duration	91.11	85.66	1.06	0.29

Table 6

Cross-Sectional Variance of Business Property Tax Balance Table

Covariate	Est.	SE	t	p-value
area	-16.83	15.95	-1.06	0.31
alt_max	-58.63	104.67	-0.56	0.58

Table 7

Falsification Tests on Pre-Treatment Political Outcomes, RDD Estimates

	Spline third	Spline fourth	Spline second	LLR (h)	LLR ($h/2$)	LLR ($2h$)
Panel A. Estimations without covariates						
Number of lists	−0.178	−0.231	−0.121	−0.033	−0.128	−0.336
	(0.449)	(0.544)	(0.349)	(0.506)	(0.668)	(0.365)
Observations	783	783	783	137	67	284
Panel B. Estimations with covariates						
Number of lists	−0.034	−0.202	−0.124	0.069	0.070	−0.244
	(0.348)	(0.419)	(0.290)	(0.351)	(0.502)	(0.292)
Observations	783	783	783	137	67	284

Notes: Election years between 1985 and 1992; municipalities between 10,000 and 20,000. Dependent variable: number of lists, i.e., party lists competing under proportional representation in this pre-treatment period (1985–1992). Estimation methods: spline polynomial approximation as in equation (1), with third, second, and fourth polynomial respectively; local linear regression as in equation (2), with bandwidth $h = 1{,}000$, $h/2$, and $2h$, respectively. Estimations in panel B also include the following covariates: macro-region dummies, area size, altitude, transfers, income, participation rate, elderly index, family size, mayor's duration in office (in days), mayor's second-term dummy. Robust standard errors clustered at the city level are in parentheses.

Table 8: Impact of Runoff on Political Outcomes, Difference-in-Differences Estimates

	Estimations without covariates (1)	Estimations with covariates (2)
Number of candidates	1.186	1.159
	(0.300)	(0.300)
Number of lists	2.303	2.259
	(0.394)	(0.392)
Lists/candidates	0.284	0.282
	(0.170)	(0.170)
Opposition lists	1.787	1.746
	(0.308)	(0.308)
Mayor's lists	0.143	0.152
	(0.181)	(0.181)
Observations	2,027	2,027

Notes: Election years between 1993 and 2007; municipalities between 10,000 and 20,000. Dependent variables: number of candidates running for mayor in the first round; number of lists supporting mayoral candidates in the first round; lists/candidates ratio; opposition lists supporting the losing candidates; mayor's lists supporting the winning candidate. Estimation methods: difference-in-differences specifications with municipality and year-of-election fixed effects, as in equation (3). Estimations in column 2 also include the following (time-varying) covariates: transfers, income, participation rate, elderly index, family size. Robust standard errors are in parentheses.

Figure 2

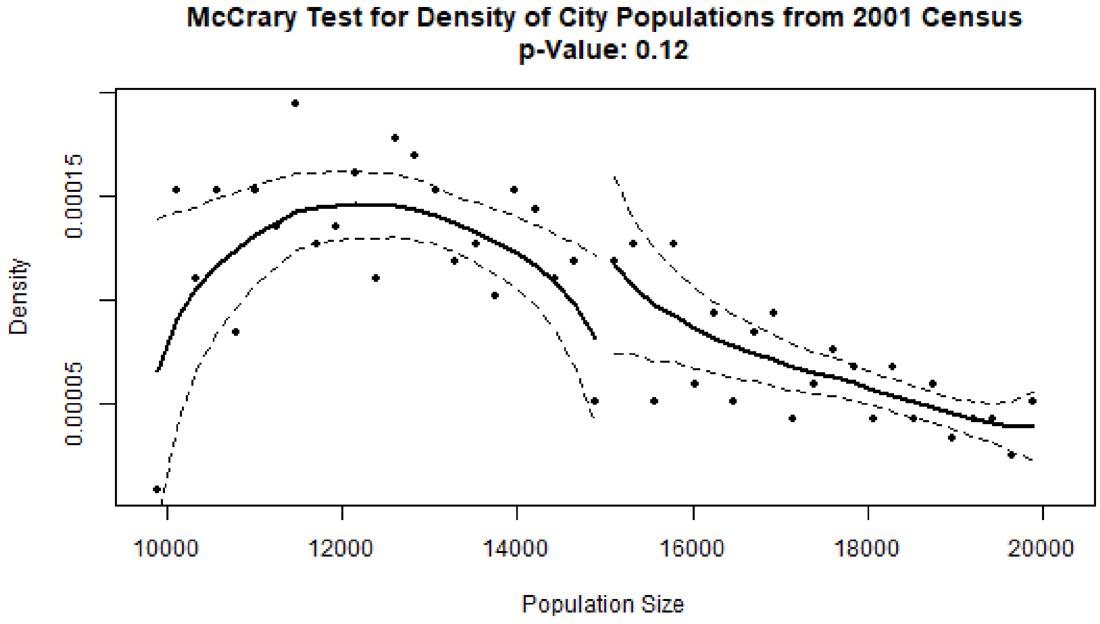


Figure 3

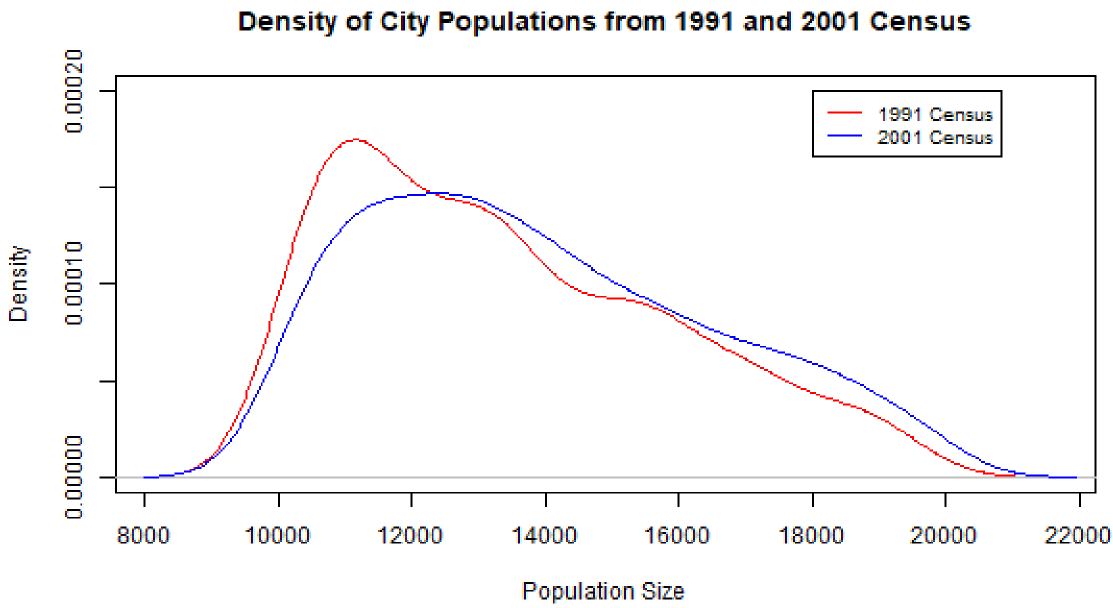


Figure 4

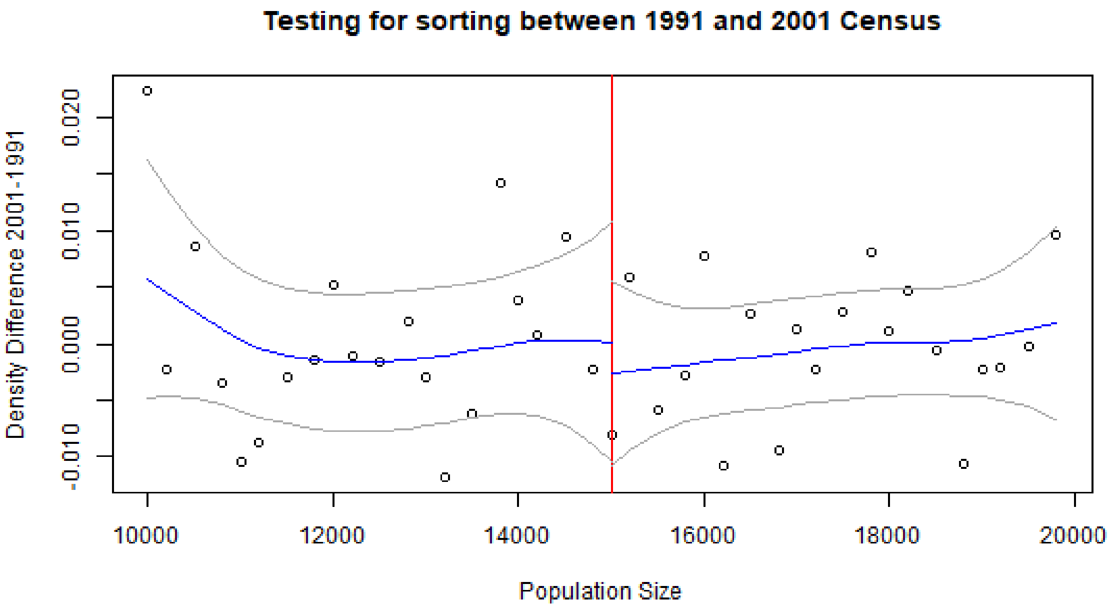


Figure 5

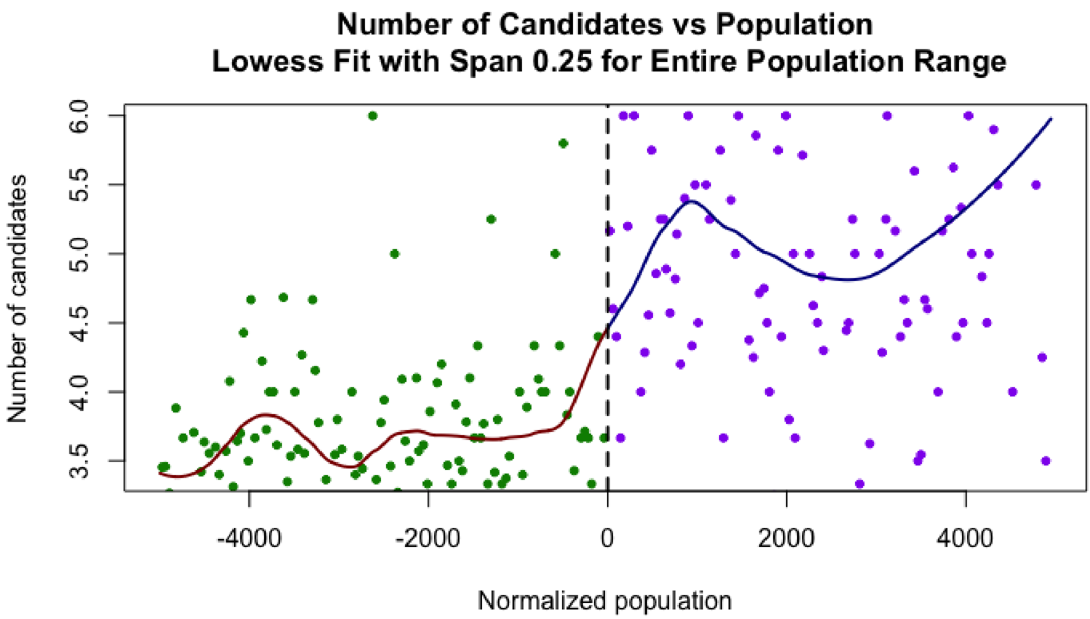


Figure 6

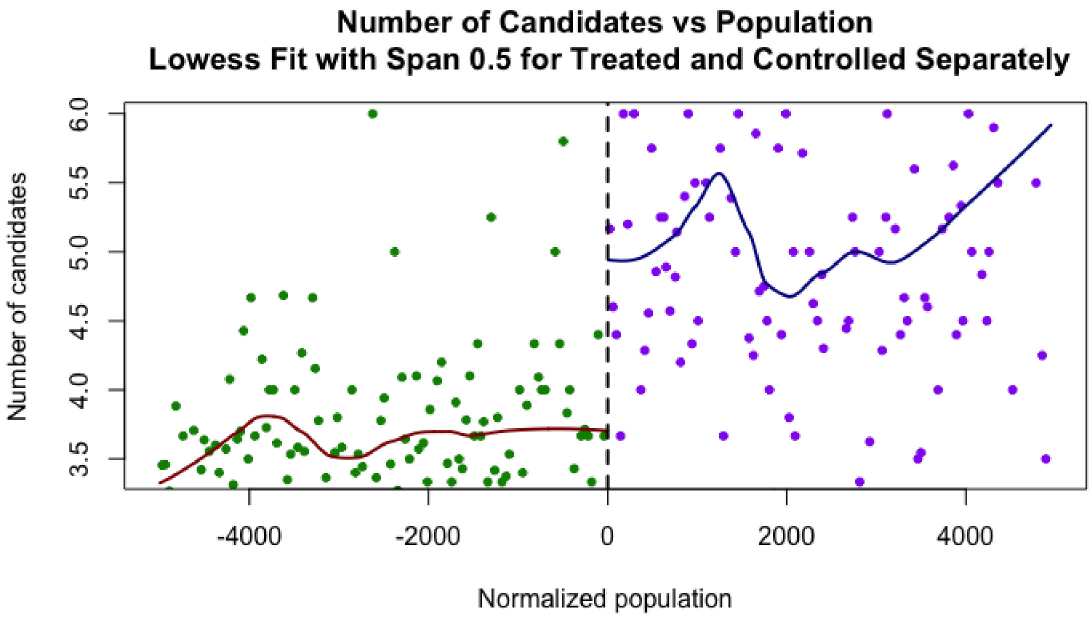


Figure 7: Contour plots showing sensitivity to hypothesized confounding at different bounds

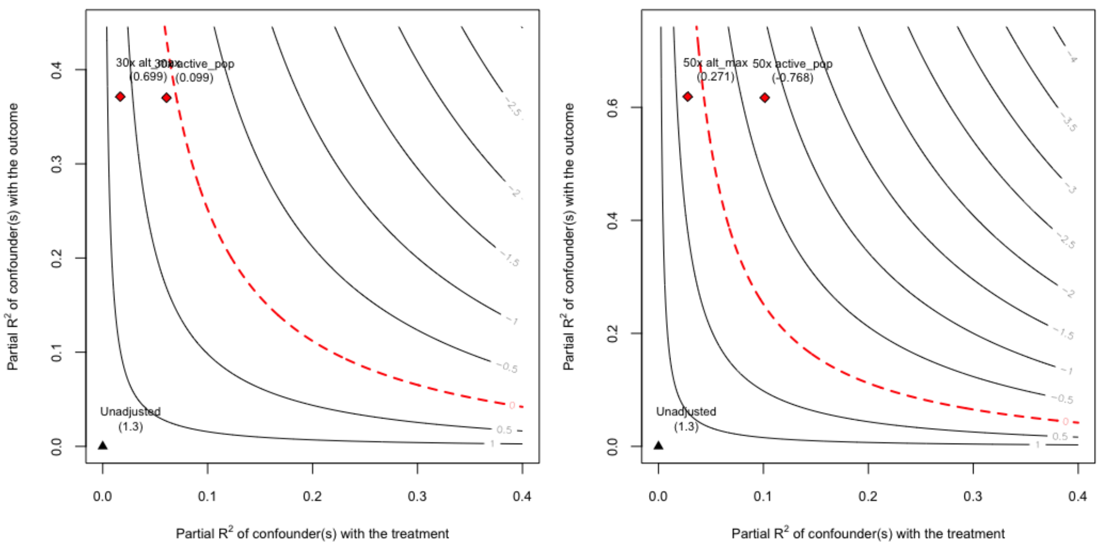


Figure 8: Contour plots showing sensitivity to hypothesized confounding.

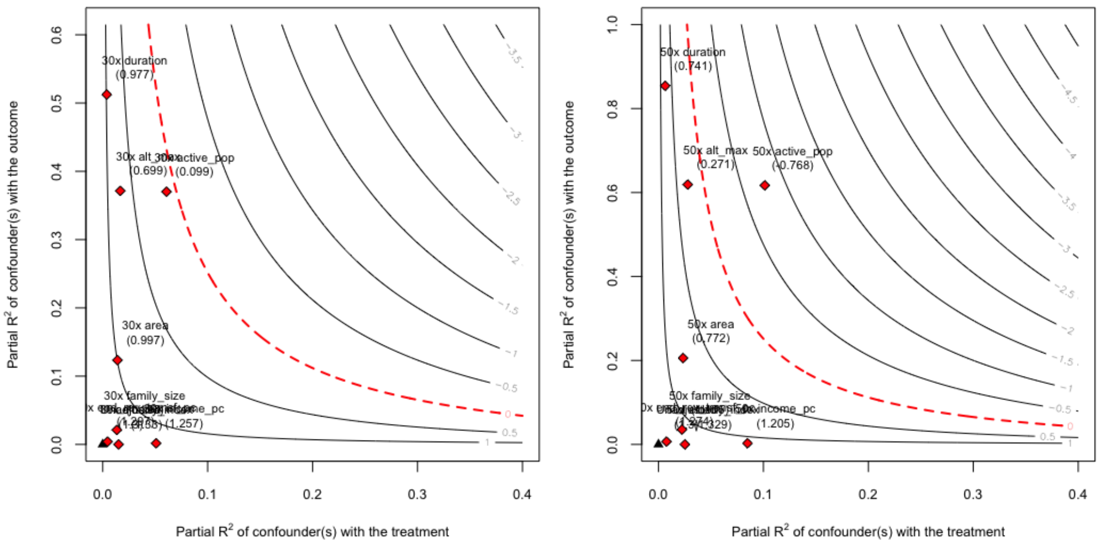
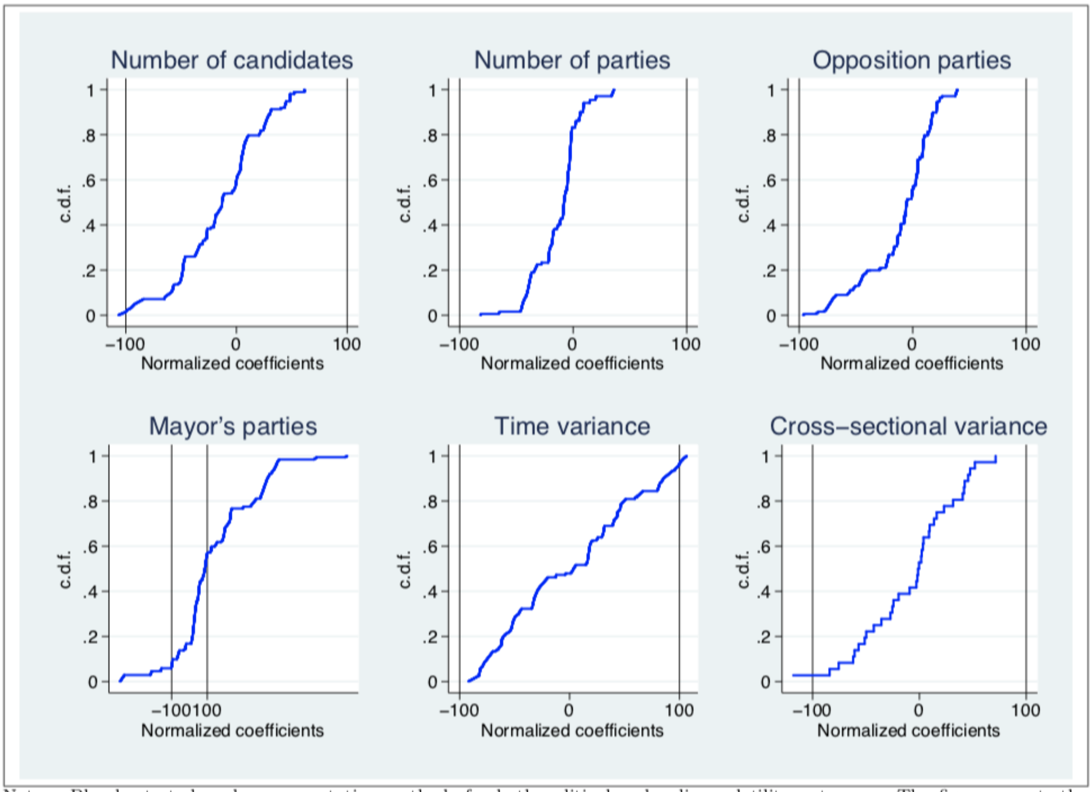


Figure 9: Placebo tests for political outcomes and policy volatility



Notes. Placebo tests based on permutation methods for both political and policy volatility outcomes. The figure reports the empirical c.d.f. of the normalized point estimates from a set of RDD estimations at 1,000 false thresholds: 500 below and 500 above the true 15,000 threshold (namely, any point from 13,501 to 14,000 and any point from 15,501 to 16,000). Only for the cross-sectional variance of the business property tax (where units of observations are 100-inhabitant bins), we consider 80 false thresholds: 40 below and 40 above the true 15,000 threshold (namely, any bin from 10,000 to 14,000 and any bin from 16,000 to 20,000). Each (false) estimate is normalized over the (true) baseline estimate from the paper; that is, a normalized coefficient equal to 100 indicates that the (false) estimate is exactly equal to the (true) baseline estimate. Dependent variables: *No. of candidates* running for mayor in the first round; *No. of lists* supporting mayoral candidates in the first round; *Opposition lists* supporting losing candidates; *Mayor's lists* supporting the winning candidate; *Time variance* (i.e., variance across terms averaged over the entire sample period) and *Cross-sectional variance* (i.e., variance across municipalities averaged over bins of 100 inhabitants) of the business property tax rate. Estimation method: spline polynomial approximation with 3^{rd}-order polynomial.

Figure 10

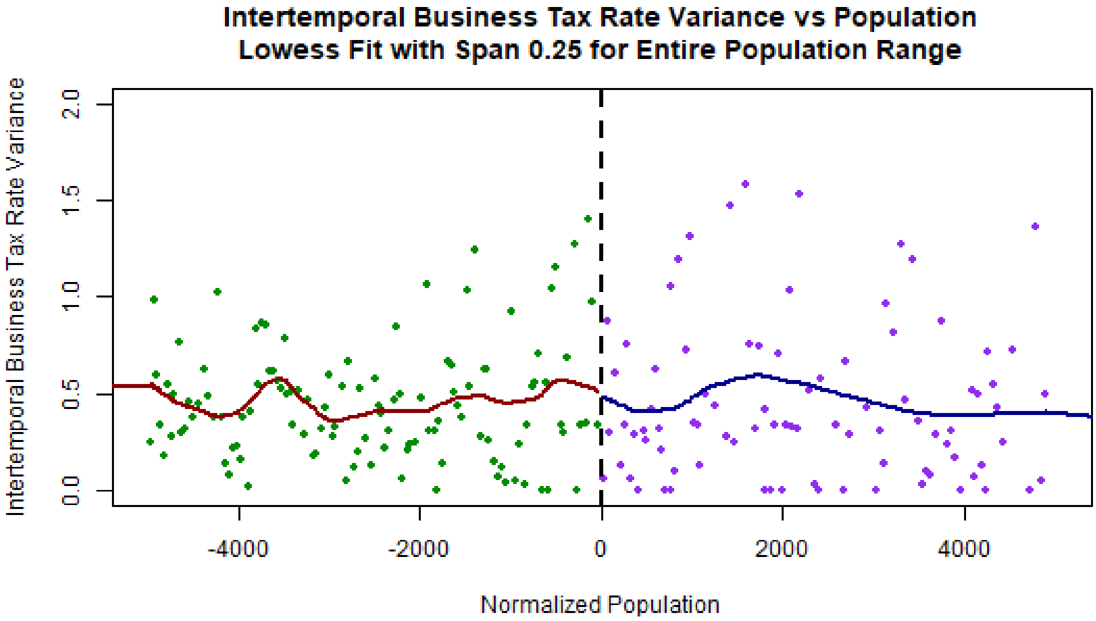


Figure 11

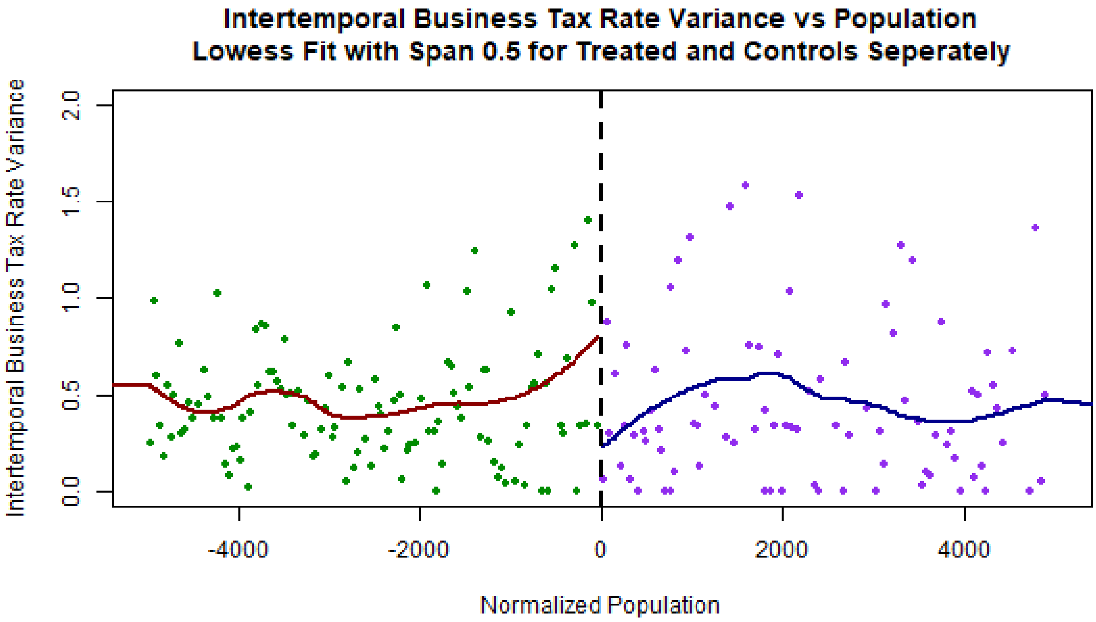


Figure 12

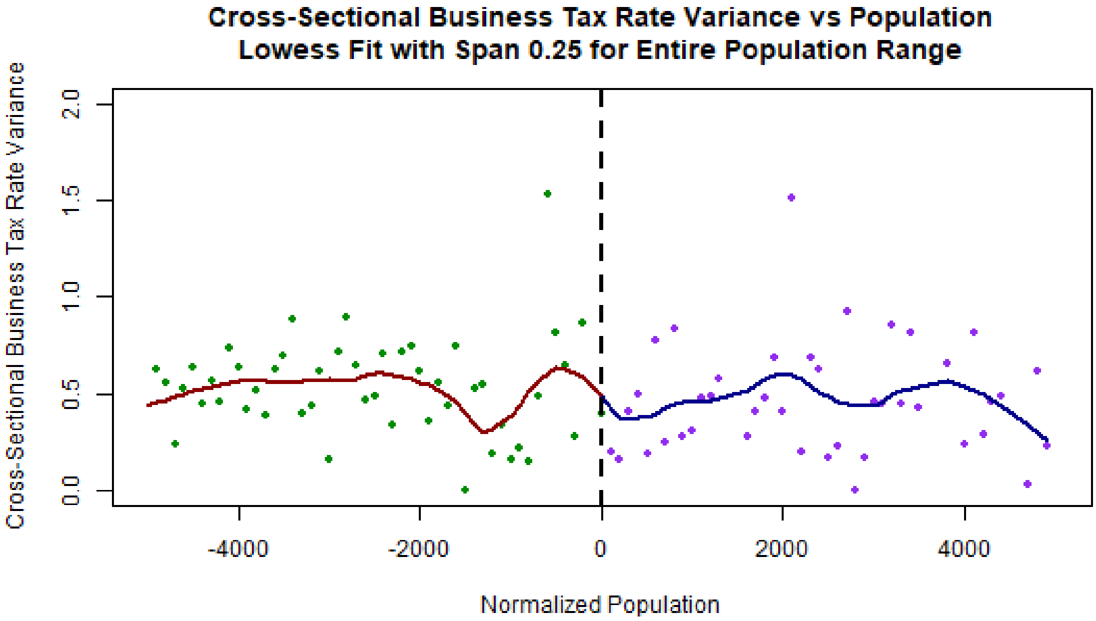


Figure 13

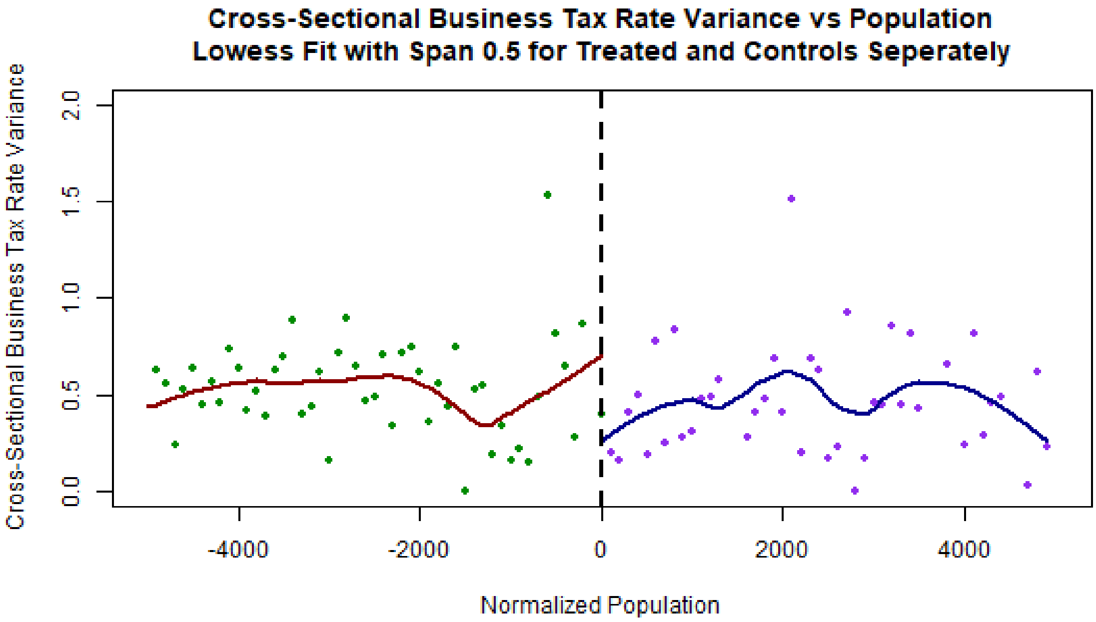


Figure 14

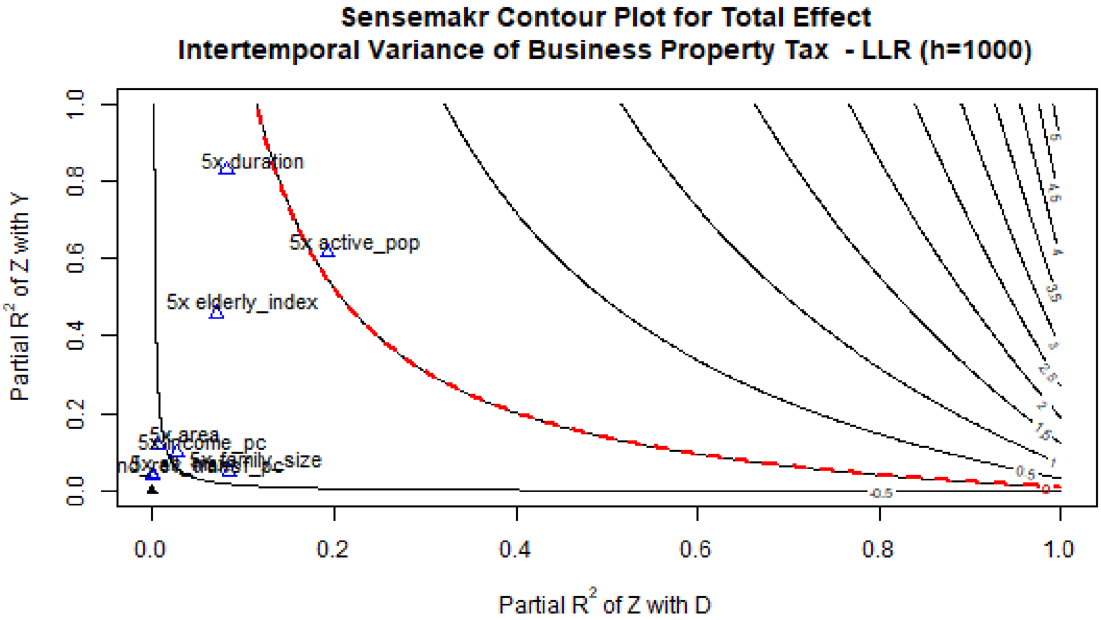


Figure 15

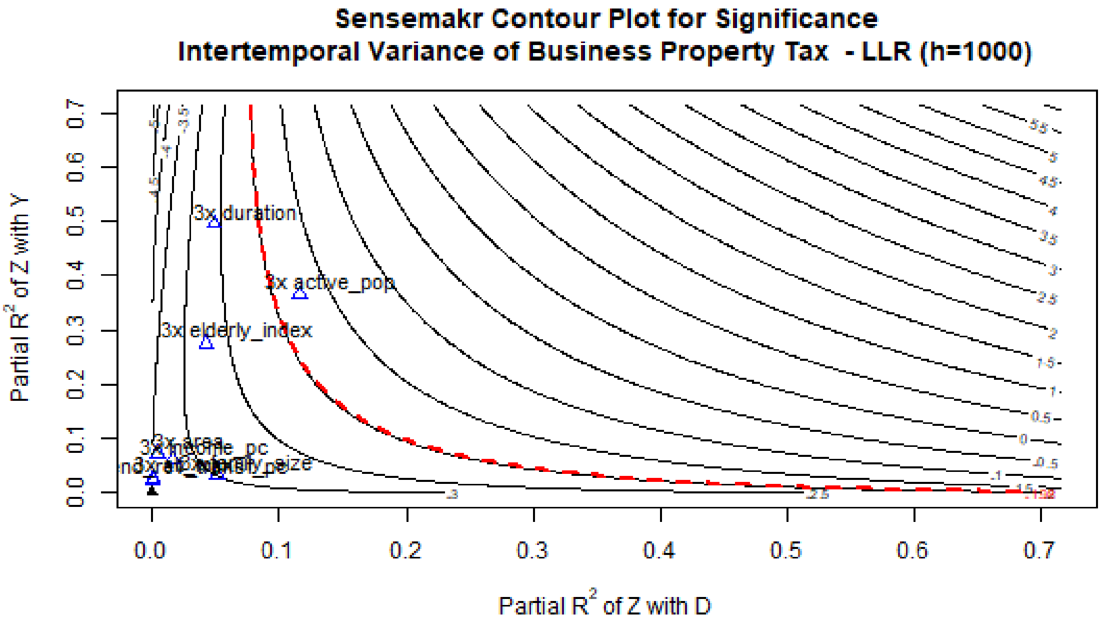


Figure 16

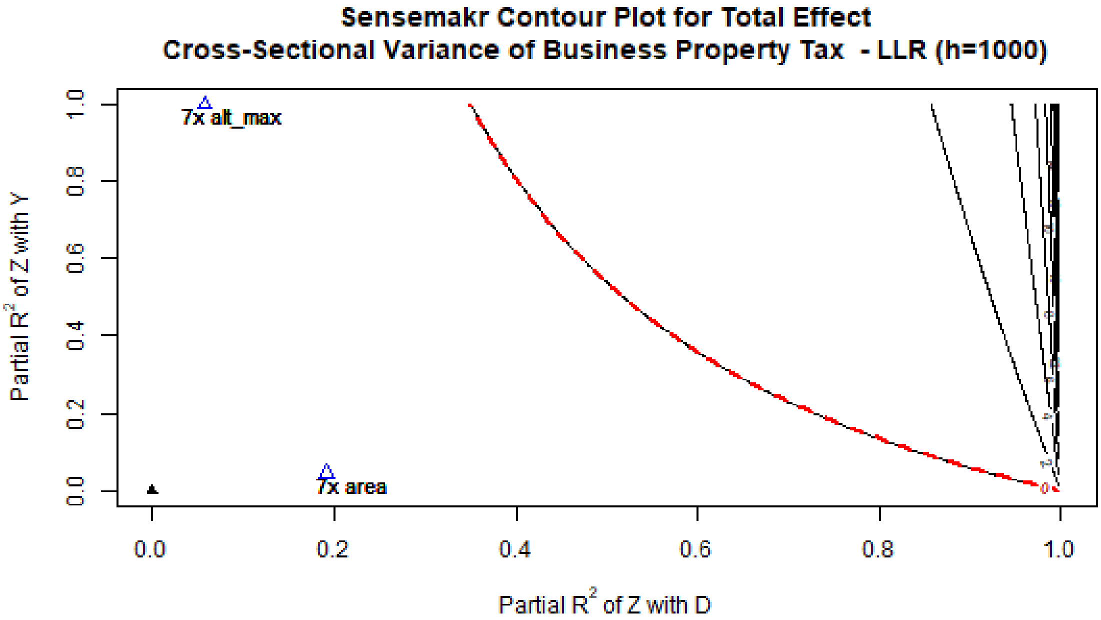


Figure 17

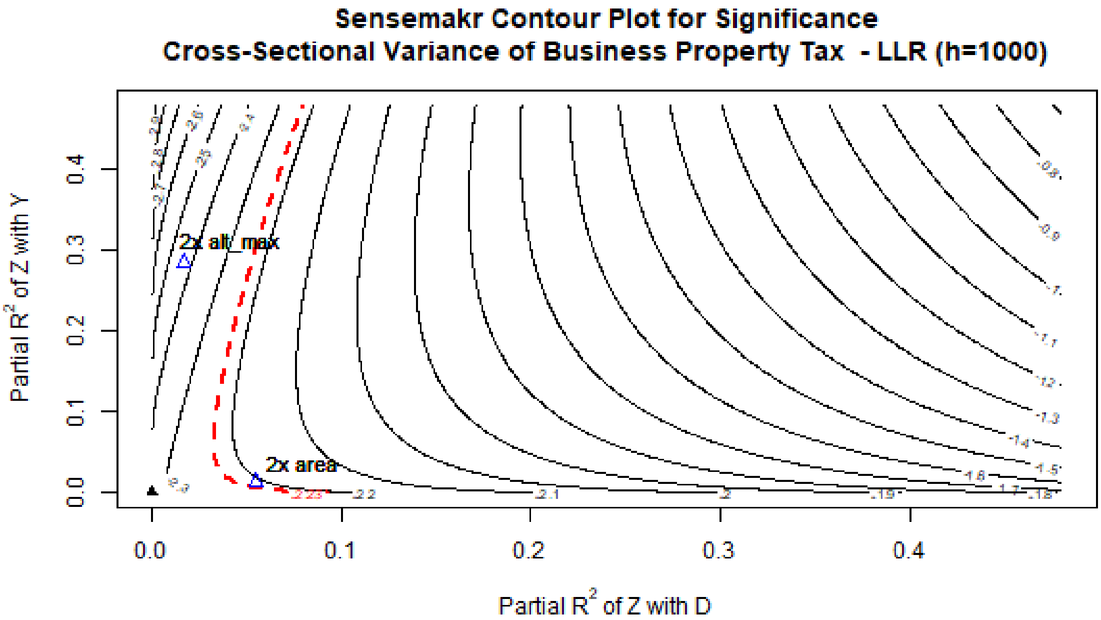
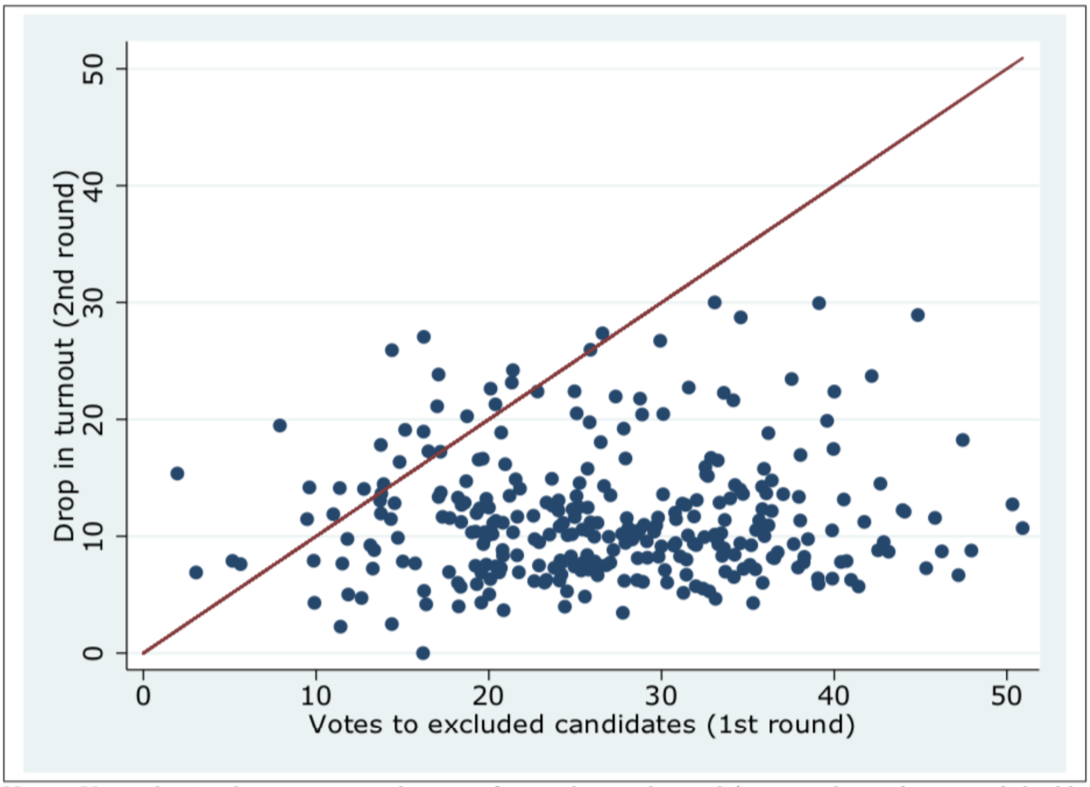


Figure 18



Notes. Vertical axis: drop in turnout between first and second round (expressed as a fraction of eligible voters). Horizontal axis: total votes for the excluded candidates in the first round (expressed as a fraction of eligible voters). Municipalities between 15,000 and 20,000 only.