
# An impact assessment of measures for gender rebalancing in local elective assemblies.

*Ugo Finardi, Elena Ragazzi*, and Lisa Sella*

CNR-Ircres

Via Real Collegio 30

10024 Moncalieri (TO)

*Corresponding author elena.ragazzi@ircres.cnr.it

Paper to be presented at the special session S51 Counterfactual methods for regional policy evaluation

## Abstract

The general objective of the paper is to present the results of an evaluation of the effects produced by the regulatory measures introduced in recent years for the purposes of gender rebalancing in the elective offices.

Women still encounter huge difficulties in being included in political decision-making. Even when they have the chance of participating to political competition as candidates, this is not a guarantee that they are elected and assigned to important political positions.

Many countries have introduced rules and mechanisms to accelerate the rebalancing process. The general feeling is that their effectiveness may vary a lot, because they need to be well designed and effectively implemented to achieve good results, so asking an important evaluation question.

Different types of mechanisms may be designed to ensure gender balance in politics.

The main adopted options are: **Gender shares** in the lists of candidates; **Double/**triple **gender preference**; **Alternate list**. The three mechanisms can be paired, and their effectiveness depends on their The impact evaluation is applied to the case of municipalities. In the other cases, the lack of a credible counterfactual

forces to limit oneself to statistical analysis of observed change. On the other hand, in the depicted situation, the voting regulation for the administrative elections is a perfect context where to evaluate the effectiveness of gender rebalancing mechanisms.

The paper presents the results of an evaluation based on RDD approach around the threshold present in the law (5.000 inhabitants). The dimension of the window around the discontinuity is established through different techniques. The paper includes also falsification tests on covariates.

The analysis shows very robust results: the impact is significant and positive for all the adopted approaches, even though the size of the impact varies.

Keywords: counterfactual impact evaluation, gender issues, regression discontinuity design

JEL codes: K38, J16, D72, H89

# 1 Introduction

## 1.1 Gender issues in politics

The general objective of the paper is to present the results of an evaluation of the effects produced by the regulatory measures introduced in recent years for the purposes of gender rebalancing in the elective offices.

Following the 2017 report on equality between women and men in the European Union, women continue to be under-represented in decision-making positions at all levels (European Commission 2017), and the gender shares in elective assemblies are far from representing the demographic situation.

Equality in decision-making no longer seems a distant goal for a few countries that are close to gender balance. However, great regional deviations persist in Europe and worldwide (Inter-Parliamentary Union 2017), and in many countries, the share of women in politics is extremely low. Italy is located close to the average EU value.

The average share of women members in the single/lower houses of national parliaments in the EU has increased from 22.1 % in 2004 to 28.7 % in November 2016. Therefore over time, a tendency towards better equilibrium can be detected, but it is very slow, with an average increase of just over half a percentage point per year. Women still encounter huge difficulties in being included in political decision-making. Even when they have the chance of participating to political competition as candidates, this is not a guarantee that they are elected and assigned to important political positions.

## 1.2 Policies for gender balance in politics

Many countries have introduced rules and mechanisms to accelerate the rebalancing process. The general feeling is that their effectiveness may vary a lot, because they need to be well designed and effectively implemented to achieve good results, so asking an important evaluation question.

Different types of mechanisms may be designed to ensure gender balance in politics.

Only a minority of countries all over the word (23) adopt quotas in their most binding form, i.e. with the provision of seats reserved for women in assemblies. This happens above all in emerging constitutional democracies in Africa and Asia, where such a binding measure was necessary to build a new vision of the role of women. [1]

In more mature democracies, the choice is generally to design the voting rules to ensure the participation of the minority gender to electoral competition, and through this to increase women shares in elective assemblies. This happened in Italy too, as in most European countries, driven by the agreed necessity to respect the will of the voter.

Nevertheless, one should note that the concept of “will of the voter” is becoming in Italy a lash concept at the national level, because with the two last electoral reforms the role of political parties is becoming more and more important in deciding which representatives are going to be elected. On the other hand, at the local level (regions and municipalities) the vote is still expressed through preferences and so the outcomes of the elections are directly influenced by overall mentality.

The main options adopted in voting regulation are:

- **Gender shares** in the lists of candidates. The list must be made up of not more than a certain quota of either of the two sexes. The rule is more or less compelling depending on the size of the compulsory share but even more on the severity of the sanctions applied if the rule is not observed (cancellation of the candidates of the exceeding gender, pecuniary fee, cancellation of the whole list);
- **Double/**triple **gender preference**, which allows voters to express two/three candidates as long as they are of different sexes. If this rule is not respected, the second/third preference is cancelled;
- **Alternate list** (also called zebra lists), which provides that the electoral lists are completed by alternating the name of a man with that of a woman. If the voting happens through the expression of preferences, this mechanism is equivalent to a 50% gender share, while it affects directly the election outcome in case of fixed lists.

The three mechanisms can be paired, and their effectiveness depends on their application with different voting regulation.

In the depicted situation, the voting regulation for the administrative elections is a perfect context where to evaluate the effectiveness of gender rebalancing mechanisms, because

- The outcome in terms of elected representatives is not a mechanical output of the voting rule (such is the gender share observed in candidates when a gender share limitation is fixed), but is mediated by a change in the parties’ and voters’ mentality.
- The features of the gender regulation for the elections of municipalities (time enforcement, threshold) offer the chance for counterfactual impact evaluation
- The number of municipalities (7,964) allows statistical assessment of the impact

[1] On the other hand, gender shares are very frequently introduced even in mature democracies for executive assemblies.

## 1.3 Past studies on gender balance in politics

Some authors have tackled the topic of gender balance in politics in the past under different points of view, both in general and related to specific electoral results. The corpus of scientific works specifically addressing quantitative research on gender balance in politics is not so vast notwithstanding the long research path on gender issues in political representativeness, and more in specific on the effects of gender quota laws. Within this theoretical framework, besides addressing specific cases of application of gender quotas in electoral laws, we will more specifically discuss the methodology exploited by authors in evaluating their effect. We do this in order to better frame our research effort.

Two studied extra-european cases are those of South Korea and Argentina. (Yoon & Shin, 2017) study the case of South Korean Municipal election with the aid of descriptive statistics, and with specific reference to 2006 electoral reform. They conclude saying that the reform has positively impacted on the number of female members in municipal councils, leading to changes in party recruitment and candidate selection. Another similar analysis has been performed by (Jones, 1998) on the case of Argentina's provinces. The author exploits an OLS model using data of 83 provinces, holding the percentage of elected women as dependent variable. The analysis of regression results show that "gender quota laws have had a significant positive effect on the percentage of women legislators elected" (p. 17).

Coming closer to the main topic of our work we discuss now, more specifically, studies on European cases of gender balance in elections. A first case to address is that of Ireland. This European case is described in the work of (Brennan & Buckley, 2017). These authors exploit the case of Irish legislative gender quotas. Such quotas were introduced in 2012, and the first election to be held under the regulation were 2016 general elections. Authors analyse data using both a specific index of marginality of candidates with respect to their placement in electoral seats, and a logistic regression. According to authors, results show that "the political parties did embrace the quota and that the integration of the formal gender quota in candidate selection processes somewhat mollified localism and informal, male-gendered institutional norms" (p. 32). Nevertheless, resistance to gender quotas is still present. This problem was addressed, prior to the 2016 general elections, also by (Buckley, Marian, & White, 2014).

(Verge & Wiesehomeier, 2018) tackle instead the female political representation in Spanish closed-list proportional system. They perform an empirical analysis to measure the strategic discrimination applied against woman in these specific systems. In Spanish political system "gender party quotas" (that is, voluntary gender quotas enforced by parties) did exist prior to the 2007 gender quota law (60/40 quotas). The dataset covers nine national elections (1986-2016). The methodology exploits OLS regressions at party level using as dependent variable the number of female candidates. Results show that women are discriminated across all political parties.

Polish elections are addressed in some recent papers. (Millard, 2014), after describing the historical path of gender issues in Polish politics, comments data on numbers of candidate and elected women in Polish political parties. He then concludes that "2011 election confirms the importance of list placement, incumbency, party magnitude, and ideology in explaining the effects of Poland's new quota system" (p. 9), while "the role of the open list is less clear". Nevertheless "still it seems premature to judge the new quota system as a failure" (p. 10). A quantitative study is instead performed by (Gorecki & Kukołowicz, 2014) who base their analysis on a comprehensive database of candidates. The Polish system is an open-list system: that is, while gender quotas are mandatory (35/65), the ranking of candidates in lists is completely devoted to the will of the parties. The dataset encompasses data on all the candidates to the 2007 (last pre-quota) and 2011 (first post-quota) elections (6187 and 7035 respectively). A logistic regression models the list

ranking of candidates, and a negative binomial the vote patterns. The results do not differ much from the statements of (Millard, 2014). In fact, in an open-list proportional representation, the “paradoxical effect” of gender quotas is an increase in the number of female candidates coupled to a lower electoral performance. Similar results are obtained also by (Jankowski & Marcinkiewicz, 2019) who show that quotas had little impact on the number of preferential votes cast for female candidates.

The Norwegian case of gender quota reform is the subject of the study of (Geys & Sørensen, 2019). Gender quotas exist in Norway since 1992, though parties introduced internal quotas prior of law enforcement. Authors employ a difference-in-difference methodology, comparing the two periods before and after the reform. To this end, they introduce the concept of “quota shock”, that is, the change in proportion of female representatives in the municipality executive boards. They conclude that the reform generated a growth in the share of female representatives in executive boards, but had only very weak effects on the fraction of women in other positions (municipal councils, mayors, etc.).

The French parity law is investigated by (Achin, 2018). The author comments on election results concluding that on the one side parity law has increased numeric representation of women in politics but, on the other side, discrimination mechanisms are still present and hinder the leading role of women in assemblies, also making women’s careers more fragile. Moreover, “the implementation of the parity law has contributed to reproducing a conservative gender order […] and has failed to tackle the preservation of male power and in-group sociability” (p. 196). Also Germany has enforced a parity law relatively early. (Hennl & Kaiser, 2008) analyse and compare sub-national (German Länder) elections, held with different electoral systems. Their aim is to disentangle the systems of list nominations in order to understand the relation between presence in “safe” list places and being elected. The cross sectional database is analysed with a multiple OLS-regression to find out that “the share of women in safe list positions is rather a function of voluntary party quotas and a minimum party magnitude of two” (p. 334).

The most relevant studied case to address in this specific theoretical framework is obviously the Italian one. (Palici di Suni, 2012) studies the Italian case, addressing specifically the 1990s reform that introduced gender quotas in 1993. It must be noted that the law was soon withdrawn in 1995 (due to unconstitutionality). Palici di Suni concludes his work saying that “Ensuring greater gender balance in Italian politics requires a change of mindset on the part of politicians and citizens and a renewal of the political system” (p. 391). The work of (Palici di Suni, 2012) is one among those (not so many) works that tackle the issue of the effects of gender quotas in Italy. Actually most of them, to the best of our knowledge, exploit data of the 1993-1995 reform, as (Palici di Suni, 2012) did. These data are for instance studied by (De Paola, Scoppa, & Lombardo, Can gender quotas break down negative stereotypes? Evidence from changes in electoral rules, 2010) and by (De Paola, Scoppa, & De Benedetto, The impact of gender quotas on electoral participation: Evidence from Italian municipalities, 2014). In the former, authors exploit the natural experiment methodology. Their econometric analysis shows that female representation has increased more in those municipalities that have been affected by the reform with respect to those where the reform had never had effect (as no elections were held in the two years of the reform). In the latter of the two works, the authors exploit a difference-in-difference methodology, and analyse also those elections held *after* the withdrawn of the reform. Results sadly show that “the positive effect of gender quotas on turnout is limited to the period in which they were in force, while the effect disappears after their removal” (p. 154).

The effects of the same electoral law is also investigated in the study of (Catalano Weeks & Baldez, 2015). In specific the aim in this case is to understand if candidates that are nominated in consequence of gender

quota law are less qualified. Results show that, on the contrary, "quota women are not less qualified than men or than other women in terms of education, occupational achievement, or previous political experience at the local level" (p. 139).

The Italian case is tackled also by (Sartori, Tuorto, & Ghigi, 2017) with an original perspective. The aim of the authors in fact is not limited to political participation gender gap but, on the contrary, expands its vision to the general exclusion of women from the public sphere. From the methodological point of view the paper models data derived from ISTAT (Italian National Statistical Institute). The four dependent variables – visible political participation, invisible political participation, participation in social activities, and involvement in leisure activities – are modelled using distinct negative binomial regressions for men and for women. The results show that time constraints are effective in hindering the participation of women in both political and non-political activities.

## 2 Object of the evaluation

The evaluation exercise we present here is based on an evaluation report prepared by the National Research Council[2] for the Italian Department of Institutional Reforms (belonging to the Presidency of the Council of Ministers), aiming at studying the impact of gender balancing norms at different institutional levels. The report (in Italian) is available at: https://www.irpps.cnr.it/wp-content/uploads/2018/03/Sveglie_rapportofinalePartePrima.pdf

The report concerned all political election, including Italian representatives to European Parliament, Italian Parliament (Deputies chamber and Senate), Regions and Municipalities. The time-span of available data covers the period 2000-2016 (2009 onwards for the municipalities, before that year data by gender were not available), and allows observing changes in the target variables: share of female candidates, share of female elected representatives, and women' success index.

This paper focuses on the elections of municipal assemblies. The law, approved at the endo of 2012 and applied for the first time for 2013 elections, establishes the following prescriptions:

- Parties or civic movements in municipalities with a population **smaller than 5.000** people are just asked to present lists with candidates of both genders (and no sanction is imposed in case the rule is not respected). Electors in municipalities can express a single preference when voting. No gender voting mechanism is present for this category.
- For municipalities **over** the threshold of **5,000 inhabitants** there is a prescription on the gender composition of lists: parties have to present a list with not more than two thirds of either of the two sexes. In case the list doesn't comply to this prescription there are sanctions that depend again on the size of the municipality:
    - For municipalities with a size included between 5,000 and 15,000 inhabitants, there is just the cancellation of the candidates of the exceeding gender. Candidates are cancelled until the share is respected or until the minimum number of candidates in a list is reached. If this happens, the **list is not cancelled** and it remains unbalanced respect to the required gender share.

[2] In particular, two institutes participated to the service: IRPPS and IRCRES.

- For municipalities with a size exceeding 15,000 inhabitants, candidates are cancelled until the share is respected. If doing so the minimum number of candidates in a list is reached, the **list is cancelled.** So the lists are obliged to respect the gender quota, otherwise the list is canceled.

- Concerning the voting procedures, for municipalities **over 5,000 inhabitants** there is the double gender preference, which allows voters to express two candidates as long as they are of different sexes. If this rule is not respected, the second preference is cancelled.

This voting regulation, which prescribes different provisions following the size of the municipality, is a perfect context to apply a regression discontinuity design (RDD). The hypothesis that allows the application of RDD is that the municipalities just above or under the threshold, are extremely similar to each other, differing (on average) only for the application of the law. In fact, the attribution to the main group of treated (municipalities immediately above the threshold of 5,000 inhabitants applying the legislation) or to the counterfactual (municipalities with a population less than 5,000 inhabitants) is independent of the variables that influence the dependent variable (the share of elected) and can therefore be considered random. The falsification tests, showed in section 5.1, confirm the respect of this hypothesis.

In RDD counterfactual evaluation, the interpretative limitation of the approach regards its external validity. In fact, it allows verifying whether there is a positive impact of the legislation and to quantify it, but only for classes of municipalities that do not differ too much from the application threshold. Indeed, descriptive analysis shows that there are significant differences in the gender balance in municipalities of different sizes.

Going more into the details, the described policy design results in two different thresholds:

- Around the municipality size of 5.000 inhabitants, the gender share and the double preference are introduced. From previous analysis, we expect that the impact around this threshold is positive and significant.

- Around the municipality size of 15,000 inhabitants, more severe sanctions to the non-compliant lists are introduced. Even if we have never verified this impact, we expect that the impact around this threshold non-significant for elected representatives, while it could have a low effect on the gender share of candidates. Moreover, we expect that this impact diminish as long as time passes, since political parties become more smart in addressing their objectives while respecting the rule. The study of this impact, which is not independent from the first threshold, represent a future direction of research.

# 3 IMethodological framework

The impact evaluation is limited to the case of municipalities and it is based on the Regression Discontinuity Design (RDD) framework (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). This approach allows quasi-experimental causal inference when the treatment is not randomized, but it is assigned when a running variable exceeds some known cutoff. In the present case, the running variable is represented by resident population, since the electoral norm is applied when population exceeds 5,000 units. This is a sharp RD

design, where the probability of treatment changes abruptly from 0 to 1 at the cutoff: treated municipalities are larger than 5,000, and control municipalities are smaller. Moreover, since further reinforcement rules act when population is larger than 15,000, a second cutoff can be identified. It will be specifically investigated a future extension of the work.

The RD framework assumes that units near the cutoff are valid counterfactuals, i.e. municipalities are not able to precisely manipulate their population. This non-sorting assumption can assume two different versions, that originate two different frameworks of analysis. This section sketches these two approaches using the potential outcome frame[3], i.e. each unit *i* has two potential outcomes $Y_i(1)$ and $Y_i(0)$, respectively denoting the outcome of the unit when the treatment is either received or not. In our case, $Y_i(1)$ is the share of female council members when gender electoral mechanisms are applied and $Y_i(0)$ the share of female council members when the norm is not applied. Thus, the observed outcome is

$$Y_i = Y_i(0) \cdot (1 - D_i) + Y_i(1) \cdot D_i = \begin{cases} Y_i(0) \; if \; D_i = 0 \\ Y_i(1) \; if \; D_i = 1 \end{cases}. \tag{1}$$

## 3.1 Local polynomial estimation methods

The first RDD approach is based on the assumption that if the density of the running variable is continuous at the cutoff, the (observable and unobservable) characteristics of units near the cutoff should not differ so much, thus implying the comparability of treated and control groups, as well as the continuity of the conditional expectation functions of the potential outcomes.

In the case of municipalities, it is reasonable to assume that population is continuous at the cutoff, but we can test this proposition by providing a set of falsification tests. In particular, there are three common falsification approaches (Cattaneo *et al.*, 2017, 2018): the first one is global in nature and it graphically inspects the continuity of outcome variables away from the cutoff over the full support of the running variable (Calonico *et al.*, 2015). The second falsification approach tests for manipulation of the running variable by verifying through binomial test whether the number of treated and control units is similar near the cutoff (McCrary, 2008). In fact, if units within a certain windows were randomly assigned to treatment as Bernoulli trials with a pre-specified probability $q$ (generally $q$=1/2), then the number of effective treatment and control units should follow a binomial distribution. The third falsification approach is based on placebo tests, that verify whether pre-intervention covariates and post-intervention variables that are not affected by the treatment do exhibit a zero treatment effect (Lee, 2008).

The continuity assumption justifies the identification at the cutoff of a local average treatment effect (ATE), i.e. the difference between the average potential outcome under treatment and control at the cutoff $r$, and the use of local polynomial techniques for estimation and inference (Hahn *et al.*, 2001). In fact, if potential outcomes are continuous at the cutoff, abrupt changes in the observed outcomes can be attributed to the sharp RD assignment. Hence, potential outcomes are regarded as random variables and the observations as a random sample from a (super)population.

[3] Each unit is assumed to have a potential outcome for (a subset of) all possible treatment assignments. The treatment effects are defined in terms of the underlying potential outcomes, which can be modelled either as fixed or random. In the first case, the large-sample approach applies, while in the second case randomization inference is used (Rosenbaum, 2010).

In this context, the main parameter of interest is the ATE at the cutoff *r*: $\tau_S = E[Y_i(1) - Y_i(0)|R_i = r]$, which is estimated within an interval by either parametric or nonparametric identification strategies. In practice, the two conditional regression functions $E[Y_i(1)|R_i = r]$ and $E[Y_i(0)|R_i = r]$ are either modelled or non-parametrically approximated near the cutoff relying on large sample theory (Calonico *et al.*, 2014). A fundamental concern is whether the (unknown) polynomial regression functions are well approximated on each side of the cutoff.

In particular, flexible parametric techniques fit by least square regression a polynomial function of order *p* within a window *W*. The full-treatment interaction model is given by

$$Y_i = \alpha + \tau_S D_i + \beta_1 \bar{R}_i + \cdots + \beta_p \bar{R}_i^p + \varphi_1 \bar{R}_i D_i + \cdots + \varphi_p \bar{R}_i^p D_i + \varepsilon_i, \qquad (2)$$

where $\bar{R}_i = R_i - r$ is the running variable re-centred around the cut-off. This approach is subject to both misspecification bias and neighbourhood selection, since both *p* and *W* are selected *ad hoc* rather than by data-driven procedures.

On the contrary, nonparametric local polynomial techniques are based on nonparametric large-sample inference and approximations. In particular, the bandwidth *h* is chosen by means of data-driven nonparametric procedures, the RD point estimate is asymptotically Mean-Squared-Error (MSE) optimal, and inference procedures are based on robust statistics that explicitly incorporate the effect of the nonparametric smoothing bias (Calonico *et al.*, 2014).

In practice, the estimation procedure is similar to the parametric one, apart from the use of kernel weights increasing the relative weights of observations closer to the cutoff. Practically, the function in equation 2 is weighted by a kernel function $\mathrm{K}(\bar{R}_i/h)$, where *h* is the data-driven bandwidth, $\bar{R}_i$ is the re-centered running variable, and a triangular kernel has the form $\mathrm{K}(u) = 1 - |u|$. However, the data-driven selection of the bandwidth *h* accounts for the presence of misspecification bias in both point estimation and inference. Concerning point estimation, the MSE-optimality criterion allows to balance the trade-off between bias and variance[4] of the point estimator $\widehat{\tau_S}$ (Imbens and Kalyanaraman, 2012; Calonico *et al.*, 2014). Concerning inference, it is known that the MSE-optimal bandwidth $h_{MSE}$ is too large to counteract the misspecification bias in the distributional approximation of the estimator. Then, Calonico *et al.* (2014) propose to directly incorporate the bias in the distributional approximation by calculating p-values based on a robust bias-corrected t-test statistic.

## 3.2 Local randomization estimation methods

The second RDD approach derives from a bit stronger assumption of local randomization, i.e. the treatment is assigned as-if-random in a small window near the cutoff, justifying the use of experimental methods for estimation and inference (Cattaneo *et al.*, 2015). In this case, the main concern is about window selection, rather than approximation of the regression functions. Furthermore, this approach can accommodate cases where the running variable has a direct effect on the potential outcomes, i.e. when standard local SUTVA[5] is violated (Cattaneo *et al.*, 2017).

[4] The smaller *h*, the smaller the estimation variance but the larger the small-sample bias. On the contrary, larger *h*s reduce the small-sample bias but increase the variance.
[5] It is the Stable Unit Treatment Value Assumption (SUTVA), i.e. units do not interfere with each other.

In particular, inside the selected window it is assumed that: potential outcomes are non-random; the distribution of the observed running scores does not depend on potential outcomes, i.e. no selection into treatment occurs; the assignment mechanism is known. Within this framework, the sharp null hypothesis of no effect can be tested using finite-sample randomization-based inference[6] (Fisher, 1935). This means that the treatment has no effect on any unit inside the window and that under this hypothesis an exact p-value can be computed for any observed value of the test statistic.

In practice, first of all both the appropriate window $W_0$, where local randomization is assumed to hold, and a model of potential outcome adjustment, e.g. constant ($p$ = 0) or linear ($p$ = 1) regression, are chosen, which give the adjusted outcomes $\tilde{Y}$. Then, a treatment assignment mechanism is assumed, e.g. complete randomization. Finally, a test statistic is selected, e.g. difference-in-means, and the finite-sample exact p-value for the sharp null hypothesis is calculated via permutation of the treatment status of units (Gerber and Green, 2008).

When the difference-in-means is used, the ATE can be estimated as

$$\tau_{W_0} = \frac{\sum_{i \in I} \tilde{Y}_i D_i}{N^+_{W_0}} - \frac{\sum_{i \in I} \tilde{Y}_i (1 - D_i)}{N^-_{W_0}}, \tag{3}$$

where $I = \{i: R_i \epsilon W_0\}$ and $N^+_{W_0}$, $N^-_{W_0}$ are the number of units in the treatment and control group, respectively. The window $W_0$ is selected around the cutoff as an interval where covariates are balanced between treated and control units (Cattaneo *et al.*, 2015). Practically, sharp null hypothesis tests are run for all covariates within different windows and $W_0$ is one of the largest where the minimum p-value across all tests and across smaller windows is above a certain cutoff.

## 3.3 Critical issues

In conclusion, the main limitation of the RDD framework concerns the external validity of results. In fact, RDD allows verifying whether there is an impact of the legislation only for the class of municipalities near the cutoff. Preliminary descriptive analyses of our dataset show significant differences in gender balance between municipalities of different size (CNR-IRPPS, 2018) and previous research (Ragazzi *et al.*, 2018) suggested strong and robust evidence of a positive impact of gender mechanisms in voting regulation based on a combination of gender quota and double gender preference. In particular, preliminary results showed a greater impact in areas characterized by greater unbalance, thus suggesting a process of territorial convergence. Nevertheless, local-specific features still maintain a strong effect.

In this paper, we use different approaches to RDD to investigate the robustness and sensitivity of results to different assumptions and estimation strategies. Moreover, controlling for determinants of women participation in electoral competition (Benati, Falavigna and Sella, 2018), estimates are more precise.

# 4 Data

Our data cover several information on Italian municipalities in the time span 2009-2016, that includes the electoral reform. In particular, pre-2013 data are mainly exploited in falsification testing, while point

[6] Large-sample approximations of the relevant test statistic can be applied, too. In this case, potential outcomes are interpreted as random variables.

estimates and inference are performed on a pooled cross-section in the span 2013-2016. The full dataset was obtained by matching information from different sources: electoral data on candidates were provided by the Italian Ministry of Interior[7]; electoral data on municipal councils are publicly available[8]; contextual data on municipalities are released by Istat, the Italian Statistical Office.

In details, the outcome variable is represented by the share of women elected in municipal councils, while the share of candidate women is an output, which is directly prescribed by the electoral norm. The running variable is the resident population, which is expressed in logarithm: the first cutoff is 8.517 (5,000 inhabitants), while the second is 9.616 (15,000 inhabitants).

**Table 1 – Descriptive statistics for running, output and outcome variables**

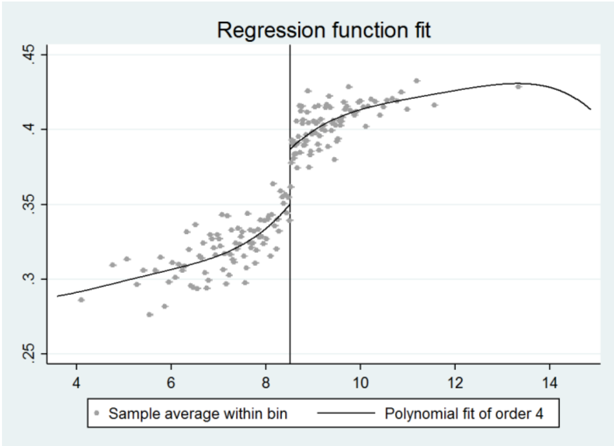
| | Running variable | | Output variable | Outcome variable |
|---|---|---|---|---|
| | Population | Ln(pop) | Share of female candidates | Share of female council members |
| Mean | 7.547,861 | 7,803 | 0,346 | 0,304 |
| Standard dev. | 53.325,777 | 1,347 | 0,101 | 0,131 |
| Min | 36 | 3,584 | 0,000 | 0,000 |
| 10th perc. | 457 | 6,125 | 0,200 | 0,143 |
| 25th perc. | 998 | 6,906 | 0,292 | 0,231 |
| 50th perc. | 2.353 | 7,763 | 0,361 | 0,308 |
| 75th perc. | 5.809 | 8,667 | 0,417 | 0,387 |
| 90th perc. | 13.585 | 9,517 | 0,458 | 0,462 |
| Max | 2.864.731 | 14,868 | 0,778 | 1,000 |
| Obs | 7.179 | 7.179 | 6.393 | 7.323 |

Table 1 reports descriptive statistics, while Figure 1 plots raw data (panels a and c) and RD plots over the support of the running variable, returning a graphical representation of our problem. In particular, RD plots are binned means of female candidates (panel b) and female council members (panel d) with quantile-spaced bins that are chosen optimally to mimic data variability (Calonico *et al.*, 2015). The smoothed RD plots show a clear upward jump of polynomial fits right at the population cutoff, i.e. the shares of both female candidates and elected are higher immediately to the right of the cutoff in municipalities that apply gender mechanisms in voting. This is a first heuristic evidence of the treatment effect we are estimating.

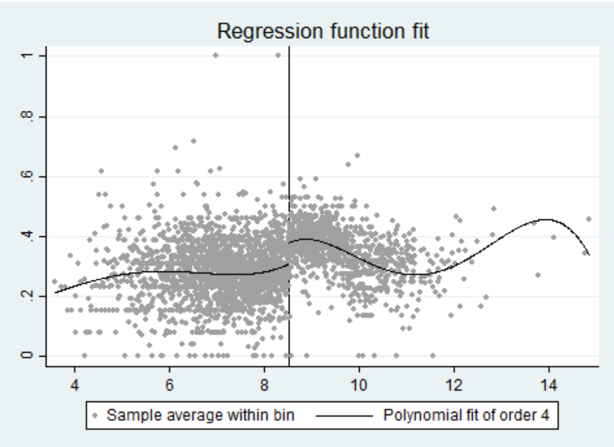
[7] Data were provided in the context of the SVEGLIE evaluation project, see section 2.
[8] https://dait.interno.gov.it/elezioni/anagrafe-amministratori

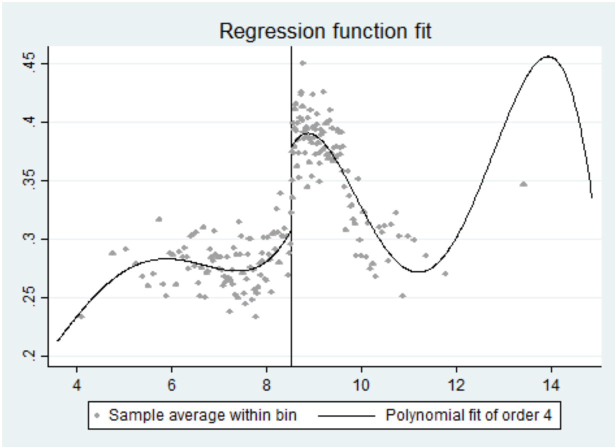
**Figure 1 – Scatter and RD plot of output and outcome variables.**

Female candidates (a) scatter plot, $N^{-}$=4.480, $N^{+}$=1.822 (b) RD plot, $J^{-}$=86, $J^{+}$=88


Female elected (a) scatter plot, $N^{-}$=5.127, $N^{+}$=2.052 (d) Elected, RD plot, $J^{-}$=93, $J^{+}$=107




*Notes:* Vertical lines represent the first cutoff. In panels (a) and (c) $N^{-}$ and $N^{+}$ denote the sample size for control and treatment groups, respectively; dots represent raw data. In panels (b) and (d) bins are quantile spaced and their total number $J^{-}$ and $J^{+}$ is chosen to mimic variance; dots represent the sample average of the output/outcome variable within each bin. In all panels solid lines show $4^{th}$ order polynomial fits using control and treated units separately.

Finally, some covariates are chosen to describe socio-demographic and economic characteristics of municipalities either in falsification testing or in nonparametric polynomial regression. In particular, we use a mix of time-variant and census variables that significantly explain women's participation in political activity (Benati, Falavigna and Sella, 2018), but are independent from the norm. Table 2 shows the descriptive statistics.

**Table 2 – Descriptive statistics for covariates**

| | Female share in population | Average family dimension | Female activity rate, 2011 | Per capita income, 2011 | Female degree rate, 2011 | Male degree rate, 2011 | Migrants share, 2010 |
|---|---|---|---|---|---|---|---|
| Mean | 0,506 | 2,312 | 40,749 | 20,496 | 7,946 | 6,494 | 0,065 |
| Standard dev. | 0,017 | 0,266 | 7,557 | 3,023 | 4,569 | 4,378 | 0,045 |
| Min | 0,304 | 1,210 | 10,450 | 11,998 | 0,000 | 0,000 | 0,000 |
| 10th perc. | 0,489 | 1,980 | 30,440 | 16,894 | 4,651 | 3,383 | 0,014 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 25th perc. | 0,499 | 2,160 | 35,335 | 18,464 | 5,907 | 4,537 | 0,028 |
| 50th perc. | 0,508 | 2,330 | 41,490 | 20,295 | 7,394 | 5,889 | 0,056 |
| 75th perc. | 0,515 | 2,480 | 46,190 | 22,198 | 9,163 | 7,527 | 0,091 |
| 90th perc. | 0,523 | 2,620 | 49,730 | 24,108 | 11,100 | 9,552 | 0,125 |
| Max | 0,624 | 3,460 | 71,720 | 53,589 | 85,493 | 95,948 | 0,340 |
| Obs | 7.179 | 7.179 | 7.124 | 7.124 | 7.117 | 7.118 | 7.120 |

# 5 Results

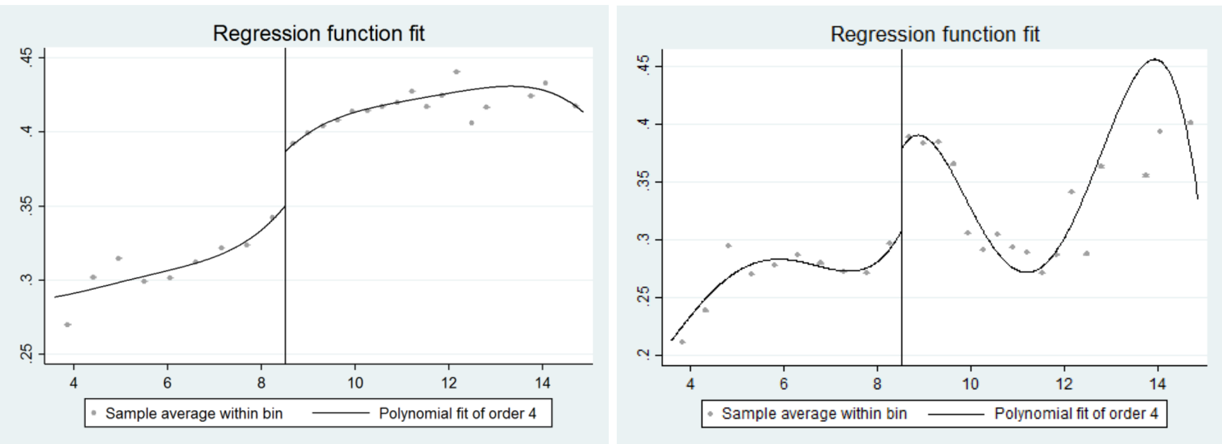
This section presents the results of estimation and inference in both local polynomial and local randomization RDD frameworks. In particular, preliminary falsification tests check the validity of the RDD approach. Then, a set of local polynomial techniques are applied, both parametric and non-parametric.…

## 5.1 Falsification tests

Falsification tests are performed following the three most common approaches described in the methodology. Figure 2 proposes a global data-driven graphical inspection of the continuity away from the cutoff over the support of the running variable of both the output (panel a) and outcome (panel b) variables. RD plots are constructed using the Integrated Mean Square Error (IMSE) optimal number of disjoint evenly-spaced bins to approximate the underlying regression functions under repeated sampling (Calonico *et al.*, 2015). By this approach, we compare a global polynomial fit to a local sample-means fit over disjoint bins in order to detect possible discontinuities away from the cutoff. The plot does not suggest any additional discontinuity, but it shows some difference between the polynomial and the sample-mean fits in the largest cities, which could derive from the second threshold.

**Figure 2 – RD plots using IMSE-optimal evenly-spaced approximation of outcome and output variables**

(a) Female candidates, $J^-$=9, $J^+$=20 (b) Female council members, $J^-$=10, $J^+$=20




*Notes:* Vertical lines represent the first cutoff; solid lines show 4th order polynomial fits using control and treated units separately; dots represent the sample average of the output/outcome variable within each bin.

The second falsification test focuses on the distribution of the running variable for observations near the cutoff *r*. The null hypothesis of random assignment is rejected for a bandwidth $h > 0.225$, i.e. we can significantly hypothesise random assignment in municipalities with population in the range [3933;6262].

**Table 3 – Binomial falsification test based on the running variable (population).**

| h | Ln(pop) | Population | $N^-_w$ | $N^+_w$ | p-value |
|---|---|---|---|---|---|
| 0,006 | [8,511;8,523] | [4970;5030] | 10 | 15 | 0,424 |
| 0,028 | [8,489;8,545] | [4862;5142] | 53 | 40 | 0,213 |
| 0,046 | [8,471;8,563] | [4775;5235] | 88 | 65 | 0,075 |
| 0,061 | [8,456;8,578] | [4704;5315] | 113 | 94 | 0,211 |
| 0,077 | [8,44;8,594] | [4629;5400] | 144 | 118 | 0,122 |
| 0,096 | [8,421;8,613] | [4542;5504] | 172 | 144 | 0,129 |
| 0,111 | [8,406;8,628] | [4475;5587] | 199 | 169 | 0,130 |
| 0,125 | [8,392;8,642] | [4413;5666] | 224 | 195 | 0,171 |
| 0,145 | [8,372;8,662] | [4325;5780] | 249 | 221 | 0,213 |
| 0,160 | [8,357;8,677] | [4261;5868] | 273 | 248 | 0,293 |
| 0,184 | [8,333;8,701] | [4160;6010] | 319 | 273 | 0,064 |
| 0,203 | [8,314;8,72] | [4081;6125] | 347 | 298 | 0,059 |
| 0,225 | [8,292;8,742] | [3993;6262] | 374 | 323 | 0,058 |
| 0,246 | [8,271;8,763] | [3910;6395] | 412 | 348 | 0,022 |

*Notes:* Cutoff is $r = 8.517$ and window $W = [r - h; r + h]$. Binomial test p-values are computed using exact binomial distribution with probability $q$=1/2. $N^-_w$ and $N^+_w$ denote the sample size for control and treatment groups.

Finally, the third falsification approach is based on placebo tests and depends on the method we use to estimate the ATE. Hence, placebo test results are proposed contextually to RDD estimates.

## 5.2 Parametric local polynomial approach

This approach estimates by OLS regression the function in equation 2. In particular, $\tau_S$ is the parameter of interest, which is estimated by linear ($p$ = 1) and quartic ($p$ = 4) models within three heuristic windows ($W$ = {0.2108; 0.4464; 1.0218}) around the cutoff and on a global scale, which is not recommended but it represents a useful comparison. Table 4 shows significant positive impacts of the legislation on both the output and outcome variables across different models. In particular, the effect on the share of female candidates is about 3.5% and quite similar across models and windows, while the effect on elected women is quite heterogeneous in the range 3.8%-7.7%.

**Table 4 – Flexible parametric RD methods applied to output, outcome and placebo variables.**

| | Linear model (p=1) | | Quartic model (p=4) | |
|---|---|---|---|---|
| Order polynomial | 1 | 1 | 4 | 4 |
| h | 0,1054 | 0,2232 | 0,5109 | Global scale |
| Population range | [4500;5556] | [4000;6250] | [3000;8334] | [33;2.864.731] |
| ***Share of female candidates*** | | | | |
| **RD treatment effect** | **0,032**** | **0,035***** | **0,040**** | **0,036***** |
| Parametric 95% CI | [0,002;0,061] | [0,014;0,056] | [0,006;0,073] | [0,024;0,049] |
| Parametric p-value | 0,036 | 0,001 | 0,021 | 0,000 |
| $N^-_w$; $N^+_w$ | 187; 157 | 372; 318 | 888; 694 | 4480; 1822 |
| ***Share of female council members*** | | | | |
| **RD treatment effect** | **0,038*** | **0,077***** | **0,036** | **0,070***** |
| Parametric 95% CI | [-0,004;0,081] | [0,045;0,108] | [-0,011;0,084] | [0,052;0,088] |
| Parametric p-value | 0,077 | 0,000 | 0,129 | 0,000 |

| $N^-_w$; $N^+_w$ | 211; 183 | 419; 356 | 1012; 782 | 5127; 2052 |
|---|---|---|---|---|
| ***Falsification tests - parametric p-value*** | | | | |
| Female share, 2011 | 0,927 | 0,252 | 0,811 | 0,030 |
| Average family dimension, 2011 | 0,067* | 0,045** | 0,053* | 0,036** |
| Female activity rate, 2011 | 0,201 | 0,396 | 0,281 | 0,053 |
| Per capita income, 2011 | 0,344 | 0,974 | 0,499 | 0,196 |
| Female degree rate, 2011 | 0,235 | 0,068* | 0,128 | 0,062* |
| Male degree rate, 2011 | 0,354 | 0,020** | 0,301 | 0,069* |
| Migrants share, 2011 | 0,892 | 0,029** | 0,738 | 0,002*** |

*Notes:* Cutoff is $r$ = 8.517 and window $W = [r - h; r + h]$; $N^-_w$ and $N^+_w$ denote the sample size for control and treatment groups. OLS estimations are performed with cluster-robust standard errors.
*** p<0,01; ** p<0,05; * p<0,1.

Apart from the global approach, the window producing the most significant estimates considers municipalities with population between 4000 and 6250 individuals. However, the falsification tests in the bottom panel show some criticalities. They report p-values from RD estimation for a set of covariates that should not be affected by the norm, i.e. their ATE should be indistinguishable from zero. However, four covariates over seven show statistically significant effects, casting some doubts on the ATE estimates within the bandwidth $h$ = 0.2232.

## 5.3 Nonparametric local polynomial approach

This approach uses data-driven techniques to select the proper bandwidth (Calonico *et al.*, 2017). Results in Table 5 are based on a local constant ($p$ = 0) and a local linear ($p$ = 1) polynomial approximations. In fact, local linear approximations are usually applied, but local constant approximation allow better comparison with local randomization results. Both the MSE-optimal and an heuristic bandwidth are applied.

**Table 5 – Robust bias-corrected local polynomial methods applied to output, outcome and placebo variables.**

| | Constant model (p=0) | | Linear model (p=1) | |
|---|---|---|---|---|
| | $h = \hat{h}_{MSE}$ | $h = \hat{h}_{FP}$ | $h = \hat{h}_{MSE}$ | $h = \hat{h}_{FP}$ |
| Order polynomial | 0 | 0 | 1 | 1 |
| h | 0,238 | 0,160 | 0,855 \| 0,697[†] | 0,160 |
| Population range | [3941;6344] | [4261;5868] | [2126;11757] | [4261;5868] |
| ***Share of female candidates*** | | | | |
| **RD treatment effect** | **0,037*** ** | **0,036** ** | **0,035*** ** | **0,033** |
| Robust 95% CI | [0,020;0,046] | [0,006;0,059] | [0,019;0,046] | [-0,016;0,059] |
| Robust p-value | 0,000 | 0,015 | 0,000 | 0,261 |
| $N^-_w$; $N^+_w$ | 400; 337 | 274; 248 | 1536; 1058 | 274; 248 |
| ***Share of female council members*** | | | | |
| **RD treatment effect** | **0,081*** ** | **0,079*** ** | **0,077*** ** | **0,051** |
| Robust 95% CI | [0,054;0,099] | [0,013;0,089] | [0,051;0,098] | [-0,035;0,067] |
| Robust p-value | 0,000 | 0,009 | 0,000 | 0,546 |
| $N^-_w$; $N^+_w$ | 449; 376 | 306; 278 | 1419; 1022 | 306; 278 |
| ***Falsification tests - robust p-value*** | | | | |
| Female share, 2011 | 0,119 | 0,981 | 0,165 | 0,923 |
| Average family dimension, 2011 | 0,112 | 0,049** | 0,057* | 0,320 |
| Female activity rate, 2011 | 0,354 | 0,315 | 0,263 | 0,176 |
| Per capita income, 2011 | 0,922 | 0,385 | 0,733 | 0,441 |
| Female degree rate, 2011 | 0,189 | 0,133 | 0,131 | 0,444 |
| Male degree rate, 2011 | 0,070* | 0,161 | 0,051* | 0,985 |

| Migrants share, 2011 | 0,036** | 0,592 | 0,035** | 0,622 |
|---|---|---|---|---|

*Notes:* Cutoff is $r$ = 8.517 and window $W = [r - h; r + h]$; $N^-_w$ and $N^+_w$ denote the sample size for control and treatment groups. Point RD estimators are constructed using triangular kernel; robust p-values are constructed using bias-correction with robust standard errors; $\hat{h}_{MSE}$ is the MSE-optimal bandwidth selector; $\hat{h}_{FP}$ is chosen *ad hoc*.
† This value is MSE-optimal bandwidth for the outcome variable share of female council members.
*** p<0,01; ** p<0,05; * p<0,1.

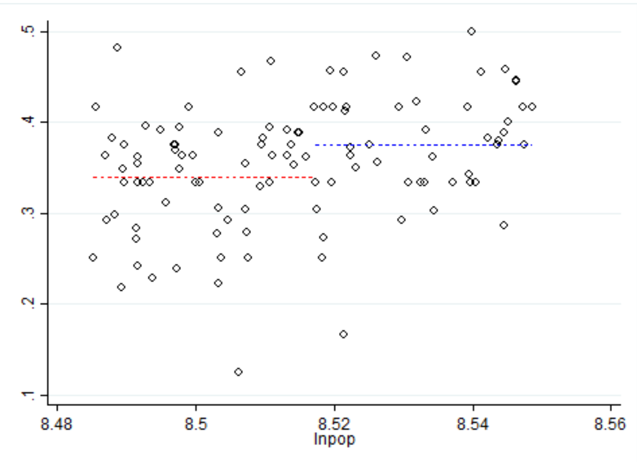
It is evident that the nonparametric framework reports much more homogeneous results. In particular, significant point estimates of the effect on the share of female candidates are in the range 3.5%-3.7%, while that on female council members is estimated between 7.7% and 8.1%. Falsification tests do not generally detect treatment effects in the covariates, suggesting a better performance of the data-driven bandwidth selection.

## 5.4 Local randomization approach

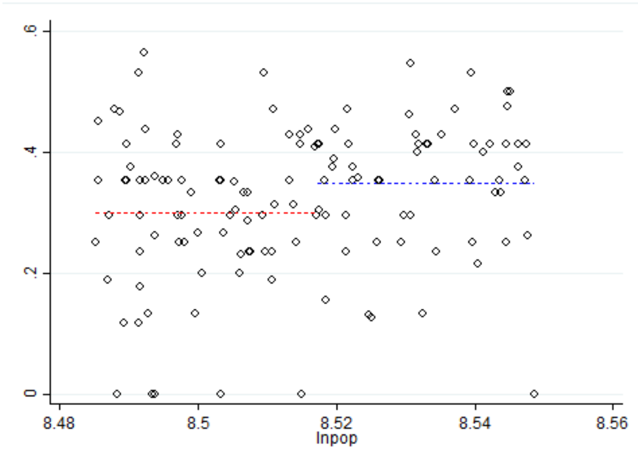
This approach estimates the ATE in equation 3. The automatic procedure for window selection (Cattaneo *et al.*, 2016) based on tests of the sharp null hypothesis of no effect on covariates selects a bandwidth $h_0$ = 0,032. As expected, it is notably narrower than the MSE-optimal one selected by nonparametric local polynomial techniques. Using $W_0$, Figure 3 plots the difference in average levels of the share of candidates (panel a) and elected women (panel b) between treated and control groups as the distance at the cutoff between the blue and red dotted lines. The sample size is quite similar in the two groups[9] and the plot does not suggest any relationship between the running and the output/outcome variables.

**Figure 3 – Scatterplot of the output and outcome variables against the running variable in the local randomization window.**

(a) Female candidates, $N^-$=62, $N^+$=47




(b) Female council members, $N^-$=72, $N^+$=57




*Notes:* Cutoff is $r$ = 8.517 and window $W_0 = [r - h_0; r + h_0]$ = 0,064; $N^-_w$ and $N^+_w$ denote the sample size for control and treatment groups, respectively. Dotted lines represent the average output/outcome within the control (red) and treatment (blue) groups. The window is selected using as placebo covariates the variables in table 2 and the Kolmogorov-Smirnov statistic.

The ATE point estimates on both the share of candidate and elected women are described in Table 6. It reports results for both untranformed ($p$ = 0) and transformed ($p$ = 1) variables according to three different bandwidths: the local randomization $h_0$, the MSE-optimal $h_{MSE}$, and the heuristic $h_{FP}$.

[9] A binomial test does not reject the null hypothesis that the probability of treatment is 0,5.

Table 6 – Local randomization methods applied to output, outcome and placebo variables.

| | No transformation (p=0) | | | Linear transformation (p=1) | | |
|---|---|---|---|---|---|---|
| | $h = \hat{h}_0$ | $h = \hat{h}_{MSE}$ | $h = \hat{h}_{FP}$ | $h = \hat{h}_0$ | $h = \hat{h}_{MSE}$ | $h = \hat{h}_{FP}$ |
| Order polynomial | 0 | 0 | 0 | 1 | 1 | 1 |
| h | 0,032 | 0,238 | 0,16 | 0,032 | 0,855 \| 0,697† | 0,16 |
| Population range | [4843;5163] | [3941;6344] | [4261;5868] | [4843;5163] | [2126;11757] | [4261;5868] |
| ***Share of female candidates*** | | | | | | |
| **RD treatment effect** | **0,037***** | **0,041***** | **0,037***** | **-0,009** | **0,036***** | **0,033***** |
| Fisher's p-value | 0,005 | 0,000 | 0,000 | 0,475 | 0,000 | 0,000 |
| $N^-_w$; $N^+_w$ | 62; 47 | 399; 377 | 274; 278 | 62; 47 | 1538; 1058 | 274; 248 |
| ***Share of female council members*** | | | | | | |
| ***RD treatment effect*** | **0,049**** | **0,083***** | **0,084***** | **0,026** | **0,077***** | **0,071***** |
| Fisher's p-value | 0,021 | 0,000 | 0,000 | 0,223 | 0,000 | 0,000 |
| $N^-_w$; $N^+_w$ | 72; 57 | 449; 376 | 306; 278 | 72; 57 | 1417; 1022 | 306; 278 |
| ***Falsification tests - Fisher's p-value*** | | | | | | |
| Female share, 2011 | 0,852 | 0,045** | 0,008*** | 0,484 | 0,008*** | 0,833 |
| Average family dim., 2011 | 0,177 | 0,011** | 0,018** | 0,069* | 0,000*** | 0,000*** |
| Female activity rate, 2011 | 0,429 | 0,833 | 0,131 | 0,367 | 0,000*** | 0,228 |
| Per capita income, 2011 | 0,360 | 0,332 | 0,990 | 0,916 | 0,187 | 0,039** |
| Female degree rate, 2011 | 0,360 | 0,094 | 0,081* | 0,109 | 0,546 | 0,000*** |
| Male degree rate, 2011 | 0,325 | 0,027** | 0,012*** | 0,697 | 0,306 | 0,000*** |
| Migrants share, 2011 | 0,697 | 0,922 | 0,079* | 0,693 | 0,000*** | 0,003*** |

*Notes:* Cutoff is $r$ = 8.517 and window $W = [r - h; r + h]$; $N^-_w$ and $N^+_w$ denote the sample size for control and treatment groups. RD point estimators are constructed using difference-in-means statistic; randomization p-values are calculated using 1000 permutations; $\hat{h}_0$ is derived as proposed in Cattaneo *et al.* (2015); $\hat{h}_{MSE}$ is the MSE-optimal bandwidth selector; $\hat{h}_{FP}$ is chosen *ad hoc*.
† This value is MSE-optimal bandwidth for the outcome variable share of female council members.
*** p<0,01; ** p<0,05; * p<0,1.

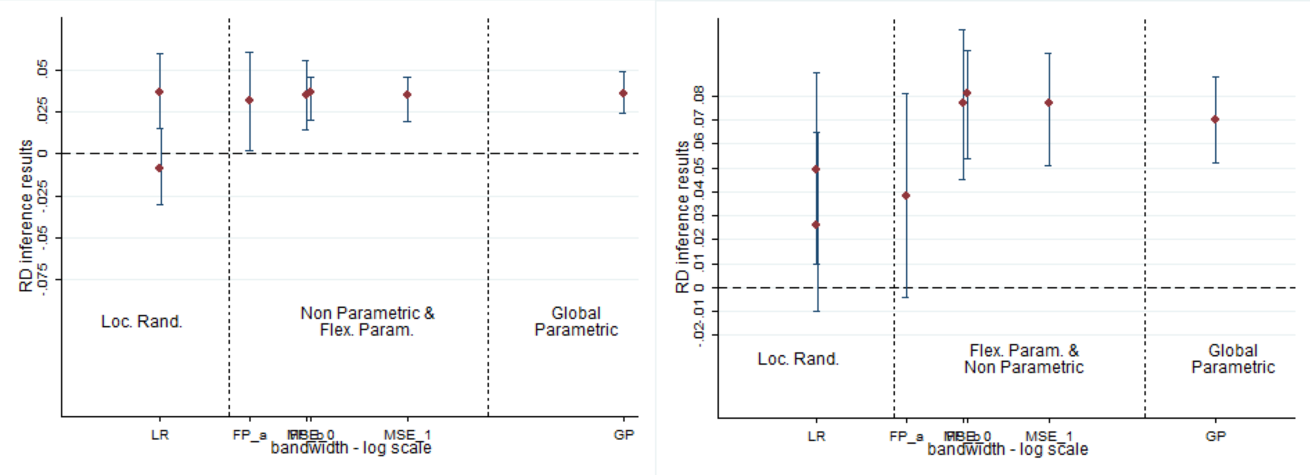
Looking at falsification tests, covariates satisfy the sharp null of no treatment effect only when $h_0$ is applied, i.e. estimation satisfies the local randomization assumptions, which is implausible in larger windows. In particular, the ATE on the candidates is 3.7%, resembling previous results. Moreover, local randomization estimates within the other windows are significant and quite similar, but less reliable from a statistical point of view because of falsification tests. On the contrary, estimates on the share of female council members are quite heterogeneous, i.e. in the range 4.9%-8.4%.

# 6 Conclusions and further research

Figure 4 summarises the results obtained with the different approaches. Red points are impact estimates; they are placed at the centre of their confidence intervals.

**Figure 4 – RD point estimates and confidence intervals, all methods.**

(a) Female candidates (b) Female council members




*Notes:* Cutoff is $r$ = 8.517; LR corresponds to point estimates and confidence intervals using $\hat{h}_0 = 0{,}032$ and no transformation ($p$ = 0); FP_a and FP_b correspond to flexible parametric estimates using *ad hoc* $\hat{h}_{FP\_a} = 0{,}1054$ and $\hat{h}_{FP\_b} = 0{,}2232$, respectively; MSE_0 and MSE_1 correspond to nonparametric specification under constant and linear model using $\hat{h}_{MSE\_0} = 0{,}032$ and $\hat{h}_{MSE\_1} = 0{,}032$; GP is the global polynomial estimation of order four using the full sample.

RDD evaluation confirms results of preliminary descriptive statistics: pairing gender quota and double preference is an effective tool to enhance the presence of women in elective assemblies. This result is very robust since the estimated impact is significant across different approaches. The impact is higher in the case of elected representatives than in the case of candidates, showing that the mechanism centres the policy aim.

The size of the impact varies among the different approaches. For example, the impact is lower in local randomisation models. The interpretation of these discrepancies will be the object of future research work. Future extension of the work will also concern:

- Sensitivity analysis (local randomization)
- Considering the second threshold (15,000 residents)
- Modeling the two thresholds together

# References

Baum M & Espírito-Santo A. (2012) Portugal's Quota-Parity Law: An Analysis of its Adoption, West European Politics, 35 (2), 319-342, DOI:10.1080/01402382.2011.648009

Brennan M. and Buckley f. (2017), The Irish legislative gender quota: The first election, Administration, (2), 15-35 doi: 10.1515/admin-2017-0013

Benati I., Falavigna G., Sella L., Gender balance in Italian electoral system: some empirical reflections on determinants, XXXIX Conferenza annuale AISRe 2018, 17-19 settembre 2018, Bolzano (IT)

Calonico, S., Cattaneo, M. D., Titiunik, R. (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295-2326

Calonico, S., Cattaneo, M. D., Titiunik, R. (2015) Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110, 1753–1769

Calonico, S., Cattaneo, M. D., Farrell, M. H., Titiunik, R. (2017) rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2), 372-404

Cattaneo, M. D., Frandsen, B., Titiunik, R. (2015) Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. Senate. *Journal of Causal Inference*, 3, 1–24

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G. (2016). Inference in Regression Discontinuity Designs under Local Randomization. *The Stata Journal*, 16(2), 331-367

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G. (2017) Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality. *Journal of Policy Analysis and Management*, 36(3), 643-681

Cattaneo, M. D., Jansson, M., Ma, X. (2018) Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1), 234-261

CNR-IRPPS (2018) *Rapporto finale dell'attività di ricerca e valutazione dell'impatto delle misure di riequilibrio di genere previste dalla legislazione vigente in materia di cariche elettive e di governo*, Presidenza del Consiglio dei Ministri: Dipartimento per le Riforme Istituzionali, https://www.irpps.cnr.it/poges/sveglie-studio-e-valutazione-dellimpatto-delle-misure-di-riequilibrio-di-genere-previste-dalla-legislazione-vigente-in-materia-di-cariche-elettive-e-di-governo/

De Paola, M., Scoppa, V., Lombardo, R. (2010) Can gender quotas break down negative stereotypes? Evidence from changes in electoral rules. *Journal of Public Economics,* 94 (5), 344-353.

De Paola M., Scoppa V., De Benedetto M.A. (2014), The impact of gender quotas on electoral participation: Evidence from Italian municipalities, European Journal of Political Economy, 35, 141-157

European Commission (2017) 2017 report on equality between women and men, Brussels

Fisher, R. A. (1935). *Design of experiments.* New York, NY: Hafner

Gerber, A. S., Green, D. P. (2008) *Field experiments and natural experiments.* In The Oxford Handbook of Political Science. Oxford University Press. Retrieved 31 Jul. 2019, from https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199604456.001.0001/oxfordhb-9780199604456-e-050.

Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69, 201–209

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79, 933–959

Imbens, G. W., Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615–635

Inter-Parliamentary Union (2017) *Women in Politics 2017*, Geneva

Jones M.P. (1998), Gender Quotas, Electoral Laws, and the Election of Women: Lessons from the Argentine Provinces, Comparative Political Studies 31 (1), 3-21

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, 142, 675–697

Lee, D. S., Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48, 281–355.

MacRae H. (2012) Double-Speak: The European Union and Gender Parity, West European Politics, 35 (2), 301-318, DOI: 10.1080/01402382.2011.648008

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142, 698–714

Meier P. (2012) From Laggard to Leader: Explaining the Belgian Gender Quotas and Parity Clause, West European Politics, 35(2), 362-379, DOI:10.1080/01402382.2011.648011

Murray R. (2012) Parity in France: A ‘Dual Track’ Solution to Women's Under-Representation, West European Politics, 35 (2), 343-361, DOI:10.1080/01402382.2011.648010

Praud J. (2012) Introduction: Gender Parity and Quotas in European Politics, West European Politics, 35 (2), 286-300, DOI: 10.1080/01402382.2011.648007

Palici di Suni E. (2012) Gender Parity and Quotas in Italy: A Convoluted Reform Process, West European Politics, 35 (2), 380-394, DOI: 10.1080/01402382.2011.648013

Ragazzi E., Finardi U., Manello A., Sella L., Measures for gender rebalancing in local elective assemblies. An impact assessment on Italian reforms, 58th congress of the European Regional Science Association (ERSA), 28-31 agosto 2018, Cork (IR)

Rosenbaum, P. R. (2010). *Design of observational studies.* New York, NY: Springer

Sartori, L., Tuorto, D., Ghigi, R. (2017) The Social Roots of the Gender Gap in Political Participation: The Role of Situational and Cultural Constraints in Italy. *Social Politics: International Studies in Gender, State & Society, 24* (3), 221-247.

Verge T. (2012) Institutionalising Gender Equality in Spain: From Party Quotas to Electoral Gender Quotas, West European Politics, 35 (2), 395-414, DOI: 10.1080/01402382.2011.648014

Yoon J. & Shin K. (2017) Opportunities and Challenges to Gender Quotas in Local Politics: The Case of Municipal Council Elections in South Korea, Asian Journal of Women's Studies, 23 (3), 363-384, DOI: 10.1080/12259276.2017.1352131