Polity Size and the Congested Budget: Evidence from Italian Municipalities *

Luca Bellodi† Massimo Morelli‡

Abstract

Once in office, politicians propose policies aimed at winning the support of their constituencies. While this form of political activism increases with polity size – i.e., the number of politicians in government – it can also clash with capacity constraints, leading to a congestion effect whereby politicians' plans are not enacted in practice. With novel data on Italian municipalities, we estimate the causal effect of polity size on a battery of planned and actual budget outcomes. We leverage a reform that introduced a new temporary population threshold where polity size changed discontinuously and estimate local treatment effects with a difference-in-discontinuities design. We document a congestion effect. Municipalities with larger polities have a larger *planned* budget which does not translate into a larger *actual* budget. The congestion effect decreases when bureaucratic capacity is high, proving how administrative capacity can be a binding constraint for politicians' incentives and behavior.

*We are thankful to Lucy Barnes, Rafael Ch, Alexander Coppock, Gianmarco Daniele, Adam Harris, Tim Hicks, Tom O'Grady, and participants at various seminars. We also wish to thank Monica Mantovani, Daniele Formiconi (ANCI), and Cristiano Durantini (Ministry of the Interior) for helpful guidance in surveying legislation and accessing data.

†Bocconi University.

‡Bocconi University, IGIER, PERICLES, and CEPR.

Many theories in political economy show how politicians have strong incentives to be active in office, with the broad aim of advancing their careers in government. Politicians might use public funds to secure re-election (Eslava, 2011), they can propose programs that benefit their district (Weingast et al., 1981), or they can boost their legislative activity to signal their competence to voters (Gratton et al., 2021). These different accounts of activism all imply that, as the number of politicians grows larger, so does the number of items on the agenda. However, the capacity of a government to enact the proposed policies is limited, and hence politicians' incentives may clash with binding capacity constraints.

In this paper, we study theoretically and empirically the interaction between polity size and bureaucratic capacity constraints and how they influence fiscal policy in the context of Italian municipalities. If at the beginning of the term of any newly elected body in two otherwise equal municipalities, the only difference is that one of them has a larger polity (i.e., a larger number of politicians in government), in such a municipality there should be more projects proposed, since everybody in an elected office wants to do something. As a result, a larger polity size should determine a larger planned budget at the beginning of any fiscal year. However, if bureaucratic capacity is binding (i.e., politicians always try to do *as much as possible*), then the ratio of actual versus planned budget at the end of the corresponding period should be lower in the presence of larger polity size. The probability of a planned policy to be successfully implemented by the bureaucracy decreases with the number of policies on the planned budget. Therefore, municipalities with larger polities have, *all else equal*, a smaller probability of implementing all the proposed policies. We derive these two intuitive predictions with a formal model which we include in Section A of the appendix.

An empirical test of this set of predictions would require an exogenous shock to polity size where its effect can be isolated from other confounding institutional features. A reform to local governments in Italy offers such an empirical opportunity. In 2011, the national Italian government passed a set of reforms aimed at controlling public expenditures in years of financial hardship. One of the measures of the reform created a

new discontinuity in polity size for municipalities above and below the 5,000 inhabitants threshold. In fact, the number of politicians in Italian municipalities changes discontinuously at various population thresholds. As a result of the reform, municipalities holding elections between the summer of 2011 and April 2014 – when the reform was repealed – with a census population above 5,000 inhabitants elected 10 councilors and the mayor could appoint up to 4 members of the executive committee. Compared to municipalities in the 3,000-5000 population band, "treated" municipalities elected 3 more councilors and 1 more member of the executive committee. Comparing differences in outcomes at the 5,000 cutoff for municipalities that held elections before, during, and after the reform allows us to disentangle the effect of polity size reform from that of pre-existing differences in politicians' wages at the same cutoff in a difference-in-discontinuities design.

With novel data on fiscal policy, government composition, and bureaucratic capacity of Italian municipalities, we estimate the effect of the reform on expenditures, revenues, and deficits, hence providing a comprehensive picture of fiscal policy. Importantly, we distinguish between what the government *plans* to spend at the beginning of the fiscal year and what the government *actually* spends at the end of the fiscal year, allowing for spending and collection capacity deficits to arise.

We find that municipalities above the 5,000 cutoff in the reform period *plan* to spend more, with expenditures increasing by more than 600 euros per capita, which is as large as 54% of the average in the data. The increase in spending is coupled with an equal-sized increase in planned revenues, leaving deficits unchanged – consistent with the rules on budget justification and coverage. However, the effects disappear when looking at *actual* budget outcomes – namely what municipalities managed to spend and collect at the end of the fiscal year. We document what we call a "congestion effect", whereby larger polities plan to spend more but do not do so in practice. We show that when bureaucratic capacity is high – measured as the share of highly educated bureaucrats – the gap between actual and planned budget decreases. Administrative capacity constraints can therefore affect politicians' ability to implement their agenda, reducing the probability of any policy being implemented.

This paper provides the first attempt at bringing causal evidence to the simple theoretical prediction that a larger polity size with fixed bureaucratic capacity reduces the probability of execution of each planned policy.

Institutional Context and Data

Municipal governments in Italy consist of a directly elected mayor – who appoints an executive committee – and a directly elected local council with quasi-legislative prerogatives. The government term lasts 5 years, hence elections are held every five years, with the precise date being set by the central government. The electoral rule follows an open-list proportional system for candidates to the local council and a plurality system for the election of the mayor. The law defines the size of the local council and a cap to that of executive committees as a function of census population.[1]

Municipalities are responsible for a whole range of services, from municipal police to housing, schooling, welfare politics, and infrastructure development, and every decision with financial implications must have financial coverage as specified in the budget, which lists the type and source of available resources and how the government intends to spend them. Expenditures are financed by municipal revenues, which consist of financial transfers from regional and central governments, local taxes and tariffs, as well as other economic activities of the municipality (e.g., sale and rent of real estate, revenues from public service provision, dividends of publicly owned firms, or traffic tickets issued by local police, to name but a few). The budget cycle starts with the executive committee presenting a budget proposal at the beginning of every fiscal year, in which it outlines the revenues and expenditures *planned* for the next three years (i.e., planned budget). At the end of the fiscal year, municipal governments approve a final budget that accounts for the *actual* expenditures and revenues incurred by the government throughout the year (i.e., actual budget). Importantly, the planned and actual budgets must be approved by the local council, which in turn has a significant influence on the fiscal decisions of the municipality.

[1]Precise size for each population band is reported in Table E.3 in the appendix.

We assembled a dataset consisting of budget data – both planned and actual – and government composition for all the Italian municipalities in ordinary-statute regions (15 out of 20), from 1998 to 2015. We web-scraped budget data from the repositories of the Ministry of the Interior and focus on three key planned and actual outcomes: expenditures, revenues, and deficit, measured as the difference between expenditures and revenues (all per capita). Data on the size of government bodies comes from the Local Administrators Database of the Ministry of the Interior.[2]

Research Design

The size of local councils and executive committees change discontinuously at seven census population thresholds: 3, 10, 30, 100, 250, 500, and 1,000 thousand inhabitants. Since the wage of the mayor, members of the executive committee, and local councilors jump discontinuously at the same population thresholds, comparing municipalities just above and below these thresholds with a cross-sectional regression discontinuity (RD) design would not allow to distinguish the effect of more politicians from that of better-paid politicians.[3] However, in the summer of 2011 the Berlusconi IV government, in an attempt to control public expenditures and as a response to the sovereign debt crisis of those years, passed a law that reduced the size of municipal government bodies and introduced a new threshold at 5,000 inhabitants, which remained into force until April 2014. While before the reform, municipalities with a census population between 3,001-10,000 inhabitants had the same number of politicians (16 councilors and up to 6 members of executive committees), the reform introduced a different size for municipalities in the 3,001-5,000 population band (7 councilors and up to 3 members of executive committee) and in the 5,001-10,000 population band (10 councilors and up to 4 members of the executive committee). The change in size occurred in a staggered fashion, for municipalities became "treated" once they renewed their government bodies in compliance with the new threshold mechanisms between August 2011 and April 2014. The change brought about by the reform is displayed in Figure 1.

[2]Description of data and sample is reported Sections B and C in the appendix.
[3]List of policies changing at population cutoffs reported in Section D in the appendix.

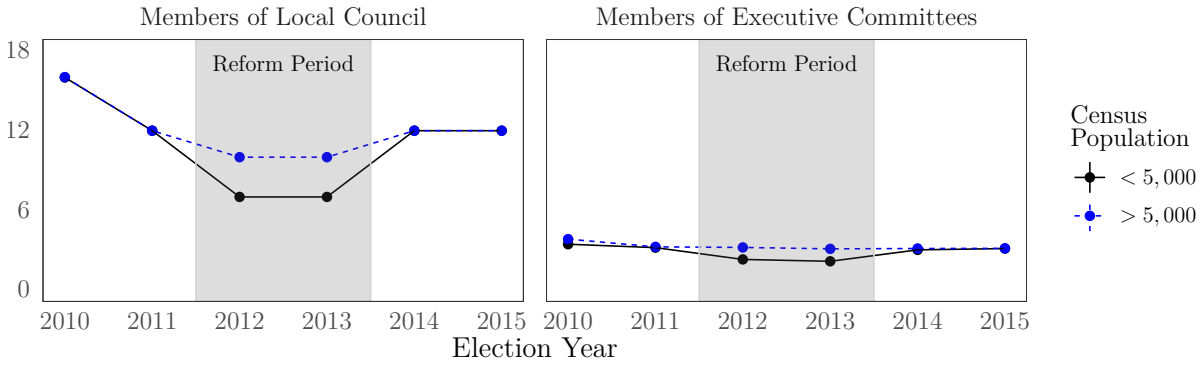

Figure 1: Average polity size for municipalities in the 3-10,000 population band. The decrease in 2011 is the result of a different reform which did not affect the threshold mechanism.

This temporary reform allows us to compare changes in fiscal outcomes at the cutoff for municipalities that held elections during three different periods: before August 2011 (pre-reform period), between August 2011 and April 2014 (reform period), and after April 2014 (post-reform period). The differences in the outcomes at the cutoff for the pre- and post-reform periods represent the effect of wage policies alone, whereas differences for municipalities in the reform period represent the compound effect of wage policies and larger polity size. By looking at the difference in discontinuities in the reform–pre-reform and reform–post-reform periods we can thus isolate the effect of polity size from the confounding effect of wage policies (Grembi et al., 2016; Eggers et al., 2018). To ensure that the reform to the size of government bodies is the only "treatment" that changes over time at the 5,000 cutoff, we limit the analysis to the calendar years 2013-2015, for a fiscal rule aimed at controlling budgetary balances for municipalities above the same 5,000 cutoff was in place between 2001-2012. Because municipalities remain treated or untreated for the entire 5-year-long government term (until new elections), in the calendar years 2013-2015 there are municipalities belonging to each of the three different groups.

The Difference-in-Discontinuities Estimator

Municipalities hold elections in three time periods, $T_t = 0, 1, 2$, which represents periods before, during, and after the reform, respectively. Under the reform, in $T_t = 1$, the RD estimator identifies the effect of both wage policies W_i and the number of politicians P_i on the outcome Y_i. The compound estimand is given by $\alpha_{RD(c)} = \lim_{\epsilon \downarrow c} \mathbb{E}[Y_i | X_i = \epsilon, T =$

$1] - \lim_{\epsilon \uparrow c} \mathbb{E}[Y_i | X_i = -\epsilon, T = 1]$. Grembi et al. (2016) show that, under additional local assumptions, information on the periods without the compound treatment ($T_t = \{0, 2\}$) allows to isolate the effect of P_i from that of W_i. Let δ_{RD} be the causal effect of W_i when $T_t \in \{0, 2\}$. To identify the causal effect of P_i, we combine both the discontinuous variation at $C_i = 5,000$ and the time variation when T moves from 0 to 1 and then from 1 to 2. The target estimand of the difference-in-discontinuities estimator (hereafter, diff-in-disc) is $\tau_{DD(c)} = \alpha_{RD(c)} - \delta_{RD(c)}$ and it is narrower than the standard RD estimand, for it is conditional on the realization of the confounding treatment. $\hat{\tau_{DD(c)}}$ thus yields the local average treatment effect of a larger polity size for municipalities that also have better-paid politicians.

Two additional assumptions compared to the standard continuity of density and potential outcomes of RD designs must hold for the diff-in-disc estimator to be unbiased. The first one is that the effect of W_i and P_i is constant over time. In other words, it requires that units just above and below the threshold would have held a parallel trend had P_i not been introduced. The second assumption is that the effect of P_i does not depend on the confounding policy W_i. One way in which this assumption would be violated is if politicians above and below the cutoff, who are paid differently, reacted differently to a change in polity size. These assumptions are indirectly tested and discussed in Section K of the appendix, where we show that the effect of wage policies is stable over time and that another reform to the size of government bodies which equally affected all municipalities does not have differential effects for municipalities above and below the 5,000 cutoff. Similarly, in Section J we present validity tests in support of the continuity of density and potential outcomes assumptions, where we show no discontinuities for several pre-treatment covariates, no discontinuity in the density function of the running variable, and no significant effects at most placebo cutoffs.

We estimate $\tau_{DD(c)}$ with local polynomial methods, fitting linear WLS regressions separately on the observations above and below the cutoff and before, during, and after the reform (in $T = 0, 1, 2$ separately). Weights are determined by the triangular kernel function based on the ratio between the distance of unit i from the cutoff C_i and the

mean-squared-error minimizing bandwidth (Cattaneo et al., 2019). The RD estimates equal the difference in the intercepts at the cutoff in every time period. Subsequently, we estimate the difference in the point estimates across the three different periods.

Results

The results are displayed in Figure 2, which shows in the top panels the RD estimates for each period – namely municipalities that held elections before, during, and after the reform – and in the bottom panel the diff-in-disc estimates, estimated as the difference in the point estimates between the reform and pre-reform periods (black coefficients) and the reform and post-reform periods (blue coefficients).[4] Consistent with the general prediction that more politicians increase the size of the agenda, municipalities above the 5,000 cutoff have larger planned fiscal outcomes. Expenditures per capita are larger by 618 and 611 euros in the reform period compared to the pre- and post-reform periods, respectively, as large as 54% compared to the average in the data for municipalities in the 3-10,000 population band in the same period. Revenues increase too and to a very similar extent (i.e., 657 and 683 euros per capita compared to the pre and post-reform period). The increased spending coupled with an equal-size increase in revenues leaves planned deficit unchanged. Having approximately four politicians more does not lead to overspending, for larger planned expenditures do not exceed larger planned revenues.

However, municipal governments fail to realize the expanded planned budget passed at the beginning of the fiscal year. When looking at the right-hand side panels, municipalities above the cutoff that held elections in the reform period have very similar budgets to municipalities above the cutoff before which were not affected by the reform. The diff-in-disc estimates are much smaller compared to the planned budget and not distinguishable from 0 at standard confidence levels. More politicians lead to a congested budget, with larger spending and revenues that nonetheless do not materialize.[5]

[4]RD plots and full regression tables are reported in Section G and H in the appendix.

[5]In Section I in the appendix we show results are robust to alternative bandwidths, estimating a single equation, excluding covariates as well as adding additional covariates, using alternative outcome variables from the National Institute of Statistics, accounting for the government term business cycle, and limiting the analysis to municipalities which held elections before 2013, for gender quotas on candidate lists started to operate based on the same population threshold from December 2012 and might confound

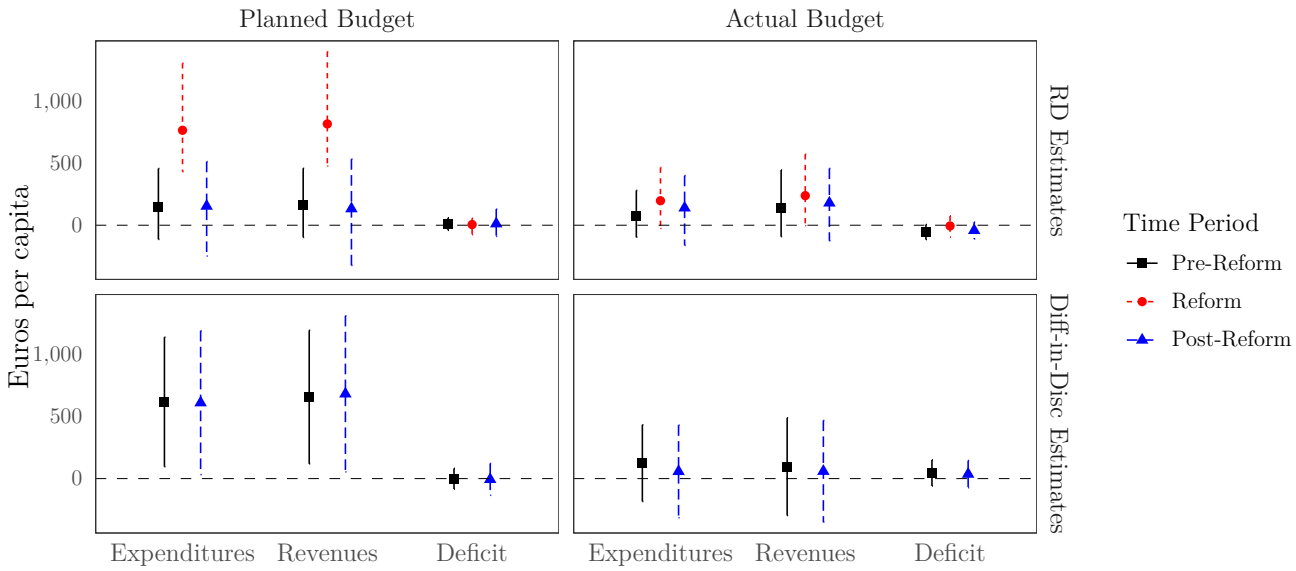

Figure 2: RD and Diff-in-Disc estimates with 95% robust confidence intervals. Covariates include: population density, surface (sq.km), surface at low, medium, and high hydro-geological risk (sq.km) – all log transformed –, gender, mayor with university degree (dummy), white-collar mayor (dummy), year and province dummies.

Mechanism

The congested budget is characterized by larger expenditures and revenues that however do not materialize in practice, reducing the spending and collection capacity of government. Collection and spending capacity deficit may be the product of capacity constraints, with bureaucracies unable to execute the larger planned budget.[6] To test whether there is empirical support for this mechanism, we replicate the main analysis on two separate samples of municipalities whose levels of bureaucratic capacity are above or below the median in the sample. We proxy bureaucratic capacity with the size of the bureaucracy (i.e., number of bureaucrats) as well as the percentage of bureaucrats with a university degree, capturing both quantitative and qualitative aspects of capacity. For these tests, the outcome variables are the ratio of actual and planned expenditures – i.e., spending capacity – and the ratio of actual and planned revenues – i.e., collection capacity. These measures are obtained directly from the National Institute of Statistics.

the effect of polity size. As for the results on revenues, we find that the results are not driven by any specific type of revenues.

[6]The congestion effect is not driven by the lower quality of politicians above the cutoff. In fact, we find that the reform had a positive effect on the selection of politicians, which we find to be more educated on average (see Table I.18 in the appendix).

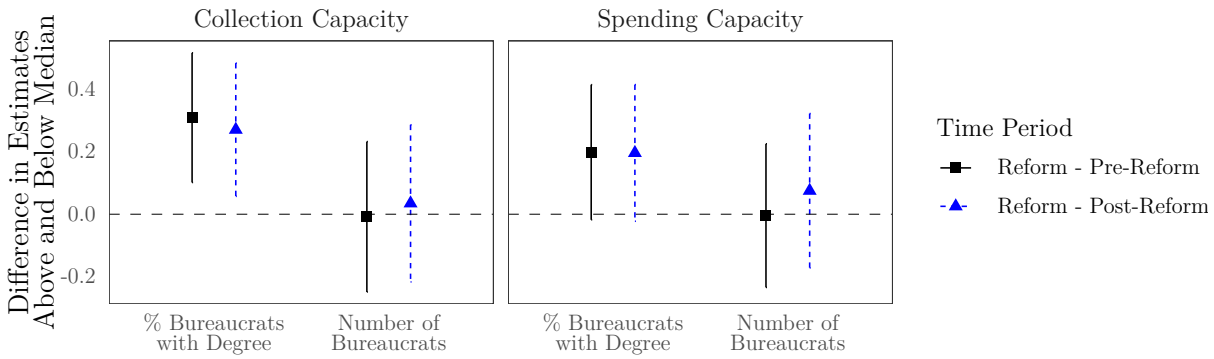

Figure 3: Difference in diff-in-disc estimates with 95% robust confidence intervals estimated across two samples (units above and below the median value of capacity indicators reported on the horizontal axis). Outcomes are collection capacity (ratio of actual/planned revenues) and spending capacity (ratio of actual/planned expenditures). Same covariates used in main analysis.

We produce difference-in-discontinuities estimates in the two samples and we compute the difference for municipalities above and below the median value, which we display in Figure 3 for both capacity variables. Spending and collection capacity increase respectively by 30 and 20 percentage points in municipalities with a larger share of bureaucrats with a university degree (although the estimates for spending capacity are less precise and statistically significant at 90%).[7] Conversely, the size of the bureaucracy does not alter spending and collection capacity, suggesting capacity deficits are more driven by the quality rather than the quantity of bureaucrats. These findings are consistent with our prediction that capacity constraints limit the ability of politicians to turn their expanded agenda into actual policy.

Discussion

In this article, we provide evidence of a clash between politicians' incentives and the existence of a binding bureaucratic capacity constraint. The larger spending proposals determined by a larger polity size are not implemented in practice, leading to spending and collection capacity reductions. These findings have important consequences for optimal institutional design. Political reforms aimed at improving policy-making by rationalizing

[7]The results are similar when building the outcomes from the budget data, although the coefficient for spending capacity is less precisely estimated (see Section H2).

the size of government bodies without considering the different levels of bureaucratic capacity might exacerbate the inequalities between local governments. Institutional reforms to the polity size of local governments should not consider population size alone, but also the quality of the bureaucracy. An increase in the number of politicians in municipalities with well-functioning bureaucracies will lead to an expanded budget, whereas it will increase the backlog of proposed policies in governments with low bureaucratic capacity, possibly worsening even those policies ultimately enacted. In fact, inefficiencies might arise if "bad" policies – which might be electorally more attractive to politicians – are more likely to make it to the final budget compared to "good" policies.

References

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (2019), *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Vol. I, Cambridge University Press.

Eggers, A. C., Freier, R., Grembi, V. and Nannicini, T. (2018), 'Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions', *American Journal of Political Science* **62**(1), 210–229.

Eslava, M. (2011), 'The political economy of fiscal deficits: A survey', *Journal of Economic Surveys* **25**(4), 645–673.

Gratton, G., Guiso, L., Michelacci, C. and Morelli, M. (2021), 'From Weber to Kafka: Political instability and the Overproduction of Laws', *American Economic Review* **111**(9), 2964–3003.

Grembi, V., Nannicini, T. and Troiano, U. (2016), 'Do fiscal rules matter?', *American Economic Journal: Applied Economics* **8**(3), 1–30.

Weingast, B. R., Shepsle, K. A. and Johnsen, C. (1981), 'The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics', *Journal of Political Economy* **89**(4), 642–664.

Online Appendix

A Simple Model for the General Predictions . . . 1

B Data Sources . . . 3

C Sample of Municipalities and Inflation Adjustment of Budget Data . . . 4

D Policies Based on Population Thresholds . . . 5

E Reforms to the Size of Municipal Government Bodies . . . 7

F Descriptive Statistics . . . 8

G RD Plots . . . 9

H Regression Tables . . . 11

- H1 Main Analysis . . . 11
- H2 Mechanism . . . 13

I Robustness Tests . . . 20

- I1 Removing Covariates . . . 20
- I2 Additional Covariates . . . 20
- I3 Alternative Bandwidths . . . 21
- I4 Additional Outcomes from National Institute of Statistics . . . 22
- I5 Government Term and Fiscal Cycles . . . 23
- I6 Gender Quotas . . . 25
- I7 Types of Revenues . . . 26
- I8 Selection Effect of the Reform . . . 27
- I9 Diff-in-Disc Estimated with Single Equation . . . 28

J Validity of RD Estimator . . . 29

- J1 Continuity of Density . . . 29
- J2 Continuity of Potential Outcomes . . . 30

K Validity of Diff-in-Disc Estimator . . . 32

A Simple Model for the General Predictions

In this section, we show formally why our prediction and results should apply to a general class of situations, with large external validity. To repeat first our starting point, any politician elected to any elected official role must have a minimum number of projects to propose and try to make happen. Indeed, typically an elected official is an "agent" subject to the scrutiny, interim or ex post, of potentially multiple principals, but always at least one type of principal. Sometimes the primary concern of an elected official is to look good to the voters who may have to reelect them, sometimes the reputation concerns are with the general public or audience they are visible to, sometimes they have to satisfy directly or indirectly the interest groups who either financed their election campaign or are actively lobbying them, sometimes the career concerns are within party, or within the electoral coalition, in order to advance their political career to some other political or public management job once the elected office is over. The fact that each elected official must have at least one of the long list of "bosses" looking at their performance should be considered obvious.

Let us now consider two municipalities. In Municipality 1, a polity of k politicians have to agree at time 0 about a set of policies to put on the planned budget, with related financial coverage. On the budget, the set of policies and their respective financial coverage take the form of expenditures and revenues. The set of potential policies is $X \equiv \{x_1, ..., x_n\}$, with $n > k$. For simplicity, let us assume that each policy $x_i, i = 1, n$ has the same tax revenue cost c, and that voters' income and willingness to pay (or government transfers to the municipality) allow a maximum of k projects to be financed, so that k is both the number of elected officials and the maximum number of projects that could be justified or covered by financial resources at the budget planning stage. Municipality 2 differs from Municipality 1 only for polity size, with $h > k$ members. Assume $k > \frac{h-1}{2}$, so that the difference is not too large, and assume that both k and h are odd numbers, in order to avoid tie breaking rules or integer problems. In both municipalities the decision is by majority rule, using demand bargaining as clarified below.

Let us assume that each politician has a benefit $B > 0$ in case their most important project gets completed or clearly on the way by the end of the time in office, leaving out the details of whether this benefit B comes from one of the principals mentioned above or from intrinsic motivations of the agent/politician. For simplicity, all other projects would yield zero utility for her. Formally, for each elected official j there exists $x \in X : \ u_j(x) = B$ and $u_j(y) = 0 \ \forall y \in X, y \neq x$.

In line with Morelli (1999), the demand bargaining procedure to reach an agreement on the planned budget works as follows: the members of the polity make a demand, sequentially, and each demand is simply a subset of X. A majority coalition and hence an agreement is formed as soon as a majority of the polity has made compatible demands, i.e., the sum of costs of the demands made by a majority coalition cannot exceed ck. Let us denote by $Y \subseteq X$ the set of policies placed on the planned budget by a committee in equilibrium.

Finally, assume that both municipalities have a bureaucratic capacity constraint and that it is the same among them. Such a bureaucratic capacity constraint can be modeled by assuming that the probability of implementation of any project x in the planned budget is equal to $P(|Y|)$,

where $|Y|$ is the number of elements of the set Y and $P(\cdot) \in (0,1)$ for every positive number of projects, decreasing in the number of projects on the budget. Intuitively, the probability that the bureaucracy manages to implement all the policies is a function of the number of policies on the planned budget (i.e., $|Y|$).

Proposition 1. *(I) If P' decreases not too sharply, the planned budget of Municipality 2 is always larger than the planned budget of Municipality 1.*

(II) Moreover, the ratio of actual over planned budget is lower (in expectation) in Municipality 2 than in Municipality 1, under the same conditions.

Proof. The unique equilibrium of the demand bargaining game in Municipality 1 displays $|Y_1| = \frac{k-1}{2}$, while in Municipality 2 it must be $|Y_2| = \frac{h-1}{2} > |Y_1|$. In both, a simple majority of members each obtains the planning of her own preferred project. The proof is trivial: whatever the order in which elected officials move in the demand game, where the order could come from any institutional or randomization protocol, the first $\frac{l-1}{2}$ members – $l = k, h$ – have each incentive to demand their own preferred project, since the probability of completion is positive and any subset of X is justifiable by a revenue coverage by assumption. Even though the addition of one demand reduces the probability of completion P of all projects marginally, the deviation to not demand such a project would yield zero. The unique equilibrium is robust to changes in the utility function, for example to assuming that each member also would derive positive utility from the projects of other members in the majority coalition, as long as at least for some members such a utility from projects of others is not too large.

Part (II) follows mechanically.

QED.

B Data Sources

To assemble the dataset, we rely on four sources of data.

1. We obtained full lists of municipalities and unique budget identifiers directly from the Local Public Finance Directorate of the Ministry of Interior. We then used the unique IDs to build URLs and scrape budget data from the on-line ministerial repository. This repository, available at `https://finanzalocale.interno.gov.it/apps/floc.php/in/cod/4` reports the data contained in the certificates of the balance sheets that the municipalities, provinces, and metropolitan cities *must* transmit to the Ministry of Economy and Finance, pursuant to Legislative Decree 118/2011. Every municipality has to communicate this data and hence non-compliance and missing data are not a concern for the definition of the sample of municipalities.

2. We merged budget data with data on local government composition (composition and biographical information on mayors and municipal politicians) obtained from the Database on Local and Regional Administrators, curated by the Ministry of the Interior. The data can be accessed at `https://dait.interno.gov.it/elezioni/open-data`. We exact matched the two datasets based on the name of the municipalities. We manually checked those unmatched municipalities and we resolved conflicts on a case-by-case basis. The Database on Local and Regional Administrators contains another unique ID for municipalities which is the one produced by the National Institute of Statistics, hence we merge all other data sources by this ID.

3. We obtained data on socio-demographic and territorial characteristics of municipalities as well as the indicators on spending and collection capacity, and the share of bureaucrats with a university degree from the National Institute of Statistics. Data available at `http://dati.statistiche-pa.it/`.

4. We accessed data on the number of municipal employees from the national account of the Italian General Accounting Office. Data available at `contoannuale.mef.gov.it`. This data does not report the unique IDs of the National Institute of Statistics, hence the dataset has been exact matched on the name of the municipalities.

5. We obtained data on the personal income certificates of municipal residents from the Ministry of the Economy and Finance (the Italian acronym of this certificate is IRPEF). Data available at `https://www.finanze.gov.it/it/statistiche-fiscali/`. We merged this data by the unique ID of the National Institute of Statistics.

C Sample of Municipalities and Inflation Adjustment of Budget Data

From the total sample of municipalities, we removed the municipalities located in the five special statute regions (Sicilia, Sardegna, Valle d'Aosta, Friuli-Venezia Giulia, and Trentino-Alto Adige/Südtirol) for the threshold mechanisms apply to those regions only insofar as they are compatible with their own special statutes, and because these municipalities are subject to financial constraints and rules that differ from those in force for the remaining 15 ordinary-statute regions.

All the outcome variables are per capita and adjusted to the 2018 inflation level. Figures are therefore expressed in real terms and are comparable over time. We applied the consumer price index for currency evaluations as of January of every year (Jan. 2018 index = 100). Time-series downloaded from the archive of the National Institute of Statistics, `www.istat.it/it/archivio/30440`.

D Policies Based on Population Thresholds

The tables below report rules which change discontinuously at population thresholds. We also report the legal sources, the time frame when the rules remained in force, and the various cutoffs at which they operated. In particular, Table D.2 shows the wage policies changing at the 5,000 cutoff.

Rule/Policy	**Time-Frame**	**Legal Source**	**Cutoffs [1,000]**
Wage Mayor	1998-2000 2001-Present	Art. 3 and Table A, Law 861/1985 Art. 1 and Table A, Min. Dec. 119/2000	3, 5, 10, 30, 50 1, 3, 5, 10, 30, 50
Wage Vice Mayor	2001-Present	Art. 3 and Table A, Min. Dec. 119/2000	1, 3, 5, 10, 30, 50
Wage Members Executive Committee	2001-Present	Art. 4, Min. Dec. 119/2000	1, 5, 50
Reimbursement to Local Councillors	1998-2000 2001-Present	Art. 10, Law 861/1985 Art. 1 and Table A, Min. Dec. 119/2000	30 1, 10, 30
Wage President of Council	2001-Present	Art. 5, Min. Dec. 119/2000	1, 15
Presence of Neighborhood Councils	1998-2000 2001-Present	Art. 13, Law 142/90 Art. 17, Leg. Dec. 267/2000	30 (optional)
Number of Financial Auditors	1998-2000 2001-2006 2007-Present	Art. 57, Law 142/90 Art. 234, Leg. Dec. 267/2000 Art. 1(732) Law 296/2006	5 5 15
Presence of Director General	1998-2000 2001-2003	Art. 51-bis, Law 142/90 Art. 108, Leg. Dec. 267/2000	15
Fiscal Rule	2001-2012	Art. 53, Law 388/2000 and other subsequent laws	5
Gender Quotas on Party Lists	2013-Present	Art. 2, Law 215/2012	5, 15
Cut to Size of Government Bodies	2011-2014	Art. 2(184) Law 191/2009, Art. 1(2) Law 42/2010	1, 3, 5, 10, 30, 50
Introduction of 5,000 Threshold	2011-2014	Art. 16(17) Decree-Law 138/2011	1, 3, 5, 10

Table D.1: Legal sources of all rules/policies based on population thresholds with reported cutoffs up to 50,000 inhabitants.

	Population Thresholds						
	1,000	3,000	5,000	10,000	15,000	20,000	30,000
Wage Policies							
Mayor	2001-Present	**X**	**X**	**X**			**X**
Vice Mayor	2001-Present		**X**	**X**			
Reimb. Councillor	2001-Present			**X**			
Exec. Comm.	2001-Present		**X**				
President of L Counc.	2001-Present				**X**		
Other Policies							
Neighbour. Councils							?
Health Center						?	
Fiscal Rules			2001-12				
Gender Quotas			2013-Present		2013-Present		
Financial Auditors			1998-06		**X**		
Electoral System					**X**		
Director General					?		
Treatment							
Local Council Size	2011-13	**X**	2011-13	**X**			**X**
Executive Comm.	2011-13	**X**	2011-13	**X**			**X**

Table D.2: Rule/Policies that change with population thresholds with respective time-frame (entire time-frame when omitted) Thresholds larger than 30,000 omitted. **X** signifies deterministic change, ? signifies possible change, namely cases in which municipalities *may* adopt the rule or implement the policy if they want to. Hospital means whether the municipality can have a hospital or health center and neighborhood councils are local councils that can be established in any neighborhood. Two important exceptions: wage/reimbursement thresholds at 1,000 were introduced only in 2001.

E Reforms to the Size of Municipal Government Bodies

In 2011, two reforms were passed aimed at reducing the number of politicians in municipal governments, with the goal of controlling public expenditures. The first reform (Reform 1) affected all municipalities, without affecting the population-threshold mechanism. A second reform (Reform 2), the one studied in this paper, introduced a new threshold of 5,000 inhabitants, which was then repealed in 2014. Table E.3 shows how the size of government bodies changed under these two different reforms.

			Election Years 2011-2013					
Threshold	**Pre-Reform**		**Reform 1**		**Reform 2**		**Post-Reform**	
[1,000]	Council	Ex. C.	Council	Ex. C.	Council	Ex. C.	Council	Ex. C.
0 – 3	12	4	10	3	6	2	10	2
3 – 5	16	6	13	4	7	3	12	4
5 – 10					10	4		
10 – 30	20	7	16	5	16	5	16	5
> 30	30	10	24	7	24	7	24	7

Table E.3: Number of local councilors and cap to size of executive committees before and after two reforms which were passed in 2011.

F Descriptive Statistics

Table F.4 reports descriptive statistics for the main variables for the total sample and the sample of municipalities in the 3-10,000 inhabitants population band.

	Full Dataset			**3-10,000 Population Band**		
N. Municipalities		8,451			2,083	
N. Observations		143,406			5,832	
Budget Item	**Mean**	**Median**	**SD**	**Mean**	**Median**	**SD**
Local Councillors	14.2	12.0	4.8	15.2	16.0	1.9
Members of Exec. Comm.	3.2	3.0	1.7	3.9	4.0	1.1
Planned Budget						
Expenditures pc	1,769.0	1,320.9	2,182.7	1,309.3	1,118.3	831.5
Revenues pc	1,754.1	1,306.4	1,974.6	1,303.6	1,108.4	849.9
Deficit pc	12.3	5.7	251.3	5.7	5.3	200.5
Actual Budget						
Expenditures pc	999.8	809.3	977.9	804.3	704.2	478.3
Revenues pc	1,035.8	825.1	1,104.6	826.4	723.2	485.7
Deficit pc	-36.1	-13.3	454.6	-22.1	-16.4	190.0

Table F.4: Descriptive statistics of main variables in the entire dataset and for the sample of units with census population between 3,001 and 10,000 inhabitants. Descriptive statistics of main variables. Deficit per capita measures are equal to the difference between total expenditures and total revenues divided by the resident population.

G RD Plots

In the figures below we report RD plots with WLS fitted lines estimated separately within the MSE-optimal bandwidth (vertical dotted line) above and below the 5,000 cutoff for planned (Figure G.1) and actual budget figures (Figure G.2).

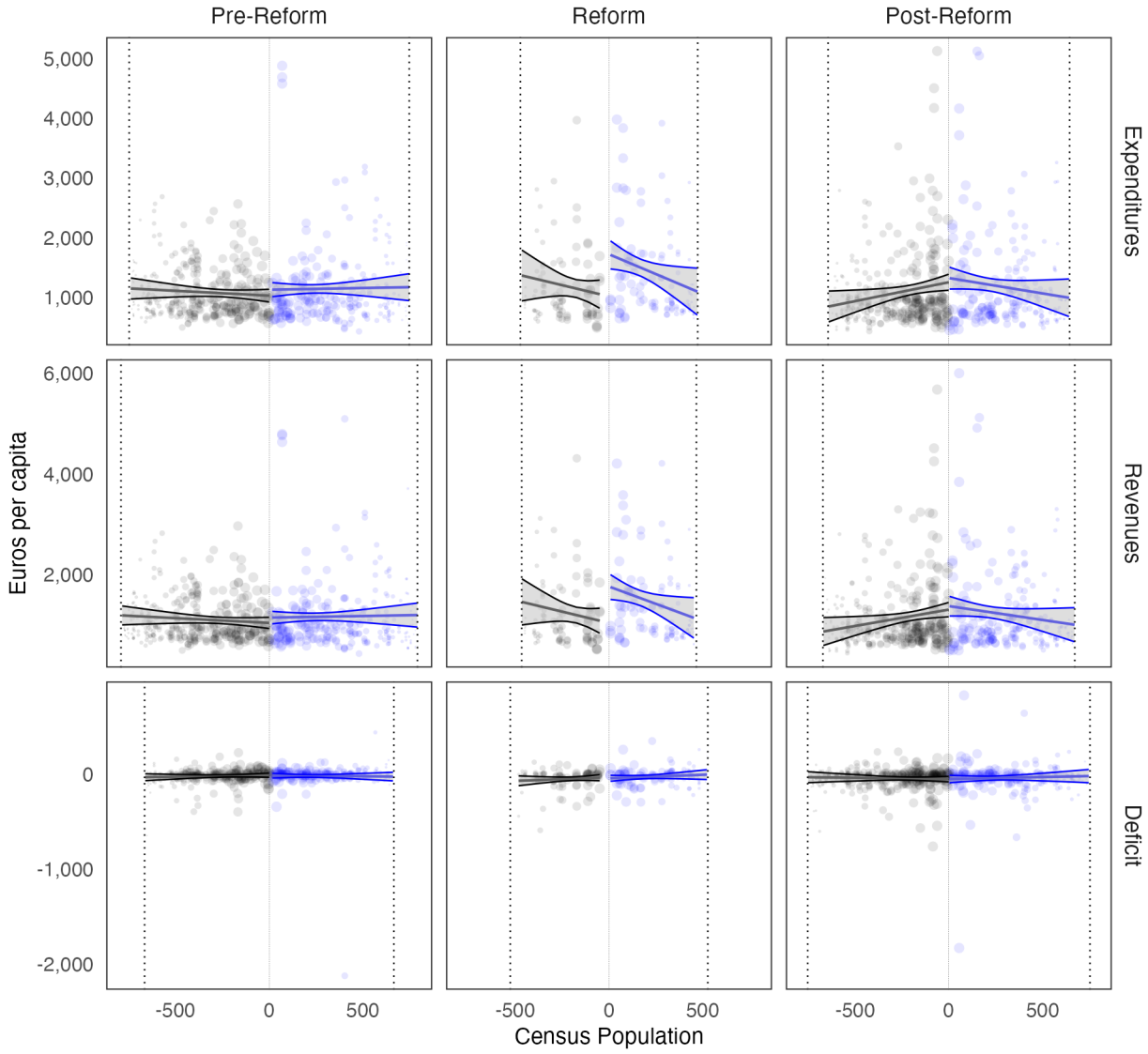

Figure G.1: RD plot with fitted WLS line estimated separately above and below the cutoff for planned budget. Each dot is a municipality-year observation, and the size of the dot is a function of the weight determined by the triangular kernel function based on the ratio of the distance of each observation from the cutoff and the MSE-minimizing bandwidth (vertical dotted line).

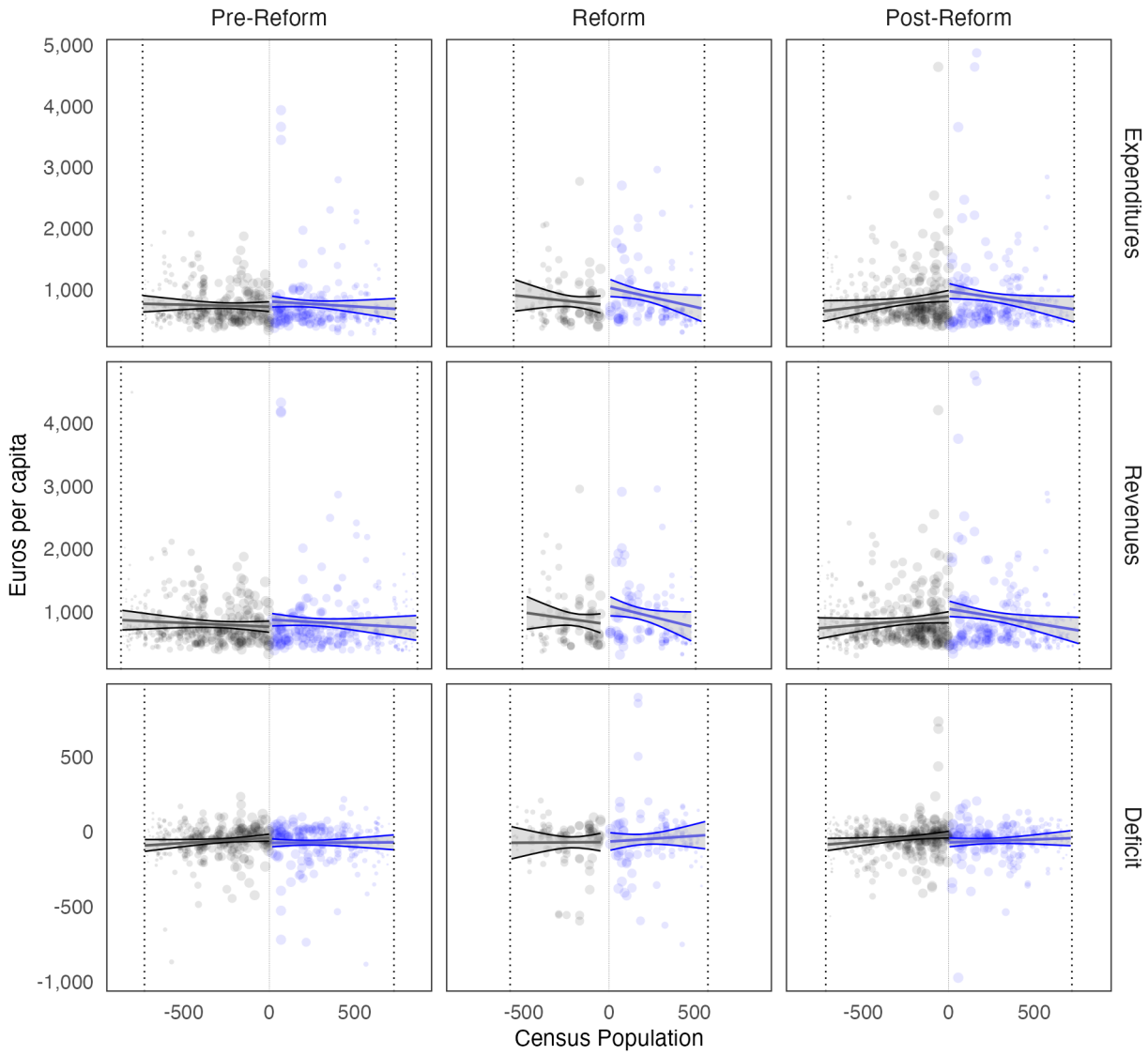

Figure G.2: RD plot with fitted WLS line estimated separately above and below the cutoff for actual budget. Each dot is a municipality-year observation, and the size of the dot is a function of the weight determined by the triangular kernel function based on the ratio of the distance of each observation from the cutoff and the MSE-minimizing bandwidth (vertical dotted line).

H Regression Tables

H1 Main Analysis

In the table below we report the RD results for the three time periods separately. These are the estimates reported in the top panels of Figure 2 in the main text. The analysis was performed with the *rdrobust* package in R (Calonico et al., 2015).

	Planned Budget			**Actual Budget**		
Outcome	Expenditures	Revenues	Deficit	Expenditures	Revenues	Deficit
Pre-Reform						
Estimate	72.8	142.2	-51.2	147.3	159.1	6.1
SE	(96.1)	(136.9)	(30.8)	(145.6)	(141.9)	(25.7)
p.value	0.339	0.197	0.075	0.239	0.204	0.710
h	1043.8	708.1	627.3	780.6	817.4	500.3
Obs. Used	751	456	400	515	549	332
Reform						
Estimate	197.4	238.5	-5.7	764.7	815.9	5.3
SE	(125.1)	(146.8)	(43.7)	(222.9)	(235)	(33.6)
p.value	0.080	0.053	0.811	0.000	0.000	0.763
h	845.7	660.7	598.2	556.4	492.6	453.1
Obs. Used	339	255	222	201	181	167
Post-Reform						
Estimate	139.7	179.6	-41.7	153.7	133.3	11.8
SE	(143.3)	(148.8)	(34.6)	(194)	(218.4)	(56)
p.value	0.408	0.265	0.219	0.497	0.632	0.721
h	631.7	656.2	766.5	573.0	591.5	734.4
Obs. Used	456	483	576	425	435	551

Table H.5: RD estimates as displayed in Figure 2 for each time period and each outcome separately. Estimates constructed using local polynomial estimators with triangular kernel and MSE-optimal bandwidth (h). Robust p.values computed using bias-correction with robust standard errors. Covariates include: population density, surface (sq.km), surface at low, medium, and high hydro-geological risk (sq.km) – all log transformed –, gender, mayor with university degree (dummy), white-collar mayor (dummy), year and province dummies.

In the table below we report the diff-in-disc estimates comparing the RD estimates in $T = 1$ (under the reform) with the RD estimates in $T \in \{0, 2\}$, pre- and post-reform, respectively. These are the estimates reported in the bottom panels of Figure 2 in the main text. The point estimate is the difference between the RD point estimates in the two periods, and the standard error of the difference has been computed with the following formula, as in Klašnja and Titiunik (2017): $SE_{DD} = \sqrt{SE^2_{RD|T=1} + SE^2_{RD|T\in\{0,2\}}}$, where $T = 1$ refers to the SE of the RD point estimate for reform period, and $T \in \{0, 2\}$ for pre- and post-reform periods, respectively.

	Reform - Pre-Rerom			**Reform - Post-Reform**		
Outcome	Expenditures	Revenues	Deficit	Expenditures	Revenues	Deficit
Planned Budget						
Diff.	617.5	656.8	-0.8	611.0	682.6	-6.5
SE	(266.2)	(274.5)	(42.3)	(295.5)	(320.8)	(65.3)
p.value	0.020	0.017	0.986	0.039	0.033	0.921
Actual Budget						
Diff.	57.6	58.9	36.0	57.6	58.9	36.0
SE	(190.2)	(209.0)	(55.8)	(190.2)	(209.0)	(55.8)
p.value	0.762	0.778	0.519	0.762	0.778	0.519

Table H.6: Diff-in-Disc estimates (i.e., difference in RD point estimates reported in Table A.5) as displayed in Figure 2 for every outcome and planned and actual budgets.

H2 Mechanism

In the tables below we report the RD estimates in the two samples of municipalities whose capacity indicator is below and above the median value. We use two outcome variables produced by the National Institute of Statistics.

- Collection capacity is an indicator computed as the ratio of actual over planned revenues.
- Spending capacity is an indicator computed as the ratio of actual over planned expenditures.

For each outcome, we first report RD estimates in the two samples of municipalities above and below the median capacity value (Table H.7 and Table H.9) and then the diff-in-disc estimates in the two samples as well as the difference in diff-in-disc estimates between the samples (Table H.8 and Table H.10).

Moreover, we replicate this analysis using collection and spending capacity indicators built from the budget data we scraped ourselves and show the results are similar.

Capacity Indicators Produced by the National Institute of Statistics

	DV: Collection Capacity (Actual/Planned Revenues)					
	Below Median			Above Median		
Time Period	Pre-Reform	Reform	Post-Reform	Pre-Reform	Reform	Post-Reform
Capacity Indicator: % Bureaucrats with Degree						
Estimate	0.00	-0.29	0.04	-0.06	-0.04	0.02
SE	(0.05)	(0.06)	(0.05)	(0.03)	(0.06)	(0.03)
p.value	0.98	0.00	0.30	0.03	0.46	0.78
h	525.13	434.07	561.69	452.23	744.38	576.80
Obs. Used	167	84	199	158	128	210
Capacity Indicator: N. Bureaucrats						
Estimate	-0.06	-0.11	0.04	-0.02	-0.09	0.03
SE	(0.03)	(0.07)	(0.03)	(0.04)	(0.09)	(0.05)
p.value	0.01	0.04	0.17	0.78	0.33	0.49
h	453.71	531.25	460.28	592.63	509.78	705.23
Obs. Used	175	115	247	158	66	139

Table H.7: RD estimates for each time period and each outcome in two samples consisting of municipalities whose capacity indicator is above and below the median. Outcome is collection capacity indicator produced by National Institute of Statistics, i.e., the ratio of actual over planned revenues. Variable proxying capacity indicator has been reported in each panels. Estimates constructed using local polynomial estimators with triangular kernel and MSE-optimal bandwidth (h). Robust p.values computed using bias-correction with robust standard errors. Same covariates used in main analysis.

	DV: Collection Capacity (Actual/Planned Revenues)					
	Below Median		Above Median		Above - Below	
Time Period	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform
Capacity Indicator: % Bureaucrats with Degree						
Difference	-0.290	-0.330	0.020	-0.060	0.310	0.270
SE	(0.08)	(0.08)	(0.07)	(0.07)	(0.11)	(0.11)
p.value	0.00	0.00	0.82	0.44	0.00	0.01
Capacity Indicator: N. Bureaucrats						
Difference	-0.060	-0.150	-0.060	-0.110	-0.010	0.030
SE	(0.08)	(0.08)	(0.10)	(0.10)	(0.12)	(0.13)
p.value	0.46	0.05	0.51	0.27	0.95	0.79

Table H.8: Diff-in-Disc estimates computed separately for below- and above-median samples. Outcome variable is collection capacity indicator produced by National Institute of Statistics, i.e., the ratio of actual and planned revenues. Same covariates used in main analysis. 'Above - Below' columns report the difference in the diff-in-disc estimates, with the standard error calculated with the following formula: $SE = \sqrt{SE^2_{Above} + SE^2_{Below}}$, where *Above* refers to the SE of the diff-in-disc point estimate for the above-median sample, and *Below* refers to the SE of the diff-in-disc point estimate for the below-median sample.

	DV: Spending Capacity (Actual/Planned Expenditures)					
	Below Median			Above Median		
Time Period	Pre-Reform	Reform	Post-Reform	Pre-Reform	Reform	Post-Reform
Capacity Indicator: % Bureaucrats with Degree						
Estimate	-0.03	-0.26	-0.01	-0.03	-0.06	-0.02
SE	(0.04)	(0.06)	(0.04)	(0.03)	(0.08)	(0.03)
p.value	0.58	0.00	0.82	0.13	0.56	0.61
h	606.43	426.70	829.33	561.52	643.10	600.65
Obs. Used	179	83	301	179	109	215
Capacity Indicator: N. Bureaucrats						
Estimate	-0.04	-0.10	0.02	-0.06	-0.12	-0.07
SE	(0.03)	(0.08)	(0.03)	(0.04)	(0.08)	(0.06)
p.value	0.11	0.09	0.40	0.18	0.12	0.15
h	824.67	489.59	520.61	597.04	436.88	663.41
Obs. Used	323	112	277	158	59	126

Table H.9: RD estimates for each time period and each outcome in two samples consisting of municipalities whose capacity indicator is above and below the median. Outcome is spending capacity indicator produced by National Institute of Statistics, i.e., the ratio of actual and planned expenditures per capita. Variable proxying capacity indicator has been reported in each panels. Variable proxying capacity indicator has been reported in each panels. Estimates constructed using local polynomial estimators with triangular kernel and MSE-optimal bandwidth (h). Robust p.values computed using bias-correction with robust standard errors. Same covariates used in main analysis.

	DV: Spending Capacity (Actual/Planned Expenditures)					
	Below Median		Above Median		Above - Below	
Time Period	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform
Capacity Indicator: % Bureaucrats with Degree						
Difference	-0.230	-0.240	-0.030	-0.050	0.200	0.200
SE	(0.07)	(0.07)	(0.08)	(0.08)	(0.11)	(0.11)
p.value	0.00	0.00	0.73	0.58	0.07	0.08
Capacity Indicator: N. Bureaucrats						
Difference	-0.050	-0.120	-0.060	-0.040	0.000	0.080
SE	(0.08)	(0.08)	(0.08)	(0.09)	(0.12)	(0.13)
p.value	0.51	0.16	0.48	0.65	0.97	0.55

Table H.10: Diff-in-Disc estimates computed separately for below- and above-median samples. Outcome variable is spending capacity indicator produced by National Institute of Statistics, i.e., the ratio of actual and planned expenditures per capita. Same covariates used in main analysis. 'Above - Below' columns report the difference in the diff-in-disc estimates, with the standard error calculated with the following formula: $SE = \sqrt{SE^2_{Above} + SE^2_{Below}}$, where *Above* refers to the SE of the diff-in-disc point estimate for the above-median sample, and *Below* refers to the SE of the diff-in-disc point estimate for the below-median sample.

Capacity Indicators Produced from Scraped Budget Data

In Tables H.11 and H.12 we replicate the same analysis this time building the measures of collection and spending capacity from the scraped budget data. Consistent with the measures produced by the National Institute of Statistics, we compute the two indicators as the ratio of actual expenditures (revenues) per capita over planned expenditures (revenues) per capita. These indicators are identical to those produced by the National Institute of Statistics, except for some missingness in both data sources. For instance, for some municipality-year pairs there is no available budget data but the National Institute of Statistics was still able to produce spending and collection capacity indicators, and vice versa.

The results are similar except for the confidence intervals of the effect of spending capacity, which are larger (p.values = .24 for the reform – pre-reform period and .28 for the reform–post-reform period).

	DV: Collection Capacity (Actual/Planned Revenues)					
	Below Median		**Above Median**		**Above - Below**	
Time Period	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform
Capacity Indicator: % Bureaucrats with Degree						
Difference	-0.27	-0.28	0.02	-0.05	0.29	0.23
SE	(0.08)	(0.08)	(0.07)	(0.07)	(0.11)	(0.11)
p.value	0.00	0.00	0.78	0.45	0.01	0.03
Capacity Indicator: N. Bureaucrats						
Difference	-0.05	-0.14	-0.06	-0.11	-0.01	0.03
SE	(0.08)	(0.08)	(0.11)	(0.12)	(0.14)	(0.14)
p.value	0.48	0.06	0.59	0.34	0.96	0.84

Table H.11: Diff-in-Disc estimates computed separately for below- and above-median samples. Outcome variable is collection capacity computed from scraped budget data as the ratio of actual and planned revenues per capita. Same covariates used in main analysis. 'Above - Below' columns report the difference in the diff-in-disc estimates, with the standard error calculated with the following formula: $SE = \sqrt{SE^2_{Above} + SE^2_{Below}}$, where *Above* refers to the SE of the diff-in-disc point estimate for the above-median sample, and *Below* refers to the SE of the diff-in-disc point estimate for the below-median sample.

	DV: Spending Capacity (Actual/Planned Expenditures)					
	Below Median		**Above Median**		**Above - Below**	
Time Period	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform	Reform - Pre-Reform	Reform - Post-Reform
Capacity Indicator: % Bureaucrats with Degree						
Difference	-0.16	-0.17	-0.02	-0.05	0.14	0.13
SE	(0.08)	(0.08)	(0.08)	(0.09)	(0.12)	(0.12)
p.value	0.04	0.02	0.78	0.59	0.24	0.28
Capacity Indicator: N. Bureaucrats						
Difference	-0.05	-0.11	-0.03	-0.01	0.03	0.11
SE	(0.08)	(0.08)	(0.09)	(0.10)	(0.12)	(0.13)
p.value	0.52	0.17	0.76	0.92	0.83	0.42

Table H.12: Diff-in-Disc estimates computed separately for below- and above-median samples. Outcome variable is spending capacity computed from scraped budget data as the ratio of actual and planned expenditures per capita. Same covariates used in main analysis. 'Above - Below' columns report the difference in the diff-in-disc estimates, with the standard error calculated with the following formula: $SE = \sqrt{SE^2_{Above} + SE^2_{Below}}$, where *Above* refers to the SE of the diff-in-disc point estimate for the above-median sample, and *Below* refers to the SE of the diff-in-disc point estimate for the below-median sample.

In the figure below we report the RD and Diff-in-Disc results reported in Figure 2 on different sub-samples of municipalities whose measures of bureaucratic capacity are above and below the median. It is clear from the top-left panel that the reform had an effect on actual budget for municipalities with a share of bureaucrats with university degree above the median value.

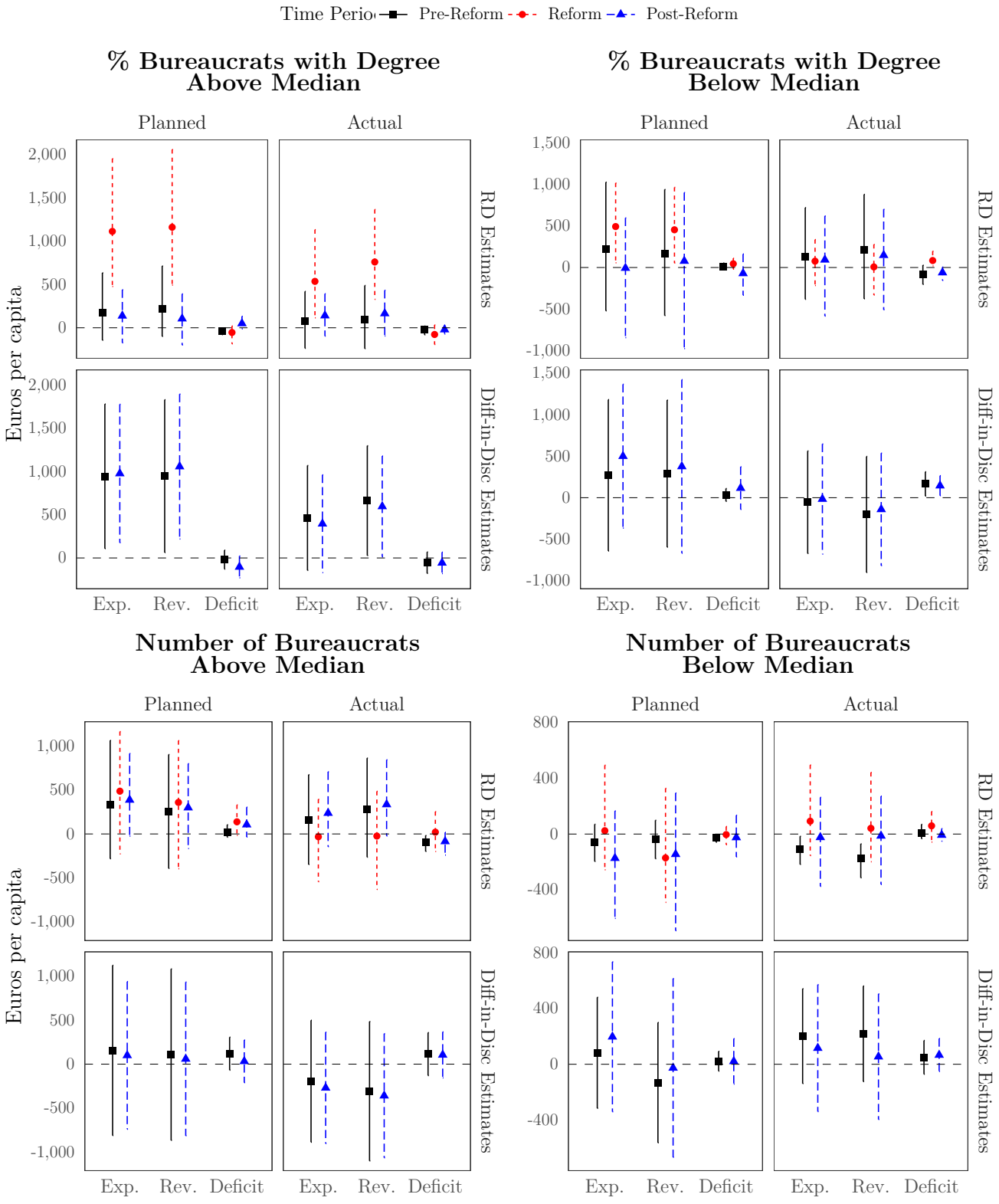

Figure H.3: RD and Diff-in-Disc estimates with 95% robust confidence intervals estimated on different samples of municipalities above and below the two measures of bureaucratic capacity. Same covariates included in the main analysis.

I Robustness Tests

I1 Removing Covariates

In the tables below we show the results are robust to omitting the covariates from the estimation (Table I.13).

	Reform - Pre-Rerom			**Reform - Post-Reform**		
	Expenditures	Revenues	Deficit	Expenditures	Revenues	Deficit
Planned Budget						
Difference	606.7	618.9	-7.2	635.3	651.3	-13.2
SE	(305.0)	(317.6)	(38.7)	(327.7)	(356.1)	(64.6)
p.value	0.047	0.051	0.853	0.053	0.067	0.838
Actual Budget						
Difference	209.2	160.0	55.2	209.2	160.0	55.2
SE	(209.7)	(216.3)	(59.7)	(209.7)	(216.3)	(59.7)
p.value	0.318	0.459	0.355	0.318	0.459	0.355

Table I.13: Diff-in-Disc estimates for every outcome and planned and actual budgets without including covariates.

I2 Additional Covariates

In the tables below we show the results are robust to including a larger set of covariates (Table I.14).

	Reform - Pre-Rerom			**Reform - Post-Reform**		
	Expenditures	Revenues	Deficit	Expenditures	Revenues	Deficit
Planned Budget						
Difference	496.6	509.0	2.1	519.4	539.4	-4.3
SE	(267.5)	(268.4)	(41.6)	(284.1)	(306.6)	(63.8)
p.value	0.063	0.058	0.959	0.068	0.079	0.947
Actual Budget						
Difference	64.8	40.4	50.9	64.8	40.4	50.9
SE	(194.0)	(205.0)	(55.0)	(194.0)	(205.0)	(55.0)
p.value	0.739	0.844	0.355	0.739	0.844	0.355

Table I.14: Diff-in-Disc estimates for every outcome and planned and actual budgets estimated including a larger set of covariates: log population density, log surface (sq.km), log surface at low, medium, and high hydro-geological risk (sq.km), gender and degree of mayor (dummy), white-collar mayor (dummy), right-wing mayor (dummy), left-wing mayor (dummy), average personal income declared by municipal residents, province and year fixed effects.

I3 Alternative Bandwidths

Figure I.4 below shows the diff-in-disc estimates are robust to using using alternative bandwidths.

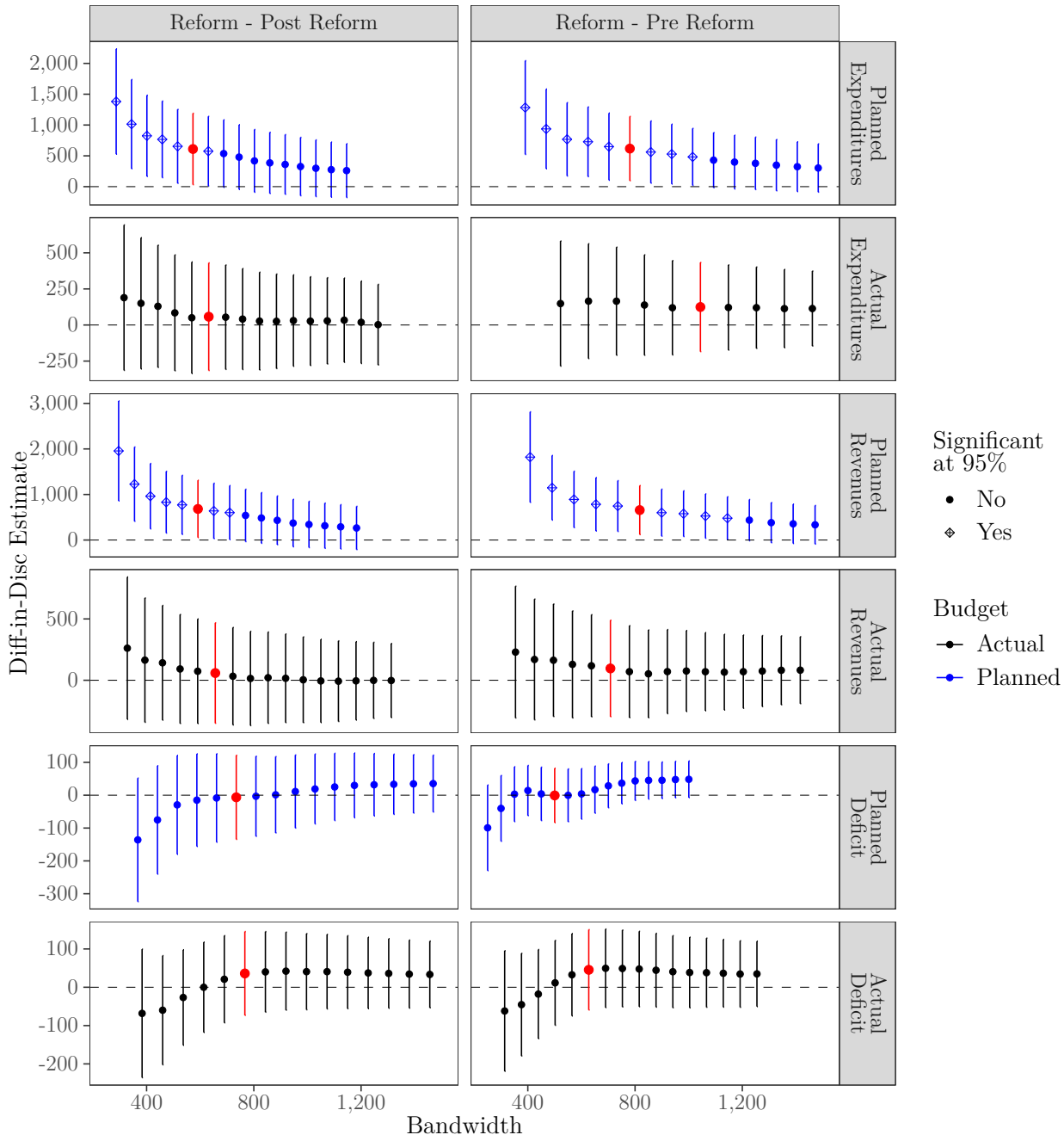

Figure I.4: Diff-in-Disc estimates with 95% robust confidence intervals estimated with alternative bandwidths. Red coefficients estimated with MSE-optimal bandwidth. Underlying RD estimates constructed using local polynomial estimators with triangular kernel and MSE-optimal bandwidth. Robust confidence interval constructed using bias correction with robust standard errors. Same covariates used in main analysis.

I4 Additional Outcomes from National Institute of Statistics

To show the results are robust to the selection of outcomes, in the table below we replicate the main analysis using two indicators built by the National Institute of Statistics capturing collection and spending capacity and remainder of administration for the same sample of municipalities. The decreased spending and collection capacity and the imprecise estimate for deficit strengthen the main results for which treated municipalities during the reform increase planned spending and revenues (the gap between planned and actual widens, as the effects for spending and collection capacity indicate) but these are not implemented in practice (no statistically significant effect on deficit).

	Reform - Pre-Rerom			**Reform - Post-Reform**		
	Spending Capacity	Collection Capacity	Deficit	Spending Capacity	Collection Capacity	Deficit
Difference	-0.16	-0.15	0.03	-0.15	-0.18	-0.29
SE	(0.06)	(0.06)	(0.14)	(0.07)	(0.07)	(0.30)
p.value	0.015	0.016	0.832	0.022	0.007	0.327

Table I.15: Diff-in-Disc estimates using alternative outcomes built by the National Institute of Statistics. Spending capacity is the ratio between actual and planned expenditures; collection capacity is ratio between actual and planned revenues; deficit is administration remainder divided by planned revenues. No covariates included.

I5 Government Term and Fiscal Cycles

Because the analysis is performed on a sample covering three calendar years, pre- and post-reform differences might be confounded by the year-of-term effect, with municipalities more ahead in the government term (and closer to new elections) more likely to spend more. We address this in two ways:

- We show the results are robust to including year-of-term dummies as a covariate (see Table I.16).
- We show that the estimates after removing the effect of the government term cycle on the outcomes are even larger and more precisely estimated. To partial out the effect of the government term on budget data we de-trend the outcomes by taking the residuals of a regression of each outcome on government year-of-term dummies and use these as outcomes in the analysis (see Table I.17). We report the outcome variables expressed in nominal terms and after being detrended in Figure I.5 below.

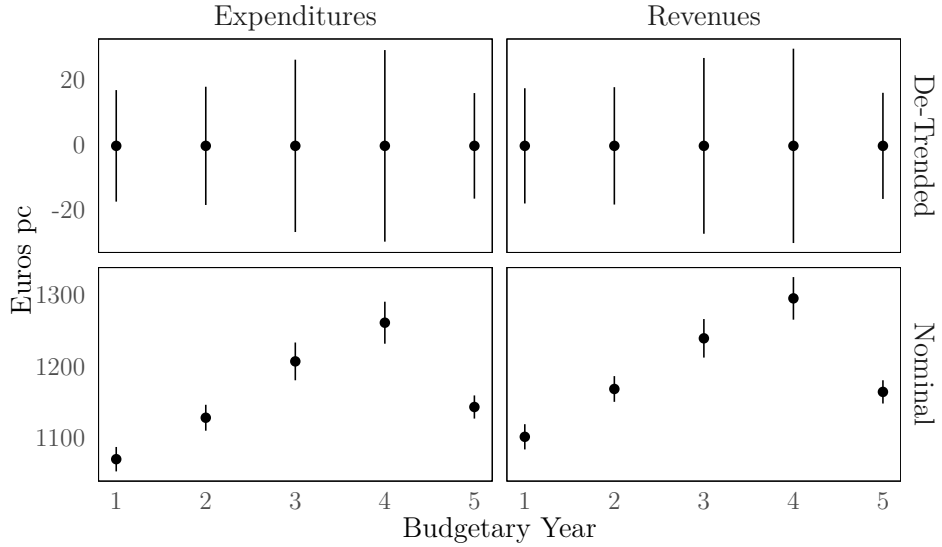

Figure I.5: Average expenditures and revenues per capita over the government term. Top panel reports de-trended outcomes, whereas bottom panel reports nominal outcomes. De-trended outcomes are the residual of a linear regression regressing the nominal outcome on year-of-term dummies (to partial out the effect of business cycles non-parametrically). Panels on the bottom show nominal averages.

	Reform - Pre-Rerom			Reform - Post-Reform		
Outcome	Expenditures	Revenues	Deficit	Expenditures	Revenues	Deficit
Planned Budget						
Diff.	649.0	711.5	-0.7	663.5	755.3	-6.5
SE	(268.9)	(276.7)	(42.3)	(299.2)	(324.3)	(65.4)
p.value	0.016	0.010	0.987	0.027	0.020	0.921
Actual Budget						
Diff.	92.0	82.9	28.2	92.0	82.9	28.2
SE	(194.1)	(209.1)	(54.6)	(194.1)	(209.1)	(54.6)
p.value	0.636	0.692	0.606	0.636	0.692	0.606

Table I.16: Diff-in-Disc estimates for each outcome and planned and actual budgets. Same covariates as in main analysis with the addition of year-of-term dummies.

	Reform - Pre-Rerom			Reform - Post-Reform		
	Expenditures	Revenues	Deficit	Expenditures	Revenues	Deficit
Planned Budget						
Difference	655.5	697.6	-2.2	649.5	730.6	-6.5
SE	(273.4)	(275.8)	(42.6)	(301.0)	(324.3)	(65.6)
p.value	0.016	0.011	0.959	0.031	0.024	0.922
Actual Budget						
Difference	76.6	63.7	34.6	76.6	63.7	34.6
SE	(194.0)	(209.0)	(55.0)	(194.0)	(209.0)	(55.0)
p.value	0.693	0.761	0.529	0.693	0.761	0.529

Table I.17: Diff-in-Disc estimates. Outcomes from RD estimates are the residuals of linear regressions of each outcome on the year-of-term variable. Same covariates included in main analysis.

I6 Gender Quotas

Because from 2013 gender quotas on candidate lists started to operate based on the same population threshold, the difference between the reform and pre-reform periods might be confounded by the presence of more female councilors in the government. This has no support in the data, as evidenced by the similar effect size of the diff-in-disc estimates in the pre- and post-reform period, which suggests that the gender composition of the local council has a negligible effect on local public finance. However, in the pre-gender quotas period (earlier than 26 December 2012), many municipalities already renewed their government bodies under the reform and became treated before gender quotas entered into force. Therefore, by limiting the analysis to municipalities that held elections before gender quotas were introduced, we can isolate the effect of having more politicians alone.

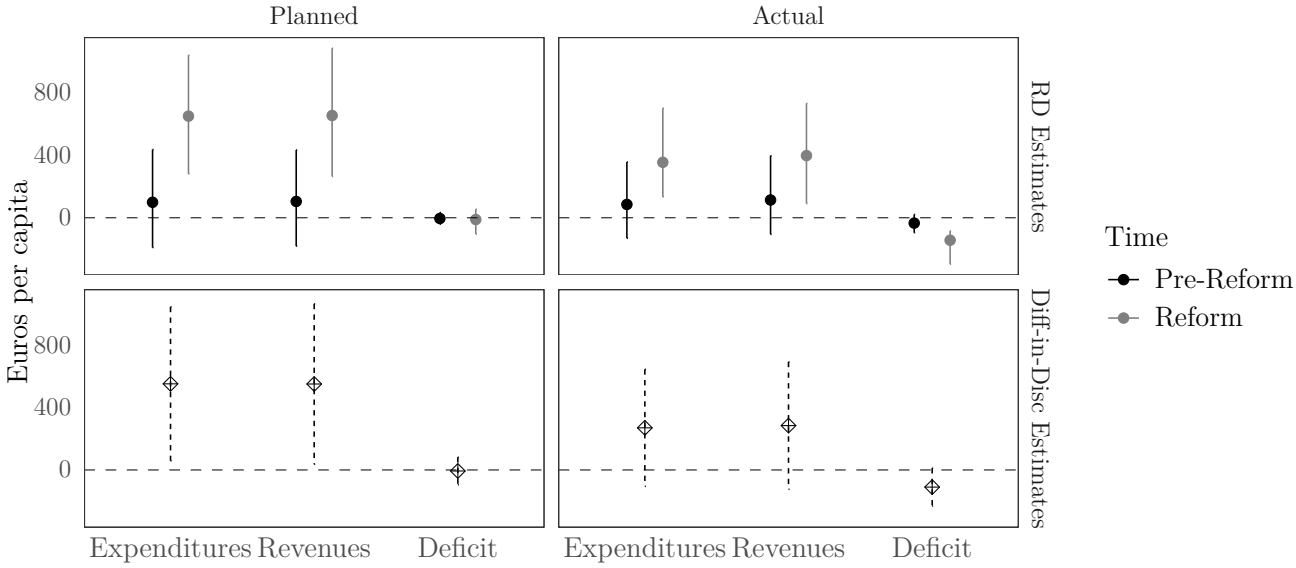

Figure I.6: RD and Diff-in-Disc estimates from sample of municipalities which held elections earlier than 26 December 2012, when the gender quotas entered into force for municipalities above the 5,000 population threshold. Estimates constructed using local polynomial estimators with triangular kernel and MSE-optimal bandwidth. Robust confidence intervals constructed using bias correction with robust standard errors. No covariates included.

I7 Types of Revenues

In Figure I.7 we show the results for planned and actual revenues per capita are not driven by one particular type of revenues.

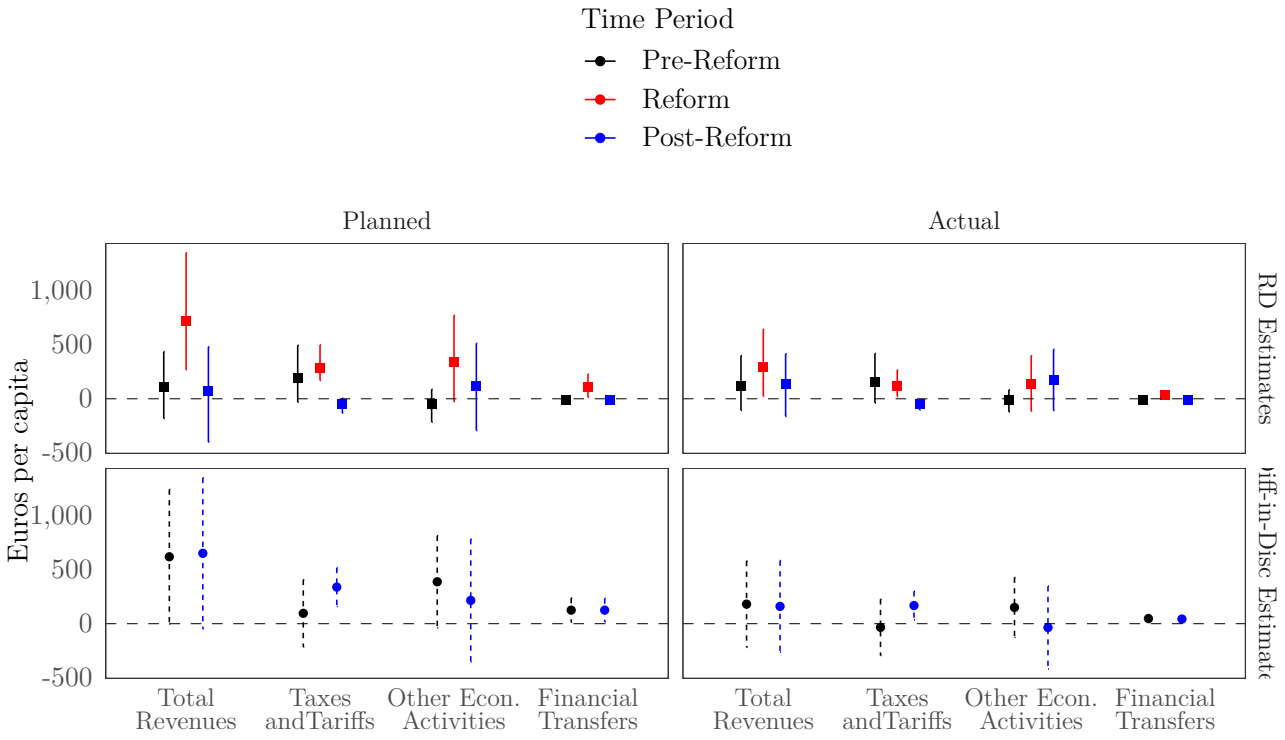

Figure I.7: RD and Diff-in-Disc estimates and robust 95% confidence intervals. Outcomes are revenues per capita and the three components thereof: taxes and tariffs, financial transfers from higher levels of government, and other economic activities of the municipalities. No covariates included.

I8 Selection Effect of the Reform

Table I.18 reports the effect of the reform on the share of councilors and members of the executive committee with a university degree.

	Reform - Pre-Rerom		**Reform - Post-Reform**	
Outcome	% Councillors with Degree	% Members of Ex. Comm. with Degree	% Councillors with Degree	% Members of Ex. Comm. with Degree
Difference	0.12	0.06	0.15	0.03
SE	(0.04)	(0.02)	(0.04)	(0.02)
p.value	0.001	0.000	0.000	0.089

Table I.18: Diff-in-Disc estimates of the effect of the reform on the share of councillors and members of the executive committee with a university degree. Same covariates included in main analysis.

I9 Diff-in-Disc Estimated with Single Equation

In Table I.19 below we show the diff-in-disc estimate by estimating a full equation with two time periods (municipalities holding elections under the reform and not, as in Grembi et al. (2016). We fit local WLS models separately on the observations above and below the cutoff and for municipalities holding elections when the reform was into force ($T = 1$) and not ($T \in \{0, 2\}$). Weights are determined by the triangular kernel function based on the ratio between the distance of unit i from the cutoff and the MSE minimizing bandwidth. We estimate the optimal bandwidth pooling all time periods but results are robust to estimating two different bandwidths in $T = 1$ and $T \in \{0, 2\}$ and then averaging the two. Units outside the optimal bandwidth receive a weight equal to zero. We then estimate the following equation:

$$\begin{aligned} Y_{it} = {} & \delta_0 + \delta_1 X^*_{it} + S_i(\gamma_0 + \gamma_1 X^*_{it}) \\ & + T_t[\alpha_0 + \alpha_1 X^*_{it} + S_i(\beta_0 + \beta_1 X^*_{it})] + \eta_{it} \end{aligned} \tag{1}$$

where S_i is a dummy for treated units above the cutoff, T_t is the post-period indicator and equals 1 when $T = 1$ and 0 otherwise, X^*_{it} is the normalised running variable ($X_{it} - 5,000$) and η_{it} the error component. The coefficient β_0 is the diff-in-disc estimator and identifies the effect of electing more politicians.

	Planned Budget			Actual Budget		
	Expenditures (1)	Revenues (2)	Deficit (3)	Expenditures (4)	Revenues (5)	Deficit (6)
Above 5,000 × Reform	587.4**	560.5**	-3.7	198.8	116.8	35.7
	(237.2)	(240.2)	(37.7)	(148.0)	(147.1)	(52.9)
Observations	1,355	1,428	1,198	1,440	1,668	995
R^2	0.03	0.03	0.00	0.01	0.01	0.02
Adjusted R^2	0.03	0.02	0.00	0.01	0.01	0.01

Table I.19: Diff-in-Disc Analysis with One Single Equation. Diff-in-Disc estimates. Estimation performed using WLS with triangular kernel and MSE-optimal bandwidth. No covariates incluided.

J Validity of RD Estimator

In this section we report tests in support of the continuity of density and potential outcomes assumptions, showing there are no discontinuities in the density function of the running variable (Figure J.8), in a set of pre-treatment covariates (Figure J.9), and at most placebo cutoffs (Figure J.10).

J1 Continuity of Density

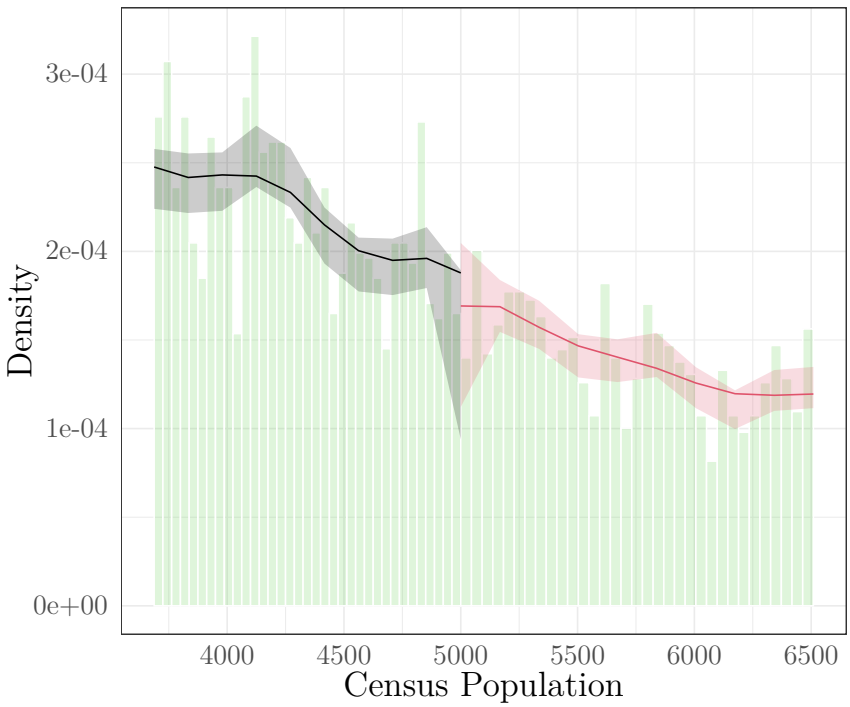

Figure J.8: Manipulation test using the local polynomial density estimators proposed by (Cattaneo et al., 2020). Histogram estimate of the running variable computed with default values in R; local polynomial density estimate (solid dark and red) and robust bias-corrected confidence intervals (shaded dark and red) computed using *rddensity* package in R. The number of observations just above the cutoff is not significantly different from the number of observations just below the cutoff (p.value = 0.50).

J2 Continuity of Potential Outcomes

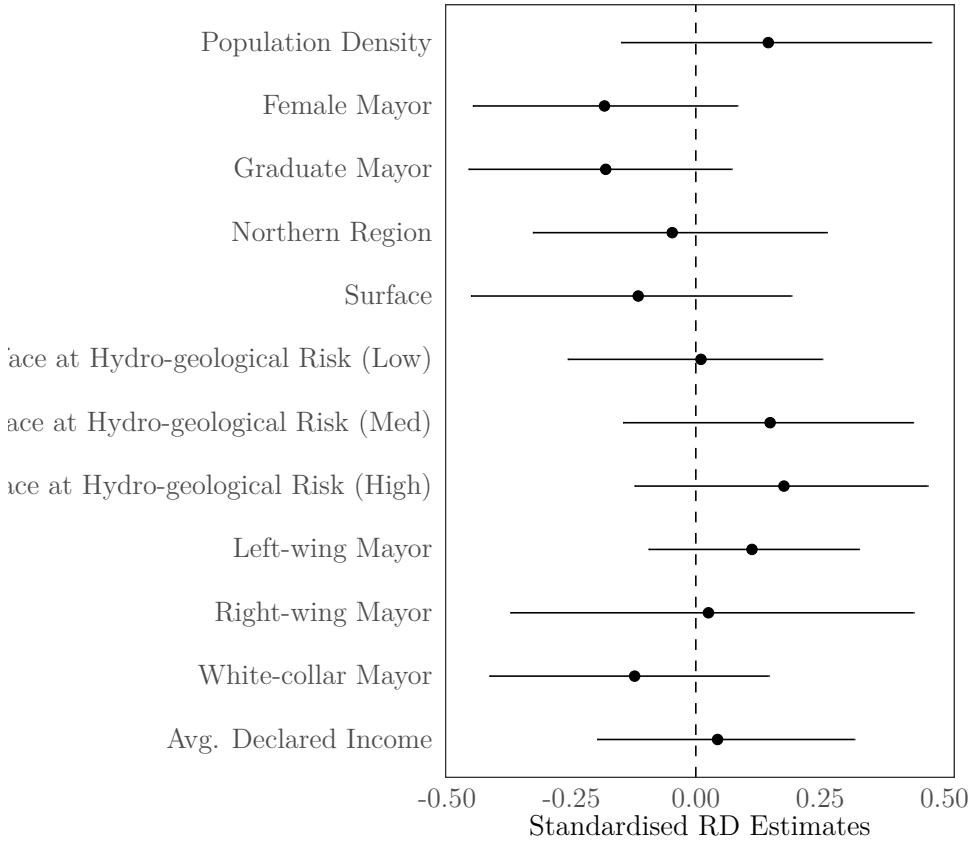

Figure J.9: Standardised RD estimates of the effect on pre-treatment covariates with 95% robust confidence intervals. Estimates constructed using local polynomial estimators with triangular kernel and CER-optimal bandwidth (as suggested by (Cattaneo et al., 2019, Ch. 5). No covariates included in the estimation. Variables used as outcomes are population density, female mayor (dummy), graduate mayor (with university degree, dummy), northern region (dummy), surface (sq.km), surface at low, medium, high hydro-geological risk (sq.km), left-wing mayor (dummy), right-wing mayor (dummy), white-collar mayor (dummy) average declared personal income of residents.

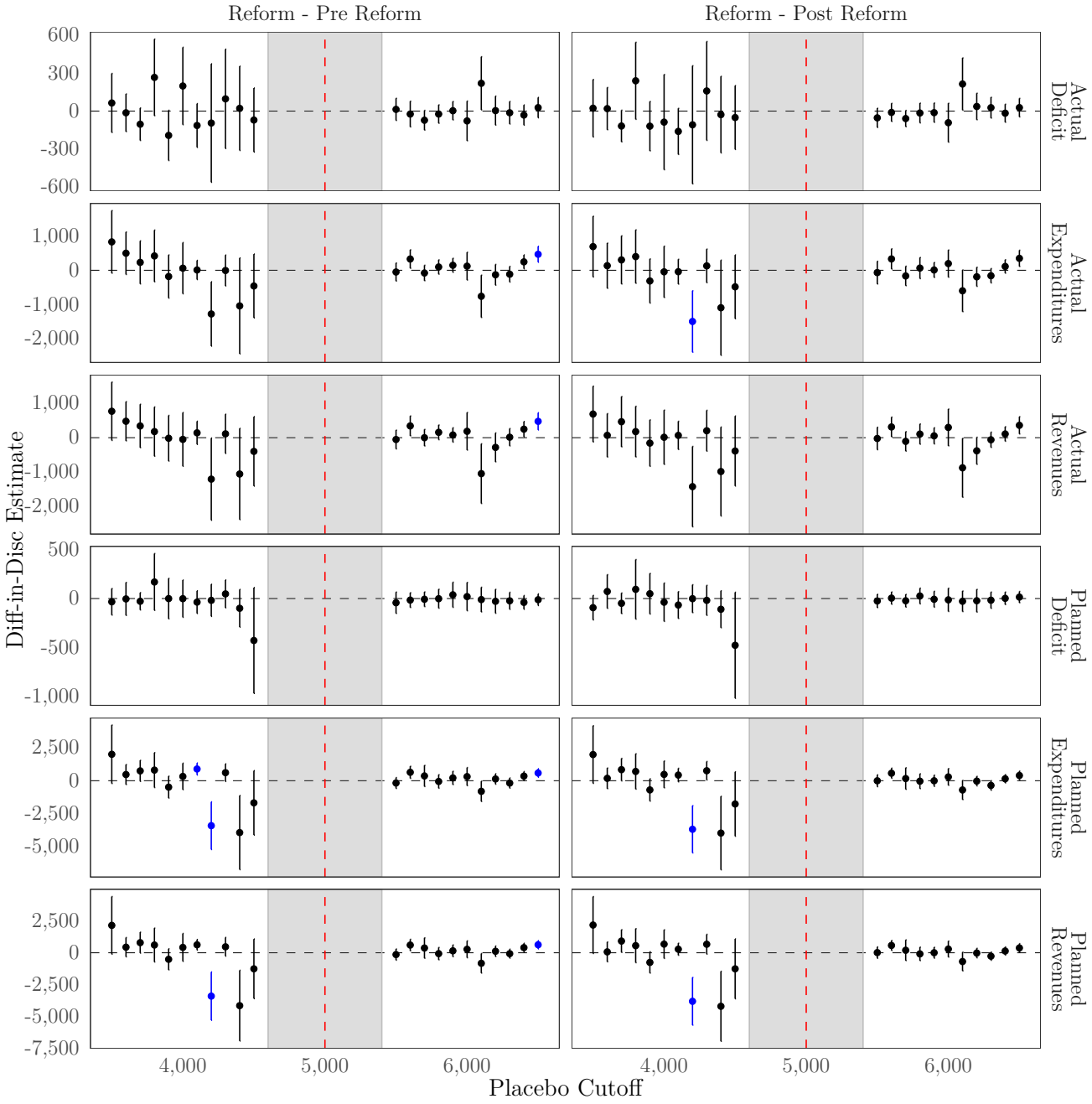

Figure J.10: Diff-in-Disc estimates with 95% robust confidence intervals. Red dashed line at the true cutoff. Blue coefficients when p.value after multiple testing adjustment smaller than 0.05. Multiple-testing adjustment performed separately for each outcome variable with Bonferroni procedure to control for the false discovery rate. Estimates constructed separately on control unit when placebo cutoff < 0, and on treated unit when placebo cutoff > 0. Placebo cutoffs very close to 0 omitted due to small sample size. Estimation performed using local polynomial estimators with triangular kernel and MSE-optimal bandwidth. Confidence interval constructed using robust standard errors. No covariates included. We fail to detect a discontinuity statistically significant effects in 96.2% of the tests.

K Validity of Diff-in-Disc Estimator

To test the assumption that politicians paid differently do not react differently to a change in the number of politicians, we compare discontinuities at the 5,000 cutoff before and after a "placebo" reform that changed the size of government bodies but *not* on a population threshold basis. Municipalities above and below the 5,000 cutoff – who are paid differently – experienced a decrease in the number of politicians by 20% as a result of the placebo reform. If a change in the number of politicians (both above and below the 5,000 cutoff) affected the effect of wage treatments on the outcomes, we should detect a significant difference in the discontinuity at the cutoff before and under the placebo reform. Figure K.11 below shows that the difference between the RD estimates in the pre and placebo reform periods is not distinguishable from 0 for all the outcomes (see Table K.20 for full regression table). A decrease in the number of politicians therefore does not change the effect of wage policies for municipalities above the 5,000 cutoff.

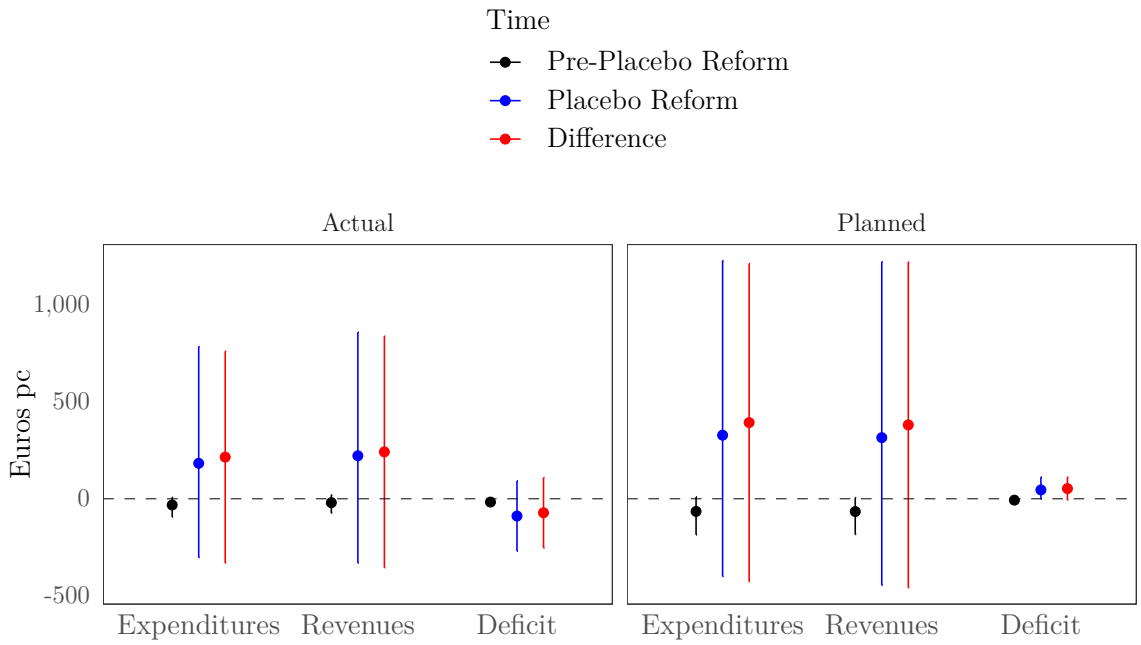

Figure K.11: Diff-in-Disc estimates with 95% robust confidence intervals estimated with alternative bandwidths. Red coefficients estimated with MSE-optimal bandwidth. Underlying RD estimates constructed using local polynomial estimators with triangular kernel and MSE-optimal bandwidth. Robust confidence interval constructed using bias correction with robust standard errors. No covariates included.

	Planned Budget			Actual Budget		
Outcome	Expenditures	Revenues	Deficit	Expenditures	Revenues	Deficit
Pre-Placebo Reform (RD)						
Estimate	-65.0	-65.8	-6.9	-31.9	-20.0	-16.8
SE	(49.9)	(48.4)	(7.4)	(25.9)	(24)	(8.4)
p.value	0.078	0.066	0.428	0.094	0.266	0.038
h	306.4	293.0	733.8	322.3	411.7	699.8
Obs. Used	2449	2301	5680	2585	3194	5412
Placebo Reform (RD)						
Estimate	328.4	315.8	45.8	183.2	222.3	-88.9
SE	(415.9)	(426.2)	(29)	(277.8)	(304)	(92.5)
p.value	0.319	0.362	0.056	0.385	0.385	0.337
h	817.7	772.8	407.4	899.1	1191.3	824.2
Obs. Used	180	164	90	216	276	180
Difference (Diff-in-Disc)						
Estimate	393.5	381.6	52.8	215.1	242.2	-72.1
SE	(418.9)	(428.9)	(30)	(279)	(304.9)	(92.9)
p.value	0.348	0.374	0.078	0.441	0.427	0.437

Table K.20: Regression table of RD results in the pre-placebo reform and placebo reform time periods as well as diff-in-disc estimates (difference in RD point estimates in the pre- and placebo reform periods) showing no statistically significant difference between the two time periods at the cutoff, suggesting that municipalities above the cutoff (paid differently) did not react differently from those below the cutoff to a same-size change in the number of politicians. No covariates included.

The "local" parallel trend assumption is indirectly tested in Figure K.12, where we estimate the discontinuities in all the outcomes for every year and show that they are highly stable in the pre reform period. Furthermore, as we show in the *Results* section, the RD effects in the reform period are very similar in the pre- and post-reform periods, suggesting that, after the reform is repealed, changes at the discontinuities return to pre-reform levels.

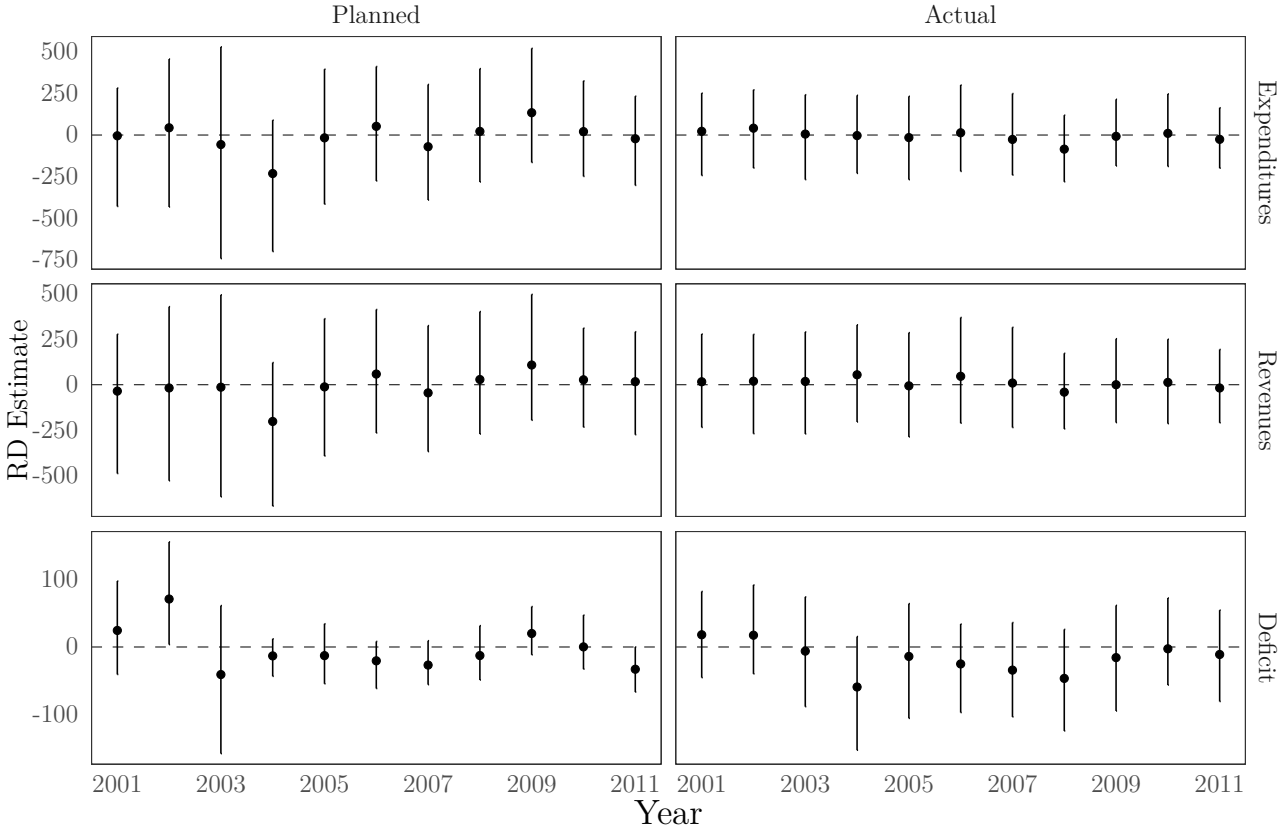

Figure K.12: RD estimates with 95% robust confidence interval for every year and every outcome in pre-treatment period (before Reform enters into force). MSE-optimal bandwidth, and triangular kernel. No covariates included.

References - Online Appendix

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2015), 'Rdrobust: An R package for robust nonparametric inference in regression-discontinuity designs', *R Journal* **7**(1), 38–51.

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (2019), *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Vol. I, Cambridge University Press.

Cattaneo, M. D., Jansson, M. and Ma, X. (2020), 'Simple Local Polynomial Density Estimators', *Journal of the American Statistical Association* **115**(531), 1449–1455.
URL: *https://doi.org/10.1080/01621459.2019.1635480*

Grembi, V., Nannicini, T. and Troiano, U. (2016), 'Do fiscal rules matter?', *American Economic Journal: Applied Economics* **8**(3), 1–30.

Klašnja, M. and Titiunik, R. (2017), 'The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability', *American Political Science Review* **111**(1), 129–148.

Morelli, M. (1999), 'Demand competition and policy compromise in legislative bargaining', *American Political Science Review* **93**(4), 809–820.