Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 34

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

ESSAYS IN POLITICAL AND GENDER ECONOMICS

Presentata da: Giovanni Righetto

Coordinatore Dottorato

Andrea Mattozzi

Supervisore

Paolo Vanin

Esame finale anno 2023

Political power and the influence of minorities: theory and evidence from Italy

Giovanni Righetto[1]

April 20, 2023

Abstract

In this paper, we analyze the relationship between minority and majority in politics, and how it can influence policy outcomes and potential conflict between parties. In particular, we focus on the consequences of a sudden increase in the political power of a minority (e.g. female politicians after a gender quota), and its potential effects on the relationship with the long-standing majority. We first present a theoretical model describing the possible consequences of such an increase in a minority's political power and show how it can increase difficulties in reaching a compromise on policy outcomes between parties. In the case of a high increase in minority's power, its demands in terms of policy outcomes increase and make the compromise costlier for the majority, which might prefer to engage in conflict. Furthermore, we empirically test these implications by exploiting the introduction in 2012 of a gender quota in Italian local elections. By the means of a Difference-in-Discontinuity strategy, we show how the generated increase in female politicians had heterogeneous effects on the level of funding for daycare, based on its differential effects on the share of post-quota women councillors. For high shares of female councillors, a decrease in expenditure for day care was observed with respect to control municipalities, while in municipalities with low shares the quota was followed by a relative increase in funding for day care.
JEL classification numbers: D74, H53, I38
Keywords: Conflict, committee, gender quotas, day care investments

---

[1] PhD Student in Economics, University of Bologna, E-mail: giovanni.righetto3@unibo.it.

I would like to thank Paolo Vanin, my PhD supervisor. I am also grateful to participants in all conferences and seminars I attended for all the comments and constructive feedback. In addition, I acknowledge the financial support of the University of Bologna.

1 Introduction

During the last two centuries, the democratization process in many countries has helped the enfranchisement of social groups which were still unable to influence the political process of decision-making. We can think for instance of the extension of suffrage to women or Afro-Americans or the recent introduction of gender quotas in politics both in developed and developing countries. The enfranchisement process has been demonstrated to benefit the under-represented social groups in the long run from many different perspectives such as employment (Aneja and Avenancio-Leon (2019)), public investments (Miller (2008)), and political responsiveness to these groups' needs (Fujiwara (2015)). On the other hand, after a minority receives larger political power there might be some other mechanisms triggered in the short run, depending on how the long-standing majority reacts to this sudden decrease in its capacity to influence the political decision-making process. For instance, while a majority group might accept to "give in" to minority's instances when the balances of power are disproportionate, it might feel threatened when the minority gains more power and demands more favourable treatment in terms of policy decisions. Indeed, growth in the representation of minorities in the political class could encourage them to increase their political demands, thus making it more difficult from the point of view of the long-standing majority to accept a compromise. In this paper, we first try to model this mechanism by sketching the preferences and behaviours of two groups of politicians, a majority and a minority group, with possible changes in the balances of political power: think for instance of female politicians after the introduction of a gender quota. The two groups of politicians have to decide on the level of a policy: in the sequential game, the majority decides on a level, and then the minority chooses whether to engage in political conflict or not. The chances of either conflict or compromise are influenced by the relative groups' shares: when the outgroup share grows, also its respective demands in terms of the implemented policy level move closer to its bliss point. In this setting, a shock in the minority's political power increases the probability that the majority decides not to compromise on policy decisions, since minority claims are too expensive. In the case of no compromise, the majority decides to vote for its bliss point on the policy level, with the minority that might end up being worse off after the increase in representation. In addition, the model shows how the salience of the policy (i.e. to which extent the public debate focuses on it) is positively related to the probability of conflict, which is on the other hand inversely proportional to the intensity of the future interactions between the two groups. The intuition is that politicians find it more profitable to conflict over issues that polarize voters, since it could be profitable in terms of electoral gains. On the other hand, the two groups prefer to avoid friction when they know that they will have to bargain frequently over other policies in the future. We test these predictions in our empirical setting by exploiting the introduction in 2012 of a new law on Italian local elections, which increased the presence of female local politicians in a subsample of small municipalities. By the means of a Difference in Discontinuity strategy, we look at the effect of this sudden increase in female politicians' representation levels and subsequently on day care-related expenditures, possibly the most "gender-sensitive" among all kinds of policies. We find an extremely interesting result: in line with our argument, the effect on day care expenditures depended on the post-law share of female councillors in treated municipalities' councils. While in municipalities with low shares of female councillors we observe a positive change in funding for day care by 9.5% with respect to the control group, in cities with large increases in female politicians, an 18% reduction in this kind of expenditure followed the introduction of the law, always in relation to control municipalities. Thus, this increase in a minority's political representation had a heterogeneous effect depending on whether the following increase in the minority group's share was such to prevent compromise with the majority. We deepen the analysis of this effect by describing differences between cities with high and low shares of post-law female councillors, to understand whether there might have been other structural elements driving our results: however, empirical evidence shows how our heterogeneous effect indeed depended on the law-induced increase in female politicians. One of our major robustness consists in assessing the heterogeneous new law's effect on other expenditure categories: in line with our reasoning, we observe a significant heterogeneous effect only on day care and not on non-gender-related categories such as transportation or tourism. This last piece of evidence is in line with our model in two different ways: first, because day care is a service for which men and women have distinct preferences (Barigozzi et al. (2019), and secondly because gender quotas can increase salience on gender topics[2], thus raising the probability of conflict. In addition, to support our empirical results, we conduct other robustness checks: first, a placebo analysis with artificial timing/threshold of the quotas' introduction to show that we are not capturing spurious correlations. Moreover, to reinforce our argument that the increase in women politicians was responsible for this heterogeneous effect, we adopt another identification strategy to demonstrate our main result. By using a 2SLS panel regression, we show how the exogenous increase in female councillors caused by the quota was followed by a decrease in day care-related expenses in the subsample of treated municipalities. Finally, we add some evidence to support our idea that the increase in the minority group size triggered political conflict. Indeed, in the municipalities with larger post-quota shares of female councillors, we also observe an increase in the probability of dissolution of the local council.

The paper is organized as follows: in Section 2 we describe the literature that we refer to, together with our contribution. After that, in Section 3 we present our theoretical model highlighting how group shares shape relationships and decision outcomes among politicians. In Section 4 we describe the new Law introduced in Italian local elections, which caused an increase in female politicians starting in 2012. In Section 5 we describe the data used and our empirical strategy, while in Section 6 we show our main results. Section 7 conducts a set of robustness checks together with additional evidence supporting our empirical findings. Finally, Section 8 concludes.

2 Literature contribution

According to the literature on increasing minorities' and politically under-represented social groups' rights, the existing evidence found a variety of positive long-term benefits, not only exclusive to these groups. For instance, women's suffrage in the US increased public

[2] According to experimental evidence (Coffman (2014)), women are more likely to expose their personal ideas in environments which are not male-dominated: an increase in the share of female councillors could have encouraged them to bring their political views on the discussion table. This is confirmed by also confirmed by local politicians in Bavaria (Baskaran and Hessami (2023): additional female councillors increased the amount of council meetings' time dedicated to gender topics.

health expenditures and helped the promotion of a large public hygiene campaign, leading to a reduction in child mortality (Miller (2008)). Moreover, the enfranchisement of the less educated electorate through the introduction of the electronic vote increased public health care spending in Brazil, facilitating in particular uneducated mothers' access to prenatal visits (Fujiwara (2015)). Another example is provided by the passing of the 1965 Voting Rights Act in the US which, by increasing the voting rights of racial minorities, reduced the wage gap between black and white workers by 5.5 pp between 1950 and 1980 (Aneja and Avenancio-Leon (2019)), mainly because of major employment of black people in the public sector. However, our main idea is that a positive shock in the rights of an under-represented group can have a short-term effect on group dynamics, leading to reduced margins of compromise between majority and minority. In this regard, our paper speaks to the strand of literature on conflicts and identity. The main paper we refer to is the work by Bonomi et al. (2021), which describes how social groups align according to their identities, generating a polarization of beliefs and conflicting views on policies' effects. An external shock, such as a migration wave or globalization, can change the dimension of conflict, for instance from an economic to a cultural conflict, leading people to align on two factions depending on their views. In this paper, we adapt their intuition to a simpler model with only one identification dimension, in which the reshaping of groups' shares can possibly trigger a political conflict. This intuition is also present in the recent work by Grossman and Helpman (2021), which shows how changes in the identification patterns of voters can alter their policy preferences. Furthermore, in this paper we also make reference to the studies on polarization by Esteban and Ray (Esteban and Ray (1994), Esteban and Ray (1999)), who described how polarization can be measured and how it might generate conflict in society. One of their major conclusions is that intra-group homogeneity and inter-group heterogeneity raise polarization levels in society. Here, our contribution is to offer a novel view of how altering the sizes of political groups can influence polarization and conflict, and ultimately even impair the investment in a publicly-funded good. Moving to our empirical part, when we focus on the potential effects of increasing the share of female politicians on the kind of policies implemented, the literature offers mixed conclusions. The theoretical predictions in these cases are not univocal: the median voter theory (Downs et al. (1957)) states that the personal characteristics of elected politicians do not matter for policy making, while the citizen-candidate model (Besley and Coate (1997)) reaches the conclusion that individual politicians' preferences matter for the kind of policies implemented. In some empirical works, the citizen-candidate model has been proven to be more plausible: for instance, there is evidence that in the US policies differ depending on whether a Republican or a Democrat wins the elections (Besley and Case (1995)). Consequently, if we think of gender quotas, we might suspect that such a sudden change in the political class' composition might have consequences in terms of policies.

Looking at the empirical literature closest to this paper, many works have exploited the exogenous imposition of gender quotas to understand whether changing the sex ratio of politicians in charge might have a causal effect on policy choices. An interesting setting in this sense is offered by the Indian mandate reservation system: since 1993, a minimum of one-third of seats plus the leadership position in randomly selected villages must be reserved for women. Policy and social consequences of this increase in female politicians have been wide, from a better representation of the policy preferences of the female electorate (Chattopadhyay and Duflo (2004)), with a larger expenditure on public goods, to a stronger reduction in the gender gap in school attendance and educational attainment (Beaman et al. (2012)). In addition, there is evidence that Indian female politicians invest more in education (Clots-Figueras (2012)) and public health infrastructures(Bhalotra and Clots-Figueras (2014)). Lastly, the generated increase in prenatal and childcare services was found to significantly reduce infant mortality (Bhalotra and Clots-Figueras (2014)). However, all these effects are not necessarily valid for other contexts: indeed, many other studies found no consequences of exogenous increases in female politicians. For instance, Bagues and Campa (2021) study the introduction of a gender quota in Spanish local elections and find no effects on either composition or size of public expenditures. Also in Norway, the exogenous increase of female politicians due to a gender quota did not alter how local administrators were using public funds (Geys and Sørensen (2019)). Other researchers exploit close mixed-gender races to study the effects of the election of a female mayor on policy choices. Also in this case, empirical results offered by the literature are mixed. For instance, it was found that in the US electing a female mayor did not affect the size and composition of public expenditures

(Ferreira and Gyourko (2014)). On the other hand, a female victory in the Bavarian local mixed-gender race was demonstrated to lead to a public child care expansion between 40% and 50% (Hessami and Baskaran (2019)). Thus, the empirical literature still did not find a consensus on whether an increase in female politicians can alter policy decisions. Hence, our contribution in this sense is to underline the importance to evaluate the post-quota interaction between male and female politicians and their respective group shares, since they can generate different policy outcomes. In other words, we demonstrate in this paper how looking at heterogeneous instead of overall effects can completely change the answer to this important question.

3 The model

We model a sequential game in which a committee has to decide the level of a policy. The committee is composed by politicians who belong to either one of two groups $G = M, F$: the groups' shares are (π_M, π_F), with $\pi_M + \pi_F = 1$ and with $\pi_F \leq 1/2$ being the share of the minority group F. The game is structured as follows: the majority M decides a level of the policy q to implement, with $q \in \mathbb{R}$, and then the minority decides whether to engage in political conflict or not (cases denoted C and N). We assume that decisions with respect to the policy level are taken through majority voting, thus the group M can impose the level that maximizes its utility. Conflict gives benefits in terms of electoral rewards, but it generates costs to both groups since it spoils the relationships with the respective outgroup's members.

The preferences over q are given by the quadratic loss function $-\frac{1}{2}(q - q_G)^2$, where q_G is group G's bliss point, and we assume without loss of generality that $q_F > q_M$. Conflict can benefit each group by an amount equal to $\frac{\phi}{2}(q - q_G)^2$, with $0 < \phi < 1$.[3] Substantially, conflicting over an issue can give more electoral benefits the more the decided policy level is distant from the group's bliss point. The parameter ϕ represents the salience of an issue, or,

[3] We need to impose this condition to prevent conflict benefits to completely overturn the utility given by the policy, which would happen in the case that $\phi > 1$.

in other words, the extent to which the public debate is centered on this particular policy. On the other hand, conflict entails also costs that are proportioned to the share of the other group and equal to $-\frac{\gamma}{2}(1-\pi_G)^2$, with $\gamma > 0$. Here, engaging in a conflict can spoil future interactions with the other group's members, and therefore creates more costs the larger the outgroup. The term γ captures the frequency of future interactions between the two groups of politicians: in other words, the longer this political class has to interact in the future, the larger will be γ. Therefore, the utility U_G of the group G depends on the policy level chosen and on the outcome of the negotiations. We have the following:

$$U_G(q, N) = -\tfrac{1}{2}(q - q_G)^2$$

and

$$U_G(q, C) = -\tfrac{1}{2}(q - q_G)^2 + \tfrac{\phi}{2}(q - q_G)^2 - \tfrac{\gamma}{2}(1 - \pi_G)^2$$

We solve the game by backward induction, starting from the best response $S_F^*(q)$ of F and then moving to the optimal level q^* of q set by M. F's optimal strategy is to choose conflict if $U_F(q, N) < U_F(q, C)$, that is if

$$\pi_F > 1 - \sqrt{\frac{\phi}{\gamma}}|q_F - q| = \overline{\pi}_F(q) \iff |q_F - q| > \sqrt{\frac{\gamma}{\phi}}(1 - \pi_F) \tag{1}$$

The intuition here is that when the minority group's share exceeds a certain threshold, its costs of conflict are such that it is convenient to engage in conflict with the majority. In addition, conflict guarantees larger electoral benefits the wider the distance $|q - q_F|$ between the policy level set by the majority and the minority's bliss point. We can thus formalize $S_F^*(q)$ as follows:

$$S_F^*(q) = \begin{cases} C, & \text{if } |q_F - q| > \sqrt{\frac{\gamma}{\phi}}(1 - \pi_F) \\ N, & \text{if } |q_F - q| < \sqrt{\frac{\gamma}{\phi}}(1 - \pi_F) \end{cases}$$

We define $\overline{q}_M = q_F - \sqrt{\frac{\gamma}{\phi}}(1-\pi_F)$, which is the policy level that makes the minority indifferent between engaging in conflict or not. We examine the case in which $q_F > q$, as the case $q_F < q$ does not happen in equilibrium[4]. Moreover, we define q^*, the best response of M, as

$$q^* = argmax_{q\in\mathbb{R}} U_M(q, S_F^*(q))$$

We can have two cases, depending on the relative position of $\overline{q}_M$ with respect to q_M.

1. Suppose $\overline{q}_M < q_M$, which can be written as $q_F - q_M < \sqrt{\frac{\gamma}{\phi}}(1-\pi_F)$: in this case, we have that the bliss point of the majority is able to prevent the conflict

$$q^* = argmax_{q\in\mathbb{R}} U_M(q, S_F^*(q)) = q_M$$

   This is the case in which the distance in preferences is sufficiently close to make it not profitable for the minority to trigger the political conflict. The majority chooses its desired policy level, and conflict is avoided.

2. On the other hand, suppose that $\overline{q}_M > q_M$, that is $q_F - q_M > \sqrt{\frac{\gamma}{\phi}}(1-\pi_F)$: here, we have a larger difference in preferences and the bliss point of the majority is not able to avoid the conflict. In this case, the majority must decide whether to choose $\overline{q}_M$, the minimal policy level that makes the minority indifferent between engaging in the conflict or not, or to vote q_M and trigger the conflict with the minority group. Thus, the majority prefers to engage in conflict in the case that $U_M(q_M, C) > U_M(\overline{q}_M, N)$, which is when:

$$\pi_F > \frac{\sqrt{\gamma}-\sqrt{\phi}(q_F-q_M)}{\sqrt{\gamma}(1-\sqrt{\phi})} = \widetilde{\pi}_F \iff (q_F - q_M) > \sqrt{\frac{\gamma}{\phi}}(1-(1-\sqrt{\phi})\pi_F)$$

   Otherwise, if $\pi_F < \widetilde{\pi}_F$, the majority prefers to vote for $\overline{q}_M$ and avoid the conflict. Therefore, depending on the level of π_F we can have either conflict or compromise

[4] The case $q^* > q_F$ cannot happen in equilibrium. To see this, notice that both $U_M(q, N)$ and $U_M(q, C)$ are symmetric around q_M. One has $U_M(\bar{q}_M, N) > U_M(q, N) \forall q \in \left(q_F, q_F + \sqrt{\frac{\gamma}{\phi}}(1-\pi_F)\right]$, and $U_M(q_M, N) > U_M(q_M, C) > U_M(q, C) \forall q \geq q_F + \sqrt{\frac{\gamma}{\phi}}(1-\pi_F)$. Therefore, the optimal q^* set by the majority is always lower than q_F.

between the two groups: in the first case, the outcome will be the majority's bliss point, while in the second case we have an outcome closer to the one preferred by the minority[5]. Note that this new threshold $\sqrt{\frac{\gamma}{\phi}}(1-(1-\sqrt{\phi})\pi_F)$ is higher than $\sqrt{\frac{\gamma}{\phi}}(1-\pi_F)$.

Therefore, we can formalize the equilibrium strategy chosen by the majority:

$$q^* = \begin{cases} q_M, & \text{if } q_F - q_M < \sqrt{\frac{\gamma}{\phi}}(1-\pi_F) \vee q_F - q_M > \sqrt{\frac{\gamma}{\phi}}(1-(1-\sqrt{\phi})\pi_F) \\ \overline{q}_M, & \text{if } q_F - q_M \in (\sqrt{\frac{\gamma}{\phi}}(1-\pi_F); \sqrt{\frac{\gamma}{\phi}}(1-(1-\sqrt{\phi})\pi_F)) \end{cases}$$

On the other hand, the equilibrium outcome decided by the minority is formalized as follows:

$$S_F^*(q^*) = \begin{cases} C, & \text{if } q_F - q_M > \sqrt{\frac{\gamma}{\phi}}(1-(1-\sqrt{\phi})\pi_F) \\ N, & \text{if } q_F - q_M < \sqrt{\frac{\gamma}{\phi}}(1-(1-\sqrt{\phi})\pi_F) \end{cases}$$

Here, we can see how conflict is influenced positively by the distance in preferences, the salience of the issue and the minority's share. Conversely, the more frequent future interactions between groups make the conflict less likely.

3.1 Equilibrium and comparative statics

We can have three possible equilibrium outcomes in terms of the implemented policy level and of the conflict/no conflict situation:

Proposition 1. *If the difference in preferences is particularly large, which is when $q_F - q_M > \sqrt{\frac{\gamma}{\phi}}(1-(1-\sqrt{\phi})\pi_F)$, the majority votes for its bliss point q_M and conflict happens.*

Here, the minority considers it profitable to engage in conflict since the large distance in preferences guarantees substantial electoral gains from fighting the majority. On the other hand, the majority chooses to implement its best-desired policy level.

Proposition 2. *If the two groups have particularly close preferences, which happens when $q_F - q_M < \sqrt{\frac{\gamma}{\phi}}(1-\pi_F)$, the majority votes for its bliss point q_M and there is no conflict.*

[5] For these two conditions to be possible we need $\widetilde{\pi}_F < 1/2$ and thus $q_F - q_M < \frac{\sqrt{\gamma}}{2}(\frac{1+\sqrt{\phi}}{\sqrt{\phi}})$. Note that $\widetilde{\pi}_F$ is always positive in this case.

In this second case, being the two bliss points close, for the minority is not profitable to engage in conflict and bear its costs. The majority has the power to implement its bliss point as the policy level and conflict is avoided.

Proposition 3. *In the case of intermediate preferences, formally when* $q_F - q_M \in (\sqrt{\frac{\gamma}{\phi}}(1 - \pi_F); \sqrt{\frac{\gamma}{\phi}}(1 - (1 - \sqrt{\phi})\pi_F))$*, the majority chooses a policy level closer to the minority's preferences and conflict is avoided: the set level is* $q^* = q_F - \sqrt{\frac{\gamma}{\phi}}(1 - \pi_F)$*.*

Overall, the intuition is the following: an increase in the share of the minority decreases its cost of conflict, thus making it more demanding in terms of policy level ($\overline{q}_M$ moves closer to q_F). On the other hand, this increases the cost of compromise for the majority, since it would have to accept a policy level more distant from its bliss point. As a consequence, it is possible that the majority decides to vote for its best-desired policy level and trigger conflict with the minority: the result, in this case, is a worse policy outcome from the minority's perspective. We can sum up the findings with respect to the minority share's π_F as follows:

Proposition 4. *Suppose* $q_F - q_M \in \left(\frac{1}{2}\sqrt{\frac{\gamma}{\phi}}, \sqrt{\frac{\gamma}{\phi}}\right)$ *and consider the effect of an increase in* π_F *from* $\pi_F^0 < \tilde{\pi}_F$ *to* π_F^1 *on equilibrium policy* q^**:*

- *there is no effect on* q^* *if* $\pi_F^1 < \bar{\pi}_F(q_M)$*;*
- q^* *strictly increases if* $\pi_F^1 \in (\bar{\pi}_F(q_M), \tilde{\pi}_F)$*;*
- q^* *weakly decreases (as a result of conflict) if* $\pi_F^1 > \tilde{\pi}_F$*, and in particular it strictly decreases if* $\pi_F^0 > \bar{\pi}_F(q_M)$*.*

In general, we have that conflict is more likely in the cases of:

- Larger difference in preferences $q_M - q_F$
- Higher salience of the public good issue ϕ
- Less frequent future interactions between the two groups γ
- Larger minority group's share π_F

The most interesting finding with respect to the whole paper's main argument is that conflict probability is directly proportional to the share of the minority. This is because when the minority gains more political power, its requests might become too expensive to accommodate for the majority. We can think for instance of the introduction of a gender quota increasing the share of female politicians: this can be followed by an increased demand for "gender-sensitive" policies[6]. These requests might become too exorbitant from the point of view of male politicians, who could prefer to establish their preferred policy level in terms of "gender-sensitive" policies, even if this means alienating a part of the political class. This is our main result: increasing the representation of a minority in the political class has the potential to damage the minority in terms of policy outcomes, at least in the short term. In the next section, we show how this theoretical finding can be confirmed empirically, by looking at the consequences of the introduction of a gender quota in Italy on the local expenditures for day care, possibly the most "gender-sensitive" among all categories of public expenditures.

4 Setting: the 2012 gender quota

In Italy, there are approximately 8100 municipalities, which represent the lowest sub-national level, after regions and provinces. Each municipality has a mayor assisted by a local council ("Consiglio comunale"), which owns the legislative power, and by an executive committee ("Giunta comunale") owning the executive power. The local administrators decide on the allocation of public funds over a large variety of categories since the provision of many public services is decentralized at the local level. There are three sources of financing for a municipality: own taxes and tariffs, transfers from the central government, and revenues from fines. Elections are held every five years, and voters can express their preference for both the mayor and local councillors. We focus here on municipalities between 2000 and 10000 inhabitants, given that outside this interval there might be different electoral rules: these cities constitute a relatively homogeneous sample that allows the implementation of

[6] This could, in turn, generate an increased salience of the gender-related topic and thus an increase in the parameter ϕ

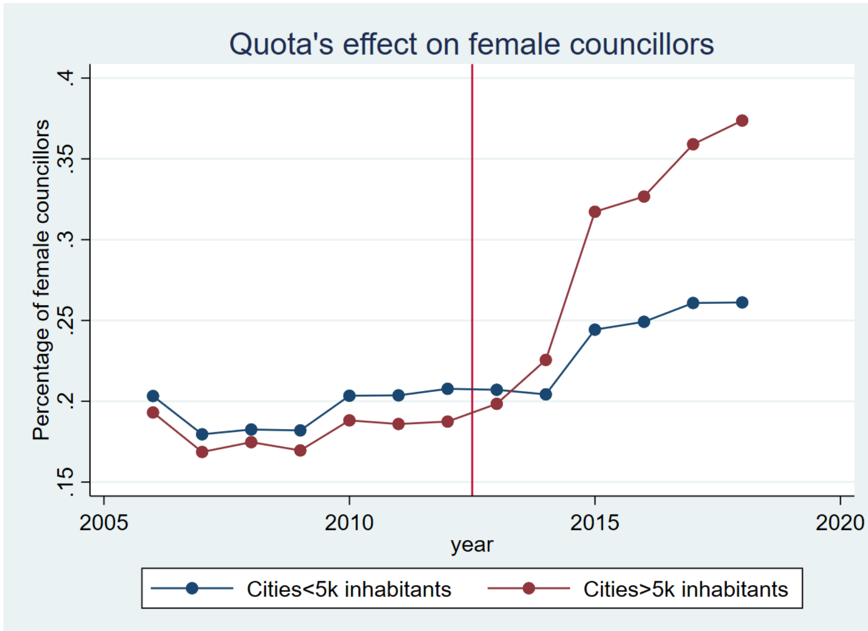
our identification strategy, as we explain in Section 6. For these cities, citizen over 18 years old can cast their vote for the mayor and, until 2013, for one local councillor, since the electoral system prescribes semi-open lists. Each mayoral candidate can be backed by one list of council candidates, and the mayoral candidate receiving the relative majority of votes obtains two-thirds of the seats that are allocated to his or her councillors. The remaining third of seats are allocated to other mayoral candidates via a proportional system. Seats are then attributed to councillors according to their vote ranking, which is relative to each party. A total of 12 councillors are elected through this system.

In this paper, we focus on Law 215/2012, which introduced two major changes for local elections across Italian municipalities with more than 5000 inhabitants. First, a gender quota was imposed on candidates' lists: a maximum of two-thirds of same-sex candidates for local councillors can be included. In addition, a second important novelty was introduced regarding the voting system: citizens could cast two votes for councillors instead of one with the condition that the two candidates were of different sex. The clear aim of the law was to increase women's presence among Italian local politicians, in order to reduce the country's gender gap in politics. According to this law, non-complying parties were punished with the removal of same-sex candidates exceeding two-thirds of the total in their lists.

The consequences of this law in terms of effects on councils' composition were studied by Baltrunaite et al. (2019), who, by the means of a regression discontinuity design, found that the law increased the percentage of female councillors by an average of 18 percentage points. They claim that this effect was mainly driven by the introduction of the double vote for councillors, since lists' compositions were not particularly affected. Moreover, they investigate whether the law provoked some additional effects in terms of elected candidates' characteristics, namely age, years of education or previous occupation: none of these characteristics seemed to have changed, not even focusing on same-sex candidates. In other words, the only effect of the law was to increase the percentage of elected female councillors, and not to alter other elected councillors' characteristics[7]. In Fig.1 we can see how the percentage of female councillors changed over the years for our municipalities of focus, namely those between 2000 and 10000 inhabitants.

[7] The law's effects in terms of turnout or of turnout by gender were also negligible

Fig. 1: % **Female councillors**



Note: Percentage of female councillors over the years in Italian municipalities with a population between 2000 and 10000 inhabitants. The red vertical line indicates the time when the law 215/2012 was introduced. Note that elections are staggered across municipalities, thus different municipalities might have had different election years. Data include 1434 out of 8100 Italian municipalities, observed between 2006 and 2018.

From Fig.1 we can notice how in municipalities interested by Law 215/2012 there was a sharp increase in female councillors, especially after 2013. On average, the percentage of female councillors passed from 18% in the pre-quota period to almost 38% in 2018 for the interested municipalities. It is important to point out that elections are staggered across municipalities, thus different municipalities might have different election years. The big spike in the increase in female councillors that we observe in 2014 is due to the fact that most municipalities interested by the quota held elections in that year, thus in 2013 their councillors were still the ones chosen with the old set of rules. We also need to specify that, even after the imposition of the new law, in almost all municipalities (93% of our sample) men detained the majority, which is required in order to pass motions on public expenditures: in other words, political power was still in the hands of male councillors. Moreover, we notice how there seemed to have been a spillover effect also for municipalities not interested by the quota, which also increased their shares of female councillors. This spillover might affect the interpretation of our empirical results in the sense that what we show in Section 6 is possibly only a lower-bound effect.

5 Data

We collected data on a total of 1422 municipalities with populations between 2000 and 10000 inhabitants, focusing on the period from 2013 to 2018. Given that Law 215/2012 is binding for municipalities with a population above 5000 inhabitants, we consider those municipalities as belonging to our treatment group, while the remaining ones constitute our control group. We decided to exclude the pre-2013 period since municipalities above the 5000 inhabitants threshold had different fiscal rules with respect to those below the threshold: they were indeed subject to the "Domestic Stability Pact"[8], that in 2013 was extended to all municipalities. Since we focus on public expenditures, our sampled municipalities need to have a homogeneous set of fiscal rules and this condition is satisfied for the period after 2013. To conduct our analysis, we merged two datasets. The first one, collected from the website of the Italian Ministry of Interior, regards characteristics of elected councillors, mayors and aldermen, namely age, sex, party and previous job. Since the units of observation of our analysis are municipalities and not single politicians, for each municipality and year we created several variables indicating the mean age and years of education of the councillors in charge, together with the percentage of female councillors. In Tab.1 we present some summary statistics for councillors, first for the whole sample and then for the treatment and control groups.

[8] The Domestic Stability Pact introduced some restrictions on municipalities' expenditures: the overall budget balance had to be proportional to a moving average of the balances of the previous years. Lack of adherence implied sanctions to municipalities, while compliance granted a reduction in interest expenses for government loans.

Tab. 1: **Summary statistics for municipalities' councils**
Whole sample

Variable	Obs	Mean	Std.Dev.	Min	Max
Mean share of female councillors	8,291	0.26	0.14	0	0.76
Councillors' mean years of education	8,064	13.93	1.37	8	18
Councillors' mean age	8,076	51.75	4.66	25	70.25

Treatment group

Variable	Obs	Mean	Std.Dev.	Min	Max
Mean share of female councillors	3,494	0.30	0.15	0	0.76
Councillors' mean years of education	3,409	14.19	1.37	8	18
Councillors' mean age	3,412	51.66	4.66	38	69

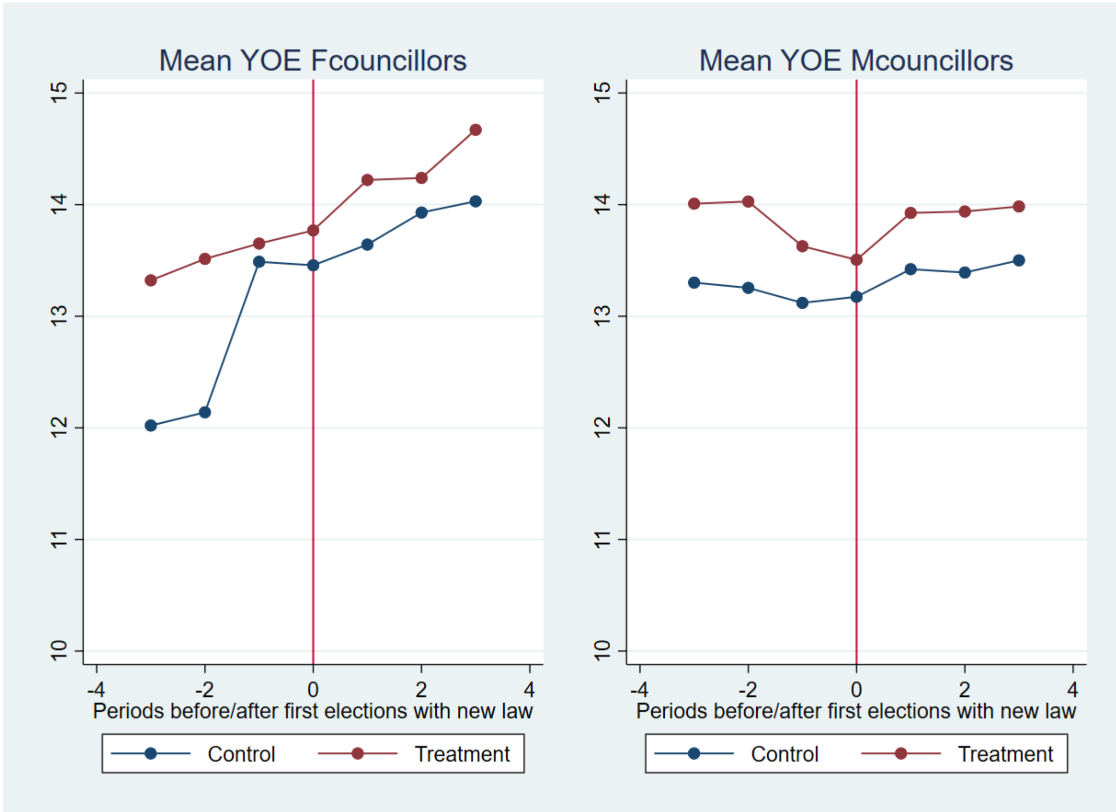
Control group

Variable	Obs	Mean	Std.Dev.	Min	Max
Mean share of female councillors	4,797	0.23	0.13	0	0.69
Councillors' mean years of education	4,655	13.74	1.34	8	18
Councillors' mean age	4,664	51.81	4.65	25	70.25

Notes: Summary statistics for sampled municipalities' councils, namely councillors' mean age, councillors' mean years of education and mean share of female councillors. The treatment group includes sampled cities with population above 5000 inhabitants, while the control group includes those below. Sampled municipalities are 1422 out of 8100 Italian municipalities, observed over the period between 2013 and 2018.

As we can observe from Tab.1, there are no striking differences between the two groups of councillors, apart from a larger female presence (due also to Law 215/2012) and a higher educational level in the treatment group. It is worth repeating that existing evidence showed that the quota did not change the educational level or average age of either male or female councillors (Baltrunaite et al. (2019)). Indeed, we can also notice this graphically: Fig.2 shows the evolution of the average educational level for male and female councillors by year, before and after the first elections under the new law. As we can see, no particular different trends appear to have been produced by Law 215/2012 for this particular variable.

Fig. 2: **% Trends in educational levels**

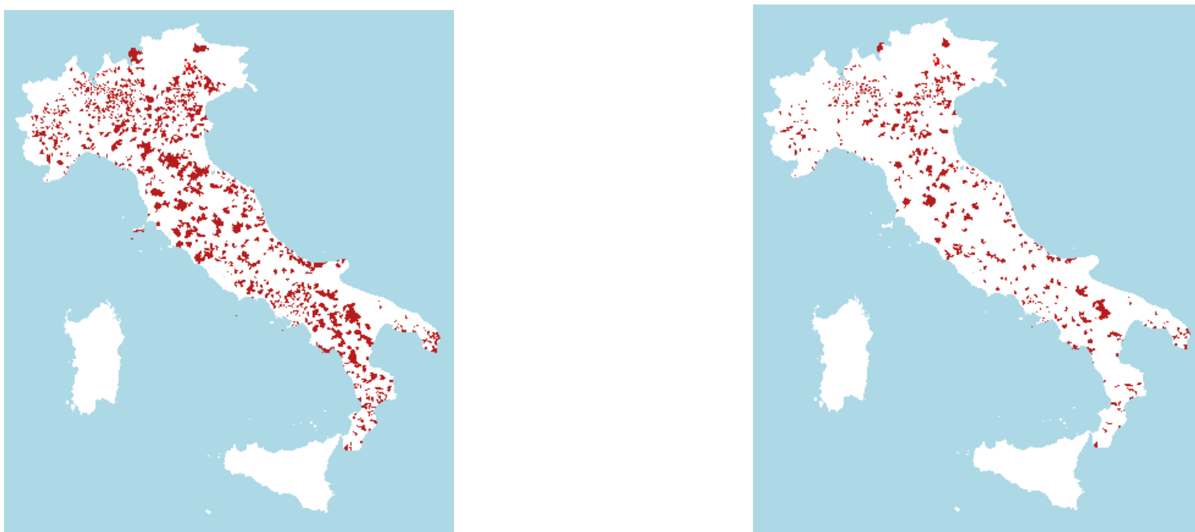


Note: Trends in average educational levels for male and female councillors. The red vertical line indicates the time when a municipality held its first elections under the law 215/2012, while the x-axis indicates the years before and after this time. The treatment group includes municipalities that were subject to the new law, while the control group the remaining ones. Data include 1434 out of 8100 Italian municipalities, observed between 2006 and 2018.

Moreover, we need to point out that most of these politicians do not have a clear ideological leaning: 95% of them belong to the so-called "liste civiche" (civic lists), which are parties not possible to locate on the right-left axis. Therefore, our sampled politicians are not influenced by any central party's guidelines, and are freer to vote on their preferred policies with respect to politicians elected at higher levels (e.g. the national Parliament or regional councils).

In Fig.3 we present the geographical distribution of sampled and treated municipalities: these two maps give us confidence that there is no geographical bias neither in our sampling or assignment to treatment procedure.

Fig. 3: **Sampled and treated municipalities**



(a) Sampled municipalities (b) Treated municipalities

0.80.8 *Notes*: Fig.3a shows all municipalities belonging to our sample, while Fig.3b shows the subsample of treated municipalities, with population higher than 5000 inhabitants. Treated municipalities were subject to Law 215/2012 since their first election after 2012. Sampled municipalities are 1422 out of 8100 Italian municipalities. Special regions' municipalities (e.g. those in the main Islands) were excluded from the sample since they are subject to different fiscal rules with respect to ordinary regions ones.

Together with the dataset on local politicians, we collected data on municipalities' expenditures from the Istat website. Since a city's local council has discretion over more than 100 spending categories, we decided to focus on the most "gender sensitive" one, for which we expect to see a quota's effect: day care expenditures. Italy is historically conservative in its family models, and traditionally childcare is performed by mothers and grandmothers (Brilli et al. (2016)). This is due to longstanding social norms on gender roles, that privilege informal child care provided by female family members with respect to formal services (Barigozzi et al. (2019)). As a consequence, both private and public day care is less provided with respect to other European countries (Brilli et al. (2016)), with public preschool investment being one of the lowest in Europe. Public investments in day care have been growing in the last decades, but Italy still remains below the European average for the percentage of GDP devoted to child care and pre-primary schools (Barigozzi et al. (2019)). We decide to focus on this particular spending category because we have evidence from the literature that female politicians are more inclined to invest in day care (Bhalotra and Clots-Figueras (2014), Miller (2008)) with respect to male politicians. Having subsidized day care helps working mothers to cope with their job duties, increasing maternal employment and generating also possible positive effects for children's well-being (Herbst (2017), Brilli et al. (2016), Miller (2008)). Thus, we might expect that the substantial increase in female councillors had a positive effect on the portion of public funds devoted to childcare. This is why we collected data on local expenditures on childcare, together with the number of available spots and users by municipality and year. The main dependent variable of our analysis, day care expenditures per capita, is resulting from three expenditure categories added up, with the total divided by the municipality's population:

- Municipality's expenditure for directly managed day care facilities
- Municipality's funding to privately managed day care facilities[9]
- Municipality's funding for day care-related expenses. This includes all subsidies to families for day care integrated services or other related expenses, for instance the "babysitter bonus"

All these three levels of expenditures are decided by local administrators and have to be approved by the local council[10]. Here, when we speak of day care facilities, we refer to nurseries for early childhood assistance, which in Italy are meant for children aged between 0 and 3. In Tab.2 we show some summary statistics for this variable, again for the whole sample and then for treated and control municipalities.

[9] Local councils can decide to directly provide this service or outsource it to privates

[10] There might be occasional state-level contributions to daycare. For instance, national law 232 of the 11th December 2016 granted a 1000 euros voucher for kindergartens to all families. We argue that this effect should not bias our results since it is not different for treated and control municipalities, and any potential bias should be absorbed by year-fixed effects.

Tab. 2: **Municipalities' day care expenditures per capita**

Whole sample

Variable	Obs	Mean	Std.Dev.	Min	Max
Day care exp. pc	8,388	7.22	12.50	0	112.95

Treatment group

Variable	Obs	Mean	Std.Dev.	Min	Max
Day care exp. pc	3,519	8.94	13.90	0	83.94

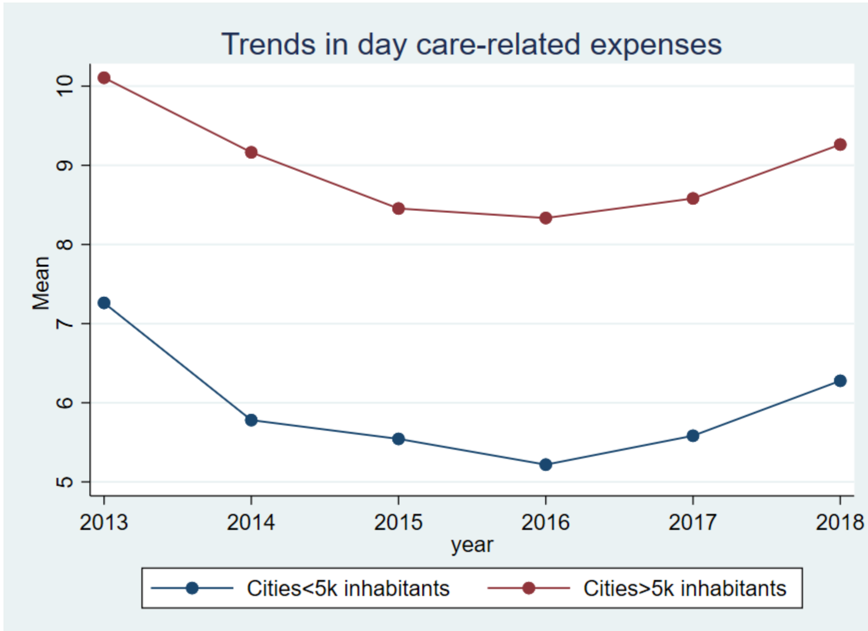
Control group

Variable	Obs	Mean	Std.Dev.	Min	Max
Day care exp. pc	4,869	5.97	11.22	0	112.95

Notes: Summary statistics for day care expenditure per capita at the municipality level. The variable, expressed in euros, includes three expenditure categories added up: expenditures for directly managed day care facilities, funding to privately managed day care facilities, and municipality's funding for day care-related expenses. These kinds of expenses represent a fraction between 0.6 and 1% of total expenses taken yearly by our sampled municipalities. The treatment group includes sampled cities with a population above 5000 inhabitants, while the control group includes those below. Sampled municipalities are 1422 out of 8100 Italian municipalities, observed over the period between 2013 and 2018.

In Fig.4 we can see the evolution of this variable for both treated and control municipalities over the observed years. As we can see, trends are extremely similar, even if treated municipalities exhibit higher levels of expenditures per capita on average.

Fig. 4: **Trends in day care expenditures for treated and control muncipalities**



Note: Trends in day care-related expenditure per capita at the municipality level over the period 2013-2018. The variable, expressed in euros, includes three expenditure categories added up: expenditures for directly managed day care facilities, funding to privately managed day care facilities and municipality's funding for day care-related expenses. These kinds of expenses represent a fraction between 0.6 and 1% of total expenses taken yearly by our sampled municipalities. The treatment group includes sampled cities with population above 5000 inhabitants, while the control group includes those below. Sampled municipalities are 1422 out of 8100 Italian municipalities.

In the following section, we explain how our identification strategy allows us to assess the effect of the introduction of Law 215/2012 on day care-related expenses.

6 Empirical strategy and main results

The major goal of this paper's section is to empirically test our theoretical predictions, and, in particular, to understand which consequences the increase in female councillors had on local expenses on day care. One of the major implications of our model is that a large increase in the political representation of a minority could have detrimental effects in terms of policy: in our case, we identify this particular change in outcome with a decrease in day care expenditures. On the other hand, when the new law did not trigger a large increase in the political power of female politicians, this should have benefited them in terms of policies, thus we might expect an increase in daycare expenditures. For testing these implications, we make use of a Difference in Discontinuity identification strategy (Grembi et al. (2016)), which combines a Regression Discontinuity design with a Difference in Difference analysis. In our case, we focus on the change in day care expenditures per capita at the threshold of 5000 inhabitants after the first elections with the new legislation on councillors' lists and the alternate vote. Thus, our treatment is represented by having a council elected after 2012 in a municipality with a population larger than 5000 inhabitants. We evaluate the effect of this treatment on our outcome of focus, which is the municipality's day care expenditures per capita. We first present the aggregate results, and then the heterogeneous treatment effect with respect to the post-quota share of female councillors. In the next subsection, we discuss our identification strategy and its validity assumptions. After that, we present our main results and some related robustness tests.

6.1 Identification strategy and validity tests

Our model takes the following functional form:

$$Daycarepc_{it} = \alpha + \beta * Treatment_i + \gamma * PostLaw_{it} + \delta * TreatPost_{it} + \zeta * normPop_i * (\eta * Treatment_i + \theta * PostLaw_{it} + \iota * Treatment_i * PostLaw_{it}) + Year_t + City_i + \Lambda * Mayor_{it} + \epsilon_{it}$$

With:

- $Daycarepc$ the sum of per capita day care-related expenses of municipality i at time t;
- $Treatment$ dummy for a municipality with more than 5000 inhabitants;
- $PostLaw$ dummy for expense taken by a council elected after Law 215/2012 entered into force;
- $TreatPost$ interaction between dummies $Treatment$ and $PostLaw$;
- $normPop$ normalized population, equal to population-5000;
- $Year$ year fixed effects;
- $City$ municipality fixed effects;

- $Mayor$ set of current mayor's controls (age, sex and level of education).

The treatment effect is identified by the coefficient δ, capturing the effect of the introduction of the law on treated municipalities around the threshold of 5000 inhabitants. We consider only municipalities belonging to a bandwidth $[normPop-h;normPop+h]$, with h computed by one of the methodologies suggested by Calonico et al. (2020), the standard MSE-optimal bandwidth[11]. Following Gelman and Imbens (2019), we perform only a linear fit and we don't use any higher-order polynomial to identify our treatment effect at the threshold.
A Difference in Discontinuity strategy allows to disentangle the quota effect from another confounder that we have at the same 5000 inhabitants' thresholds: the mayor's wage. While for municipalities under 5000 inhabitants mayors earn 2170 euros per month, in cities above this cutoff (up to 10000 inhabitants) mayors' monthly wage increases to 2790 euros. This confounder might have effects on both selection into politics and on subsequent kinds of policies adopted (Gagliarducci and Nannicini (2013)), possibly biasing our results in case we implemented a simple Regression Discontinuity design. On the other hand, a dynamic strategy such as the Difference in Discontinuity can disentangle the quota's effect from the wage's effect. Moreover, the fact that this strategy focuses on identifying the treatment effect at the threshold (thus, considering only municipalities belonging to a bandwidth) allows a more convincing comparison between treated and control units with respect to a simple Difference in Difference[12]. Three validity assumptions need to be satisfied for this identification strategy to be implemented (Grembi et al. (2016)). The assumptions are the following:

1. All potential outcomes are continuous in the running variable at the threshold
2. The effect of the confounding policy is constant over time
3. The effect of the treatment at the cutoff does not depend on the confounding policy

[11] Our results are however robust to the adoption of another, narrower, bandwidth minimizing the coverage error rate.

[12] Our major Difference-in-Discontinuity results are however robust in coefficients' signs and magnitudes to the use of a less local identification strategy such as the Difference-in-Difference

The first condition is that at our threshold of 5000 inhabitants, all potential outcomes are continuous in the running variable, which is the normalized population [13]. To check compliance with this condition, we test whether a wide set of covariates are balanced at the threshold by performing a series of pre-treatment Regression Discontinuity designs, using these covariates as outcomes. The set of covariates includes both geographic time-invariant characteristics[14] and demographic time-varying factors[15]. In Tables 3 and 4 we present our results.

Tab. 3: **RDD with time-invariant characteristics**

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	South	River	Lake	Surface	Sea distance	Altitude
RD_Estimate	-0.063	0.041	-0.069	3.034	-7.830	31.902
	(0.114)	(0.068)	(0.056)	(4.206)	(7.410)	(24.176)
Observations	3,643	3,643	3,643	3,643	3,643	3,643

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Regression discontinuity designs with municipalities' time-invariant covariates as main outcomes. The running variable is the normalized population, with 5000 inhabitants (corresponding to zero in the normalized population) as the discontinuity cutoff. Outcomes include South (dummy for municipality located in the south), River (dummy for river presence), Lake (dummy for lake presence), distance from the sea, surface extension, mean altitude level. Regressions are performed over the pre-treatment period, in other words over the period before the first elections with Law 215/2012 in force. Sampled municipalities are 1422 out of 8100 Italian municipalities, observed over the period between 2013 and 2018.

[13] Normalized population here is defined as the municipality's population - 5000.

[14] Among the time-invariant characteristics, we have South (dummy for municipality located in the south), River (dummy for river presence), Lake (dummy for lake presence), distance from the sea, surface extension, mean altitude level

[15] The set of demographic factors includes fertility rate, gender participation gap, female unemployment rate, percentage of graduated men and women and age structure (ratio of over 65 over under 15 years old residents)

Tab. 4: **RDD with time-variant characteristics**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Fertility	Gender Part.gap	FPLF	F.Unempl.	% Grad.women	% Grad.men	Age structure
RD Estimate	0.552	0.348	1.035	0.850	0.354	0.324	5213
	(1.683)	(0.299)	(0.838)	(0.677)	(0.296)	(0.302)	(0.535)
Observations	3,643	3,643	3,643	3,643	3,643	3,643	3,643

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Regression discontinuity designs with municipalities' time-invariant covariates as main outcomes. The running variable is the normalized population, with 5000 inhabitants (corresponding to zero in the normalized population) as the discontinuity cutoff. Outcomes include South (dummy for municipality located in the south), River (dummy for river presence), Lake (dummy for lake presence), distance from the sea, surface extension, mean altitude level. Regressions are performed over the pre-treatment period, in other words over the period before the first elections with Law 215/2012 in force. Sampled municipalities are 1422 out of 8100 Italian municipalities, observed over the period between 2013 and 2018.

As we can observe, there are no jumps in municipalities' characteristics at the 5000 inhabitants' threshold, thus it seems that a set of potential outcomes is continuous in the running variable at this cutoff. In addition, we argue that any manipulation in the running variable (for instance, by mayors seeking a higher salary) is unlikely. First, because the set of laws that becomes binding at population thresholds, such as Law 215/2012, take as reference value the population measured by the last census, in this case 2011 [16]. And secondly, the population is measured by independent employees from Istat[17], with no presumable interest in manipulating the true value. However, in Appendix B we perform the McCrary test to verify the presence of manipulation in our running variable at the threshold, confirming that its observational density is continuous.

Turning to the second validity assumption, we need the confounding policy, the mayor's salary, to have constant effects over time. We know that a higher salary can influence both the selection of citizens into politics and the kind of policies implemented (Gagliarducci and Nannicini (2013)). The authors who studied the effects of Law 215/2012 found no change in candidates' lists composition, in terms of age, education, or previous job (Baltrunaite et al. (2019)) after the novelties were introduced: thus, we can state that the selection effect was constant over time. Speaking instead of the effects of having different kinds of mayors on the policies implemented, our solution is to control for mayors' characteristics in our set of

[16] Censuses in Italy have been performed every 10 years since 1861 until 2019, when they became yearly.
[17] Istat is the Italian National Institute of Statistics

regressions. Each of the regressions we show in the Main Results' section includes controls for the current mayor's sex, age, and years of education, which should capture this potential confounding effect on policies.

Finally, we need to discuss the third validity assumption: there should be no interaction between the new law's effects and the confounding policy. To check whether the assumption is satisfied, we perform a series of Difference in Discontinuity regressions interacting the treatment variable with a set of variables for each mayor's characteristic. Indeed, we know that a higher salary, the confounding policy, creates incentives for different individuals to enter into politics: thus, if our treatment effect is different because of this confounder, we should observe significant interaction effects between the treatment coefficient and mayors' characteristics. However, we do not observe any significant coefficients in the interaction terms, as we show in Appendix A. Therefore, we can be confident also about compliance with the third validity assumption[18].

6.2 Difference in Discontinuity results

In this section, we first show the aggregate Difference in Discontinuity results, and then we move our focus to the relevant heterogeneity analysis. In Tab.5 we present the first results, adopting the model described in section 3, with and without Mayors' controls. To preserve space, we decided to present only the coefficient capturing the treatment effect, TreatPost.

[18] If the third validity assumption holds, the Difference in Discontinuity identifies the Average Treatment Effect, which can be generalized to the whole sample of observations. On the other hand, in case this assumption did not hold, we would have a Local Average Treatment Effect: the effect that we observe would be valid only for the treated municipalities.

Tab. 5: **Difference in Discontinuity aggregate results**

	(1)	(2)
VARIABLES	Daycarepc	Daycarepc
TreatPost	-0.715	-0.671
	(0.938)	(0.984)
Mayor's controls	NO	YES
Municipality FE	YES	YES
Year FE	YES	YES
Observations	2,690	2,629
R-squared	0.025	0.029
Number of municipality1	577	574

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Difference in discontinuity regressions. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. The dependent variable is yearly per capita expenditures on day care-related categories. Expenditures taken during election years are attributed to the council governing up to the elections' month. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

As we can observe from Tab.5, the empirical evidence tells us that the aggregate effect of Law 215/2012 on day care-related expenses was zero. Moreover, we see a null aggregate impact also if we look at the treatment effect over time, by a dynamic Difference in Discontinuity similar to the one described by Vannutelli (2021): we show this ulterior piece of evidence in Appendix C.

However, our major theoretical implications state that increasing a minority's political representation can have different policy effects depending on the subsequent size of the minority group. In our empirical setting, we would therefore need to assess the heterogeneous law's effects on day care with respect to each council's post-quota share of female councillors. The next section is devoted to analyzing these possible heterogeneous effects.

6.3 New law's heterogeneous effects on day care

Our model concludes that an increase in the share of a minority reduces the probability of compromise with the majority, increasing the chances that the majority votes its bliss point. Therefore, in our empirical context we would expect to see a decrease in day care expenditures in places where the increase in female politicians was so large that prevented bargaining on day care between male and female councillors. Therefore, we look at heterogeneous results with respect to the post-quota share of female councillors in treated municipalities. In order to do this, we proceed in the following way:

1. Generate the variable “Female post quota”, indicating the percentage of female councillors elected in the first election after the new law entered into force

2. Create percentile dummies based on the distribution of this variable across treated municipalities. To be more precise, we create the dummies for being lower than the 25th, higher than the 50th, 75th or 90th percentiles

3. Interact the treatment variable $TreatPost$ with each percentile dummy and assess the heterogeneous effects of the new law on day care expenditures. We interact the 25th dummy to look for the effect on councils with low post-quota female presence. The interactions with the 50th,75th and 90th percentile should assess the effects in councils with high post-quota shares of female councillors.

We exclude from this analysis the few treated councils in which there was a female majority, since they would not comply with our model's assumptions. For our analysis to be correct, we would need gender to prevail over ideology when voting for day care, since coalitions should be forming also with members of different parties: we believe this to be true for three reasons. First, there is evidence that female politicians have stronger preferences for day care with respect to their male colleagues also in developed contexts such as Germany or Belgium (Baskaran and Hessami (2023), Slegten and Heyndels (2020)). Subsidized day care is a service that benefits disproportionally women with respect to men, and this supports the idea of a gender difference in preferences for this kind of public expenditure (Bhalotra and Clots-Figueras (2014)). In addition, we know that the 95% of politicians in our sample belong to civic lists, without a clear ideological belonging: this guarantees them more freedom in the voting process, without having to follow any national party's guidelines. Last, voting in the Italian local councils is secret, thus there is even more freedom for councillors in choosing the alternative that they best prefer. In Tab.7 we show how the increase in female politicians affected day care expenditures differently on the basis of each council's post-quota share of female councillors.

Tab. 6: **Difference in Discontinuity heterogeneous results**

VARIABLES	Baseline interaction Daycarepc	25th percentile Daycarepc	median Daycarepc	75th percentile Daycarepc	90th percentile Daycarepc
TreatPost	3.936	-0.984	-0.104	-0.247	-0.615
	(2.541)	(1.059)	(1.049)	(1.022)	(1.017)
TreatPost*Fem. post quota	-12.211*				
	(6.522)				
Twenty-fifth p.		2.088**			
		(1.009)			
TreatPost*Twenty-fifth p.		1.863*			
		(1.130)			
Median			-1.484		
			(0.938)		
TreatPost*Median			-1.800*		
			(1.020)		
Seventy-fifth p.				1.518	
				(1.917)	
TreatPost*Seventy-fifth p.				-3.510	
				(2.444)	
Ninetieth p.					0.991
					(1.706)
TreatPost*Ninetieth p.					-2.447
					(2.121)
Observations	2,382	2,446	2,446	2,446	2,446
R-squared	0.035	0.032	0.034	0.037	0.032
Number of municipality1	544	559	559	559	559
Mayor controls	YES	YES	YES	YES	YES
Municipality FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Robust standard errors in parentheses are clustered at the municipality level
*** p<0.01, ** p<0.05, * p<0.1

Note: Difference in discontinuity regressions. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. Interaction terms indicate the post-quota share of elected female councillors, taken with respect to the overall distribution of treated cities' councils. The dependent variable is yearly per capita expenditures on day care related categories. Expenditures taken during election years are attributed to the council governing up to the elections month. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

Looking at Tab.6, we can immediately observe that the heterogeneous treatment effect, given by the sum between $TreatPost$ and $TeatPost * percentile$ takes different directions on the basis of the share of post-quota female politicians in cities' councils. The interaction between the treatment coefficient and the continuous measure of $Femalepostquota$ is negative, following our theoretical reasoning: in the councils with higher minority representation, the introduction of the law was followed by a decrease in day care expenditures compared to control municipalities. This is also confirmed when we look at the third, fourth and fifth columns, with the interaction between treatment and $Femalepostquota$ higher than the median, 75th and 90th percentile, all characterized by negative coefficients. The logic is that in these cities the new law increased the share of female councillors up to a level where a compromise with male councillors on day care expenditures would have been too expensive for the majority, who therefore decided to go with its respective bliss point and implementing an expenditure level lower than the pre-quota amount. On the other hand, we see in the second column of Tab.6 that in councils with low post-quota levels of female councillors (lower than the 25^{th} percentile), the increase in female politicians was followed by an increase in funding for day care compared to the control group. This is also aligned with our theoretical implications: with low minority representation, a compromise between the two groups of politicians is still possible and the larger political power of female politicians is able to move the policy outcome closer to their desired outcome. To give a sense of the magnitude of our effect, we have that in cities with low levels of female councillors the increase in expenditure was about 0,88 euros per capita, a 9.5% growth with respect to the pre-quota period and the control municipalities. On the other hand, in cities with $Femalepostquota > 50^{th}$ percentile, the size of the decrease was about 1,7 euros per capita, an 18% reduction with respect to the control group[19]. Recall that in all these municipalities men still constitute the majority of councillors, which is required in order to approve or reject motions on public expenditures: therefore, even after the quota, male politicians still detained the decisional power. We might be worried that the share of female councillors is persistent, or in other words that the cities with high/low shares of female councillors had different starting levels in this variable. In Appendix D, we add as control the pre-law share of female councillors: our main results are preserved even after the inclusion of this control. Moreover, in the next section we offer multiple pieces of evidence supporting the robustness of our results, starting

[19] Using Istat data, we verified whether this effect corresponded to a change in the number of available spots or in the number of subscriptions to day care facilities. We find a significant increase in public spots and subscriptions in cities with low shares of post-law female councillors, and a negative effect in those with high shares.

from a comparison between cities with high and low shares of post-law female councillors.

7 Robustness and additional evidence

7.1 Cities and post-quota shares of female councillors

A major concern for our analysis is that what we are observing is not the heterogeneous effect with respect to the post-quota share of female councillors but with respect to some other factor correlated with this variable. To exclude this hypothesis, we compare the average levels of some representative variables[20] for treated cities with high and low shares of female councillors (respectively, post-quota female councillors higher than the 75th percentile and lower than the 25th percentile), and check if the differences in means are statistically significant. In Tab.7 we can observe these comparisons.

[20] We look at variables referring to municipalities' balance sheets (e.g. Tot. expenditures), population's characteristics (e.g. Years of Education or Age Structure), geographical characteristics and local politicians' attributes.

Tab. 7: **Test for difference in means, cities with low and high levels of post-quota female councillors**

Variable	High_Share	Low_Share	Difference	p-value
Tot. Expenditures	2937784	2569516	-368268.4	.008***
Mean Y.o.E. population	8.820	8.896	.075	.300
Age Structure (M)	1.356	1.332	-.023	.608
Age Structure (F)	1.787	1.725	-.062	.349
Age Structure (Tot.)	1.564	1.521	-.043	.434
Fertility rate	39.342	38.963	-.378	.805
River	.431	.417	-.013	.824
Lake	.315	.379	.064	.262
Sea Distance	69.723	80.610	10.886	.111
Surface (km^2)	33.796	32.543	-1.253	.777
Mean altitude	227.177	231.675	4.498	.855
Population density	371.505	325.033	-46.472	.278
South	.341	.291	-.050	.383
Gender participation gap 2011	20.302	19.863	-.439	.432
Female labor force participation 2011	41.385	42.388	1.002	.256
Fertility rate	39.342	38.963	-.378	.805
Female unemployment rate 2011	12.614	12.202	-.412	.639
Female mayor	.140	.177	.037	.387
Y.o.E. mayor	15.819	15.873	.053	.874
Mean Y.o.E. councillors	14.345	14.433	.087	.582
Mean age councillors	49.599	50.620	1.0211	.027**
Mean Y.o.E. aldermen	14.434	14.475	.040	.919
Mean age aldermen	45.890	45.194	-.695	.499
Mean Y.o.E. female councillors	14.410	14.784	.374	.409
Mean age female councillors	43.063	43.360	.296	.657
Mean Y.o.E. male councillors	13.912	14.154	.242	.227
Mean age male councillors	48.009	47.778	-.230	.683
Union of municipality dummy	.319	.392	.072	.203

Note: Test for statistical differences in means of variables between group of treated municipalities with low and high shares of post-quota female councillors (post-quota female councillors<25th percentile and post-quota female councillors>75th percentile). Demographic variables refer to the whole municipality's population, or to the whole population by gender. Age ratios correspond to the ratio between the population over 60 and the population under 20 years old. Information on revenues and expenditures comes from municipalities' balance sheets. South is a dummy with value 1 in case that the municipality is located in Southern Italy. Only treated municipalities belonging to our optimal bandwidth are included.

The table tells us that there are a couple of statistically significant differences between the two groups of cities. What we observe is that cities with high levels of post-quota female councillors tend to spend more on average and have slightly younger councillors. The latter difference is not problematic, since it is only one year of difference in means and it should not bias our results. Regarding the other difference, if we include Total Expenditures as a control in our Difference in Discontinuity results, the main coefficients' magnitude, size and significance levels are preserved. Thus, we can be more confident that our main results are driven by a combination of the introduction of the new law and the subsequent shares of female councillors generated in the treated municipalities' local councils[21]. In particular, we underline how fertility rates, which are a potential huge driver of the demand for childcare, are absolutely comparable between the two groups of cities.

7.2 Placebo test - other expenditure categories

In this section, we conduct an important placebo test, verifying the new law's effect on other expenditure categories. We know from the model that the conflict outcome is more likely in the case of divergent preferences between two groups, and of more salience around the policy. In this sense, we know that men and women have different preferences with respect to day care (Barigozzi et al. (2019)), but we might also think that the introduction of the gender quota raised the salience of gender topics, which can include expenditures on day care. Indeed, according to experimental evidence (Coffman (2014)), women are more likely to expose their personal ideas in environments which are not male-dominated: an increase in the share of female councillors could have encouraged them to bring their political views on the discussion table. Thus, male and female politicians should have clashed over this issue but not on other kinds of expenditures which were not gender-related: here, we verify the soundness of this hypothesis. Given that each municipality's balance sheet contains more than 100 expenditure categories, we focus here on the 10 most relevant ones in terms of share

[21] I performed also the test in differences (i.e. comparing municipalities with high post-quota increases in female councillors and municipalities with low increases in female councillors), and the results were in line with this test. Few variables exhibited statistically significant differences in means, for instance total revenues, but not so large in magnitude to generate concerns.

of total expenditures, with the condition for them to be decided by the local council[22]:

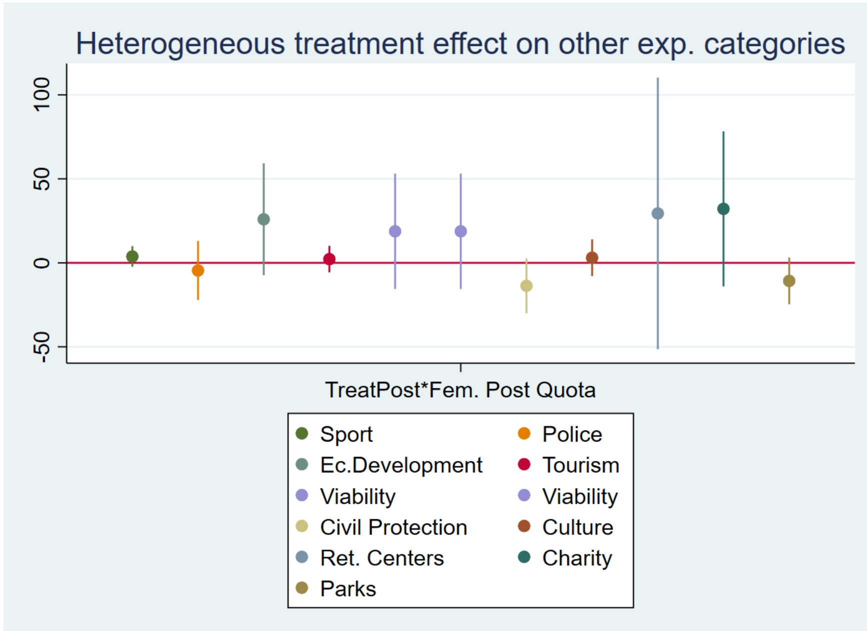
- Sport
- Police
- Economic development
- Tourism
- Public viability
- Civil protection
- Cultural events
- Retirement centers
- Charity
- Public gardens

All these expenditure categories, differently from day care, are not gender-related: therefore, there should not be a difference in preferences between male and female councillors[23]. To preserve space, in Fig.5 we exclusively show the interaction coefficients between the treatment dummy and the Female post-quota variable. Each coefficient displayed refers to the interaction between Treatpost and Female post-quota when we perform our Difference in Discontinuity regression with the outcomes indicated in the legend.

[22] We focus on current, and not capital, expenditures per capita since they are more likely to be affected by sudden changes in the political class composition. However, results with capital expenditures are not different in terms of magnitude/significance level of coefficients.

[23] The national government can contribute through its central funds to municipalities' expenditures, but funds are usually not restricted to a specific service or infrastructure. The local government has the possibility to decide how to administer the funds and to which public service financial support is most needed. All the expenditure categories in the placebo exercise are under the responsibility of the local administrators. A major exception to this setting is healthcare, which is the responsibility of the regional government.

Fig. 5: **Law's effect on other expenditure categories**



Note: Difference in Discontinuity heterogeneous effects with respect to the city council's post quota shares of female councillors. Each coefficient corresponds to the interaction between Treatpost and Female Post Quota, when we perform our Difference in Discontinuity regression with the expenditure categories indicated in the legend as the main outcomes. Confidence intervals are at the 95% level.

From Fig.5, we can see how there seem to be no significant effects from the interaction between Treatpost and Female Post Quota on any of these expenditure categories: we exclusively observe our main result on day care expenditures, given its gender-related connotation. Indeed, in line with our model, the exogenous shock brought by the gender quota moved the political debate over gender topics, thus increasing salience over its related policies and raising the chances of conflict. On the other hand, salience was not increased on other kinds of expenditures, thus the two groups of councillors did not conflict over them. In addition, this placebo is also useful to reduce the concern over other potential confounding policies (e.g. the Domestic Stability Pact) which could have driven the differential trends in day care expenditures between treated and control municipalities.

In addition to this robustness, in Appendix E we perform two additional placebo tests with artificial cutoffs and year of the law's implementation. Results from these two more tests confirm that what we observe in our main regressions are not simply spurious results.

7.3 2SLS panel regression

We perform a further robustness analysis by changing the identification strategy and recurring to an instrumental variable approach. Indeed, we might think that the Female post quota variable is endogenous to some unobserved municipality's characteristics: the shares of elected female politicians can indeed depend on a large variety of factors intrinsic to the electorate's and city's characteristics (Hessami and da Fonseca (2020)). Thus, given the possible endogeneity of this variable, in this section we instrument the percentage of female councillors with Treatpost: this way, we are able to evaluate the effect on day care expenditure of the exogenous increase in female politicians driven by the introduction of the new law. The model we adopt in this Section takes the following specification:

1st stage:

$$FemaleCouncillors_{it} =$$
$$\alpha + \beta * Treatment_i + \gamma * PostLaw_{it} + \delta * Treatpost_{it} + Year_t + City_i + \Lambda * Mayor_{it} + \epsilon_{it}$$

2nd stage:

$$Daycarepc_{it} =$$
$$\zeta + \eta * Treatment_i + \theta * PostLaw_{it} + \iota * \widehat{FemaleCouncillors}_{it} + Year_t + City_i + \kappa * Mayor_{it} + \mu_{it}$$

The coefficient ι captures the Local Average Treatment Effect of female councillors on day care expenditures, that is the effect of the law-induced increase in the share of female councillors on funding for day care[24][25]. As we explain in Section 4, the law 215/2012 had a positive effect on the share of female councillors. This positive effect is confirmed in Tab.8,

[24] This effect is "Local" in the sense that it interests only those treated municipalities that had their shares of female councillors increased by the new law

[25] For this identification strategy to be valid, we need the instrument to influence the outcome only through its effect on the councillors' sex ratio (exclusion restriction assumption). We argue that this is the case: the new law was imposed on a set of municipalities irrespective of their expenditures' composition, and it did not change how the local administrators could spend public money, plus it did not affect their revenues. Moreover, it's hard to think that the new law generated any effect on the demand for daycare, since we saw how other variables such as the number of spots did not change. Thus, we argue that being subject to the gender quota did not alter directly the municipalities' expenditures for daycare, but it affected daycare only through its effect on female councillors.

which displays our first stage in the 2SLS setting.

Tab. 8: **First stage regression**

	(1)
VARIABLES	Female councillors
TreatPost	0.099***
	(0.007)
Municipality FE	YES
Year FE	YES
Observations	7,896
Number of municipalities	1,442
R-squared	0.347

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: First stage regression in the 2SLS setting. TreatPost corresponds to the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. The dependent variable is the share of female councilors in the municipalities. All 1442 municipalities belonging to our sample are included in the regression. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

As we can see, the introduction of the law increased on average the female presence in city councils across treated cities, which was indeed the main goal of Law215/2012. In Tab.9 we show the second stage results, together with OLS base results.

Tab. 9: **OLS and second stage regression**

VARIABLES	OLS Daycarepc	2SLS Daycarepc	2SLS Daycarepc
Female councillors	-0.105	-5.555***	-5.096*
	(0.734)	(1.928)	(2.806)
treatment	-2.180***		-2.057***
	(0.726)		(0.718)
Postquota	-0.215		0.346
	(0.204)		(0.372)
Observations	7,896	7,896	7,896
Number of municipalities	1,422	1,422	1,422
Muni FE	yes	yes	yes
Year FE	yes	yes	yes

Robust standard errors are clustered at the municipality level
*** p<0.01, ** p<0.05, * p<0.1

Note: Second stage regression in the 2SLS setting, together with OLS base results. Female councillors is instrumented with Treatpost, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. The dependent variable is the per capita expenditures on day care measured at the municipality level. All 1442 municipalities belonging to our sample are included in the regression. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

As we can see, the post-law increase in female councillors decreased on average day care expenditures across treated municipalities. On the other hand, the correlation between polarization and day care expenditures is almost null and non-significant in the OLS results: this is not particularly surprising, since this correlation takes into account also control municipalities and the period before the introduction of the quota. The magnitude of the last column's effect is larger with respect to our Diff-in-Disc results: the decrease in 2SLS results is of roughly 5 euros per capita, with respect to the 1,8 euros per capita in the Diff-in-Disc setting (municipalities with polarization levels higher than the median). There might be several reasons for this: first of all, Tab.11 presents a LATE effect, which regards a subset of treated municipalities, the compliers, whose share of female councillors was increased by the law: this effect might be different from the effect at the threshold observed in the Diff-in-Disc. Moreover, we include here all municipalities in our sample, and not exclusively those belonging to a narrow bandwidth around the cutoff. Plus, the 2SLS setting allows us to overcome possible biases due to endogeneity in the Female post-quota variable: it might be that these biases are shrinking the coefficients in the Diff-in-Disc framework. However, this last piece of evidence confirms that the law-induced increase in female councillors had on average negative effects on the funding for day care, in line with our theoretical framework.

7.4 Effect on councils' dissolution

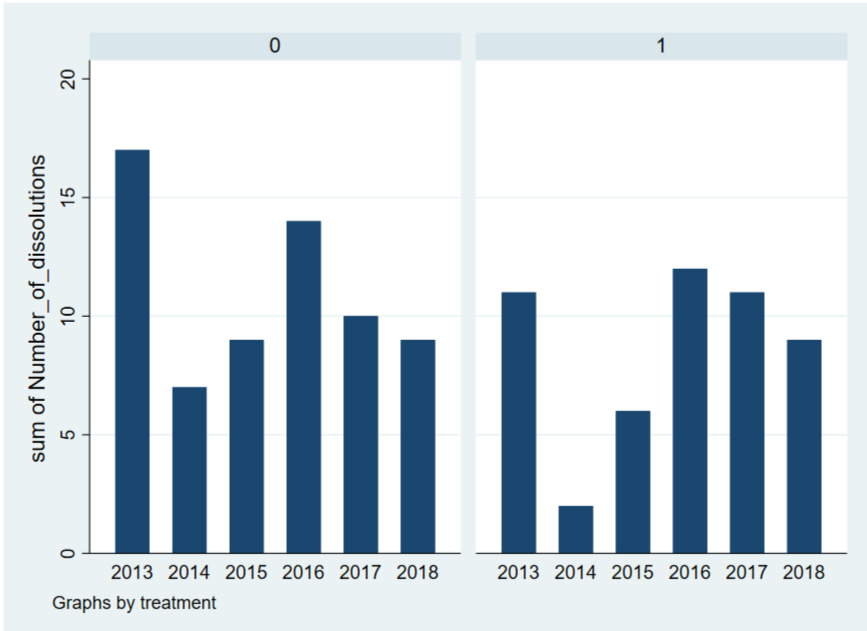
In Italy, municipal councils can be dissolved for different reasons:

- Persistent law violations and threats to public policy (e.g. mafia infiltrations)
- Mayor's forfeiture because of resignation, promotion to higher office or death
- Resignation of more than half of local councillors
- Passage of a nonconfidence motion[26]
- Prolonged financial distress

In case the council is dissolved, the municipality's local administrators are replaced with external commissioners, and local public finances are placed under supervised management until new elections are held. Some of these cases can signal a heavy divergence of opinions among local councillors: in this subsection, we focus on councils dissolved because of the resignation of more than half of the local councillors. This possibility happens mostly for persistent disagreement about the implementation of the political agenda. In Fig.6 we see trends in dissolutions for this specific reason over time, for treated and control municipalities belonging to our sample.

[26] The nonconfidence motion must be presented by at least 2/5 councillors and has to be approved by the majority of councillors in order to pass.

Fig. 6: **Number of dissolutions because of councillors' resignation**



Note: Number of local councils dissolved because of councillors' resignation by year, for treated (label 1) and control (label 0) municipalities. The whole sample of 1422 municipalities is included.

As we see, control municipalities get dissolved more frequently for this reason with respect to treated municipalities: however, yearly trends between the two groups of municipalities are similar[27]. In relative terms, the resignation of the majority of councillors is a rare event: the percentage of our sampled municipalities that experiences this kind of council dissolution is 3.8%. Nevertheless, these specific dissolutions can signal a conflict between local councillors and, in line with our theoretical reasoning, we might expect that this happened more frequently when the new law generated a larger presence of female councillors. Indeed, in our model, a larger minority presence is decreasing the probability of compromise between the two groups. In our empirical setting, this finding can translate into increased chances of conflict among councillors after the introduction of the gender quota. To assess whether Law 215/2012 raised the frequency of councils' dissolutions, we create a "Dissolution dummy" taking value 1 for the year when the local council was dissolved for the resignation of the majority of local councillors. The "Dissolution dummy" takes value 1 also for the previous

[27] Note that so far we excluded the year of dissolution for municipalities dissolved for any reason from our main analysis, since, as we pointed out, public finances are placed under a special regime in these cases.

years of the specific dissolved councils, since it identifies the future conflicting councils. This variable is the outcome of Tab.10, which shows new Difference in Discontinuity regressions with heterogeneous results aimed at verifying whether dissolutions were more frequent after the Law 215/2012 introduction.

Tab. 10: **Heterogeneous Law's effect on councils' dissolutions**

VARIABLES	(1) Dissolution	(2) Dissolution	(3) Dissolution	(4) Dissolution	(5) Dissolution
TreatPost	-0.098	0.010	-0.018	-0.002	0.000
	(0.062)	(0.026)	(0.028)	(0.027)	(0.027)
TreatPost*Twentyfifth p.		-0.081			
		(0.068)			
TreatPost*Fem. post quota	0.265*				
	(0.140)				
Twenty-fifth p.		-0.442			
		(1.388)			
TreatPost*Twenty-fifth p.		1.002			
		(1.738)			
Median			0.465		
			(0.353)		
TreatPost*Median			0.055***		
			(0.020)		
Seventy-fifth p.				0.465	
				(0.353)	
TreatPost*Seventy-fifth p.				0.028**	
				(0.013)	
Ninetieth p.					0.465
					(0.353)
TreatPost*Ninetieth p.					0.033**
					(0.014)
Observations	2,483	2,547	2,547	2,547	2,547
R-squared	0.037	0.059	0.062	0.053	0.053
Number of municipalities	515	528	528	528	528

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Difference in discontinuity regressions with council's dissolution because of majority of councillors' resignation as main outcome. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. Interaction terms indicate the post-quota share of elected female councillors, taken with respect to the overall distribution of treated cities' councils. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Data were constructed from the Ministry of Interior database. Robust standard errors in parentheses are clustered at the municipality level.

From Tab.10 we see that in councils where the Law's introduction generated a larger increase in female councillors, this led to a raised probability of council's dissolution compared to the control group: columns 1,3,4 and 5 confirm our theoretical prediction. If we interpret the dissolution as a signal of conflict, we can say that having a more balanced composition of male and female councillors increased the probability of political disagreement[28]. On the other hand, in councils with low levels of post-quota female politicians, this probability was not affected (column 2). To get an idea of the magnitude of the effect, for levels of post-quota female councillors higher than the median, the effect of the law was an increase of 3.7 percentage points in the probability of dissolution. This evidence hints at the possibility that the increase in female politicians leads to further conflict in some municipalities, not only by changing public expenditures but also ending up dissolving the local council. Again, this is a reduced form effect and we cannot exclude possible other effects leading to councils' dissolution. However, this other piece of evidence, together with the other presented ones, supports our major claim that the Law increased the probability of political conflict in some cities[29].

8 Conclusion

In the last decades, Western societies have been acting in order to increase the political power and representation of some social groups which were unable to influence the process of political decision-making. These measures, such as the extension of suffrage to Afro-Americans and the introduction of gender quotas, have been proven effective in benefiting these groups, in terms of granting them more social rights and access to welfare programs (Miller (2008), Fujiwara (2015), Chattopadhyay and Duflo (2004)). However, in this paper we highlight how

[28] This evidence is in line with Gagliarducci and Paserman (2012), who show how Italian municipalities with female mayors are more likely to be dissolved, and that this probability increases in male-dominated councils. Their evidence, in line with ours, suggests that an important determinant of group dynamics is its gender composition.

[29] We might wonder why these dissolutions did not generate ulterior inefficiencies on other kinds of expenditures apart from day care. On this matter, we cannot exclude that the political conflict extended to other kinds of expenditures, but given the large number of categories (over 100), it is possible that different conflicting councils clashed over different categories. In other words, it is hard to believe that all of these councils conflicted on the same expenditure category, apart from day care, given its gender-related connotation. Therefore, we do not observe a clear effect on sport or tourism, for instance, because it is not obvious that each conflicting council would raise the funding for these specific categories after the reduction in day care funding.

in the short run the increase of political representation of an under-represented social group might potentially generate unintended consequences, which could both benefit and damage the group. Indeed, increasing the representation of a minority can reduce the possibility of compromising with the majority, as the claims of the minority group might also be stronger and more distant from the majority's preferences. As we show in our theoretical model, the growth in the political minority group's share decreases the probability of reaching a mutual agreement on a policy for which the two groups have distant preferences. The growth of the minority's share decreases its costs of conflict and increases its demands in terms of policy outcomes: as a consequence, compromise becomes too costly for the majority which might decide to vote its bliss point as the policy level. This situation moves the final decision on the policy level to a more extreme outcome which is less preferred by the minority with respect to the outcome that would have been obtained with a compromise. The probability of compromise between groups is negatively related to the minority group's share, the salience of the issue and the difference in the two groups' bliss points with respect to the policy level. On the other hand, our model predicts that having to deal more frequently with the other group in the future raises the probability to avoid a conflict. We test some of these theoretical implications by exploiting the introduction of a gender quota in Italy in 2012, which substantially raised the share of female politicians at the local level. We assess the consequences of this shock in the political class composition on the level of investment for day care, a service for which men and women typically have different preferences (Barigozzi et al. (2019), Bhalotra and Clots-Figueras (2014)), and find that there were heterogeneous consequences with respect to the post-law share of female councillors. By the means of a Difference-in-Discontinuity strategy (Grembi et al. (2016)), we show that where the quota substantially raised the share of female politicians in local councils, there was a cut in expenditure for day care with respect to the cities not subject to the quota. On the other hand, with low shares of female councillors, the level of funding for this particular expenditure category was increased with respect to the control group. The empirical findings are in line with our theoretical model since the minority group's share influences the probability of compromise between the two groups of politicians. A series of robustness tests support the soundness of our results: we prove that there are no substantial differences between cities with high and low post-law shares of female councillors, and three different placebo tests reinforce our claims of causality. Moreover, through an instrumental variable analysis, we show that the negative effect on day care expenditures was indeed driven by the quota-induced increase in female councillors. Lastly, we show more evidence supporting our idea that an increase in the minority group size can fuel political disagreement in the short run. Indeed, the councils with larger post-law shares of female councillors were more likely to be dissolved in the years after elections with respect to councils with low scores in this variable. In conclusion, the short-run consequences of enhancing a minority's political representation can be more complex than expected. If we needed to suggest a policy implication, it would be to cautiously assess potential heterogeneity in preferences between politicians: giving more political power to a group facing another one with distant preferences could be detrimental in the short run. However, increasing the political power of minorities, as we said previously, grants consistent long-run benefits, thus we are not advocating a limit to their enfranchisement. More simply, we are suggesting to assess also possible short-run distortions that might be caused by sudden shocks in the composition of the political class.

References

Abhay P Aneja and Carlos F Avenancio-Leon. The effect of political power on labor market inequality: Evidence from the 1965 voting rights act. *Unpublished working paper*, 2019.

Manuel Bagues and Pamela Campa. Can gender quotas in candidate lists empower women? evidence from a regression discontinuity design. *Journal of Public Economics*, 194:104315, 2021.

Audinga Baltrunaite, Alessandra Casarico, Paola Profeta, and Giulia Savio. Let the voters choose women. *Journal of Public Economics*, 180:104085, 2019.

Francesca Barigozzi, Helmuth Cremer, and Chiara Monfardini. The gender gap in informal child care: theory and some evidence from italy. 2019.

Thushyanthan Baskaran and Zohal Hessami. Women in political bodies as policymakers. 2023.

Lori Beaman, Esther Duflo, Rohini Pande, and Petia Topalova. Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *science*, 335(6068):582–586, 2012.

Timothy Besley and Anne Case. Does electoral accountability affect economic policy choices? evidence from gubernatorial term limits. *The Quarterly Journal of Economics*, 110(3): 769–798, 1995.

Timothy Besley and Stephen Coate. An economic model of representative democracy. *The quarterly journal of economics*, 112(1):85–114, 1997.

Sonia Bhalotra and Irma Clots-Figueras. Health and the political agency of women. *American Economic Journal: Economic Policy*, 6(2):164–97, 2014.

Giampaolo Bonomi, Nicola Gennaioli, and Guido Tabellini. Identity, beliefs, and political conflict. *The Quarterly Journal of Economics*, 136(4):2371–2411, 2021.

Ylenia Brilli, Daniela Del Boca, and Chiara D Pronzato. Does child care availability play a role in maternal employment and children's development? evidence from italy. *Review of Economics of the Household*, 14(1):27–51, 2016.

Sebastian Calonico, Matias D Cattaneo, and Max H Farrell. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210, 2020.

Raghabendra Chattopadhyay and Esther Duflo. Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica*, 72(5):1409–1443, 2004.

Irma Clots-Figueras. Are female leaders good for education? evidence from india. *American Economic Journal: Applied Economics*, 4(1):212–44, 2012.

Katherine Baldiga Coffman. Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660, 2014.

Anthony Downs et al. An economic theory of democracy. 1957.

Joan Esteban and Debraj Ray. Conflict and distribution. *Journal of Economic Theory*, 87 (2):379–415, 1999.

Joan-Maria Esteban and Debraj Ray. On the measurement of polarization. *Econometrica: Journal of the Econometric Society*, pages 819–851, 1994.

Fernando Ferreira and Joseph Gyourko. Does gender matter for political leadership? the case of us mayors. *Journal of Public Economics*, 112:24–39, 2014.

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

Stefano Gagliarducci and Tommaso Nannicini. Do better paid politicians perform better? disentangling incentives from selection. *Journal of the European Economic Association*, 11(2):369–398, 2013.

Stefano Gagliarducci and M Daniele Paserman. Gender interactions within hierarchies: evidence from the political arena. *The Review of Economic Studies*, 79(3):1021–1052, 2012.

Andrew Gelman and Guido Imbens. Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456, 2019.

Benny Geys and Rune J Sørensen. The impact of women above the political glass ceiling: Evidence from a norwegian executive gender quota reform. *Electoral Studies*, 60:102050, 2019.

Veronica Grembi, Tommaso Nannicini, and Ugo Troiano. Do fiscal rules matter? *American Economic Journal: Applied Economics*, pages 1–30, 2016.

Gene M Grossman and Elhanan Helpman. Identity politics and trade policy. *The Review of Economic Studies*, 88(3):1101–1126, 2021.

Chris M Herbst. Universal child care, maternal employment, and children's long-run outcomes: Evidence from the us lanham act of 1940. *Journal of Labor Economics*, 35(2): 519–564, 2017.

Zohal Hessami and Thushyanthan Baskaran. Competitively elected women as policy makers. 2019.

Zohal Hessami and Mariana Lopes da Fonseca. Female political representation and substantive effects on policies: A literature review. *European Journal of Political Economy*, 63: 101896, 2020.

Grant Miller. Women's suffrage, political responsiveness, and child survival in american history. *The Quarterly Journal of Economics*, 123(3):1287–1327, 2008.

Caroline Slegten and Bruno Heyndels. Within-party sex gaps in expenditure preferences among flemish local politicians. *Politics & Gender*, 16(3):768–791, 2020.

Silvia Vannutelli. From lapdogs to watchdogs: Random auditor assignment and municipal fiscal performance in italy. *Job Market Paper: Yale University*, 2021.

Appendices

A Difference in Discontinuity: third validity assumption

For our identification strategy to correctly assess the true effect of the Law 215/2012 on day care expenditures, three validity assumptions need to be satisfied (Grembi et al. (2016)). In this subsection we discuss the third assumption, namely the absence of relevant interaction effects between the treatment and the other confounder present at the 5000 inhabitants' threshold, which is the increase in the mayor's salary. To check whether the assumption is satisfied, we perform a series of Difference in Discontinuity regressions interacting the treatment variable with a set of variables for each mayor's characteristic. Here, our reasoning is that the higher salary might influence the selection into politics and bring different mayors into charge who could interact diversely with the new councillors elected thanks to the gender quota. Thus, if our treatment effect is different because of this confounder, we should observe significant interaction effects between the treatment coefficient and mayors' characteristics. The mayor's characteristics we consider are the ones for which we have data on the Ministry of Interior's website: mayor's sex, education and age. In Tab.11 we check whether any of this interaction effects is relevant or present.

Tab. 11: **Interactions treatment coefficient and mayor's characteristics**

VARIABLES	(1) Daycarepc	(2) Daycarepc	(3) Daycarepc
TreatPost	-1.423	0.112	2.364
	(1.878)	(0.837)	(2.148)
Mayor_fem	0.224		
	(0.464)		
TreatPost*Mayor_fem	-2.150		
	(1.410)		
Mayor_age		-0.011	
		(0.020)	
TreatPost*Mayor_age		0.026	
		(0.032)	
Mayor_Ed			-0.066
			(0.072)
TreatPost*Mayor_Ed			-0.156
			(0.125)
Municipality FE	YES	YES	YES
	YES	YES	YES
Year FE	YES	YES	YES
	YES	YES	YES
Observations	3,345	3,409	3,354
R-squared	0.032	0.033	0.033
Number of municipalities	704	707	704

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

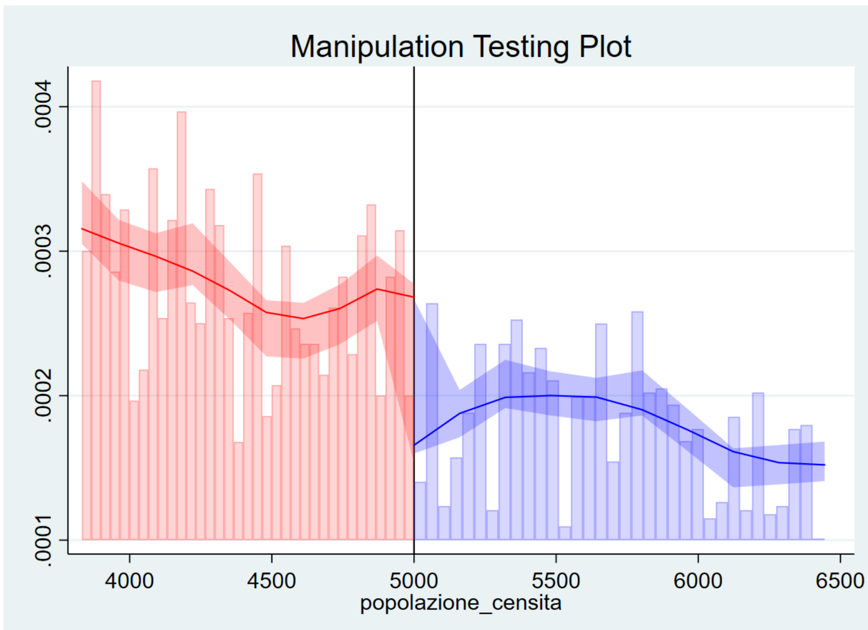
Note: Difference in discontinuity regressions. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. TreatPost is interacted with each one of the current mayor's characteristics, namely sex, age and education. The dependent variable is yearly per capita expenditures on day care related categories. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

As we can observe from Tab.13, there are no significant interaction effects between our treatment coefficient and each of the three mayor's characteristics we have data on, namely sex, age and education. We can conclude that the third validity assumption for our Difference in Discontinuity strategy is satisfied.

B McCrary test

We want to test whether there could have been manipulation in the running variable at the threshold of 5000 inhabitants. Fig.7 shows the McCrary test, which analyzes whether there is a significant difference in observational densities at the cutoff.

Fig. 7: **% McCrary test for manipulation of running variable**



Note: McCrary test assessing the manipulation of the running variable. The null hypothesis is that the two observational densities, at the right and left of the 5000 inhabitants threshold, are equal. P-value for this hypothesis is 0.95, not supporting the null hypothesis' rejection. Sampled municipalities are 1422 out of 8100 Italian municipalities, observed over the period between 2013 and 2018.

The McCrary test does not reject the null hypothesis of no manipulation in the running variable, thus we can conclude that there is no reason to believe that mayors distorted their cities' population to move it either above or below the threshold.

C Dynamic Difference in Discontinuity

The dynamic Difference-in-Discontinuity Vannutelli (2021) we exploit here interacts the dummy identifying the treatment group with each of the periods observed, 1/2/3/4 years before the first elections with the new law and 1/2/3/4 years after. This way, we are able both to observe the treatment effect in a dynamic way and to perform a placebo test, checking whether the pre-treatment periods present null treatment effects[30]. Also in this case, we focus on the reduced sample of municipalities belonging to the bandwidth computed before. Tab.12 presents the point estimates, while Fig.8 shows the Dynamic Difference in Discontinuity results graphically, to allow an easier understanding of this other result.

[30] In other words, we are empirically testing a parallel trends assumption

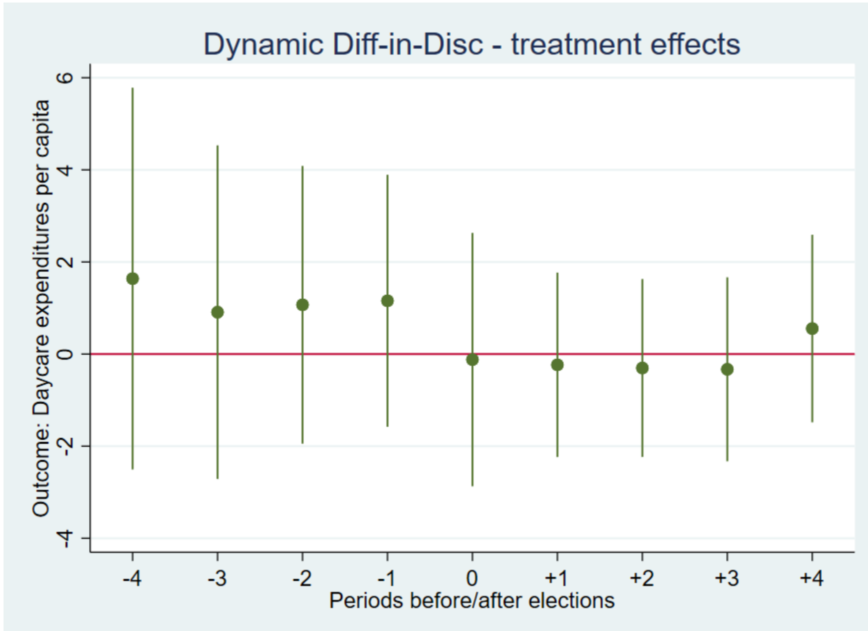
Tab. 12: **Dynamic Difference in Discontinuity results**

VARIABLES	(1) Daycarepc	(2) Daycarepc
Treated*4 years before	1.551	1.639
	(1.980)	(2.110)
Treated*3 years before	0.873	0.909
	(1.758)	(1.844)
Treated*2 years before	0.996	1.070
	(1.425)	(1.535)
Treated*1 year before	1.137	1.157
	(1.288)	(1.393)
Treated*Election year	-0.022	-0.120
	(1.303)	(1.400)
Treated*1 year after	-0.208	-0.233
	(0.949)	(1.020)
Treated*2 years after	-0.289	-0.303
	(0.918)	(0.983)
Treated*3 years after	-0.304	-0.331
	(0.960)	(1.017)
Treated*4 years after	00.516	0.554
	(0.984)	(1.037)
Mayor's controls	NO	YES
Municipality FE	YES	YES
Year FE	YES	YES
Observations	2,641	2,582
R-squared	0.022	0.025
Number of municipalities	576	573

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Dynamic Difference in Discontinuity results. The dynamic treatment effect was obtained by interacting the dummy *Treatment* with several dummies identifying the periods before and after the first election with Law 215/2012 in force. The dependent variable is yearly per capita expenditures on day care-related categories. Expenditures taken during election years are attributed to the council governing up to the election month. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

Fig. 8: **Dynamic Difference in Discontinuity results**



Note: Dynamic Difference in Discontinuity results. The coefficients' size and confidence intervals shown here correspond to the variables identifying the dynamic treatment effect, obtained by interacting the dummy $Treatment$ with several dummies identifying the periods before and after the first election with Law 215/2012 in force. The dependent variable is yearly per capita expenditures on day care-related categories. Expenditures taken during election years are attributed to the council governing up to the election month. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression.

Looking at both Tab.12 and Fig.8, we can conclude that the aggregate treatment effect has not been significantly different than zero, even if we look at it in a dynamic sense. The results we see from the Dynamic Difference in Discontinuity also exclude differences between treated and control municipalities in political budget cycles: even graphically, we do not observe recurring paths in the periods before or after the elections. This absence of differences in spending cycles contributes to supporting the idea of comparability between treated and control municipalities.

D Different level of pre-quota female councillors

A potential concern we need to discuss comes from the fact that the heterogeneity that we found might depend on pre-quota levels of female councillors, which in turn might be driven by other municipalities' characteristics also linked to day care services. Even if this particular heterogeneity should be captured by municipality's fixed effects, we perform an additional robustness and introduce the pre-quota level of female councillors as a control in our Difference in Discontinuity regressions, to exclude this potential confounder for our analysis. Tab.13 presents this additional evidence.

Tab. 13: **Difference in Discontinuity heterogeneous results**

VARIABLES	Baseline interaction Daycarepc	25th percentile Daycarepc	median Daycarepc	75th percentile Daycarepc	90th percentile Daycarepc
TreatPost	4.014	-0.831	0.075	-0.116	-0.482
	(2.559)	(1.068)	(1.065)	(1.040)	(1.033)
TreatPost*Fem. post quota	-12.242*				
	(6.549)				
Twenty-fifth p.		5.244***			
		(1.273)			
TreatPost*Twenty-fifth p.		1.875*			
		(1.132)			
Median			-5.190***		
			(1.350)		
TreatPost*Median			-1.798*		
			(1.033)		
Seventy-fifth p.				0.469	
				(2.698)	
TreatPost*Seventy-fifth p.				-3.584	
				(2.576)	
Ninetieth p.					-0.271
					(2.503)
TreatPost*Ninetieth p.					-2.426
					(2.264)
Fem. pre-quota		17.977**	23.362***	9.252	10.798
		(8.985)	(8.114)	(10.984)	(10.831)
Observations	2,382	2,446	2,446	2,446	2,446
R-squared	0.036	0.034	0.037	0.039	0.034
Number of municipality1	544	559	559	559	559
Mayor controls	YES	YES	YES	YES	YES
Municipality FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Robust standard errors in parentheses are clustered at the municipality level
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Difference in discontinuity regressions. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. Interaction terms indicate the post-quota share of elected female councillors, taken with respect to the overall distribution of treated cities' councils. The dependent variable is yearly per capita expenditures on day care related categories. Fem. pre-quota indicates the share of female councillors before the first elections with Law 215/2012 in force. Expenditures taken during election years are attributed to the council governing up to the elections month. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

What we see from Tab.13 is that our main coefficients are not modified in magnitude or significance level. Therefore, the inclusion of the share of pre-quota female councillors confirms that this variable is potentially increasing the provision of day care across municipalities but also that it should not be a source of bias for our results.

E Placebo tests - artificial cutoffs/year of implementation

In this section, we conduct two additional placebo analyses to support the soundness of our result and corroborate our thesis that Law 215/2012 had these heterogeneous effects with respect to city councils' post quota shares of female councillors. For the first placebo analysis we create two subsamples with only treated or control municipalities and artificially move the threshold of enforcement for the law to either 4000 or 7000 inhabitants. Then, we verify the absence of significant effects on day care-related expenditures. In addition, we perform another placebo test by moving the law's implementation year to 2015 and checking the resulting treatment effects[31]. The first placebo test's results are shown in Tab.14 and 15: in the first Table, the sample is composed of only control municipalities and the fictional cutoff is at 4000 inhabitants, while in the second Table, we have only treated municipalities with a 7000 inhabitants cutoff.

[31] Since it might be argued that these fictional cutoff and year are chosen arbitrarily, we performed additional placebos and observed null heterogeneous effects also with other "fake" cutoffs/years, such as 6000 inhabitants or 2014-2016.

Tab. 14: **Placebo test with artificial population cutoff - control municipalities**

VARIABLES	Baseline interaction Daycarepc	25th percentile Daycarepc	median Daycarepc	75th percentile Daycarepc	90th percentile Daycarepc
TreatPost		1.592	0.948	1.404	1.075
0.880					
	(1.874)	(1.045)	(1.304)	(1.150)	(1.122)
TreatPost*Fem. post quota	-1.102				
	(4.595)				
Twenty-fifth p.		5.570***			
		(0.326)			
TreatPost*Twenty-fifth p.		1.425			
		(1.688)			
Median			-3.304*		
			(1.987)		
TreatPost*Median			-0.474		
			(0.836)		
Seventy-fifth p.				-3.313*	
				(1.985)	
TreatPost*Seventy-fifth p.				0.173	
				(0.972)	
Ninetieth p.					-3.317*
					(1.984)
TreatPost*Ninetieth p.					3.111**
					(1.487)
Observations	1,594	1,655	1,655	1,655	1,655
R-squared	0.020	0.022	0.020	0.020	0.024
Number of municipalities	392	419	419	419	419
Mayor controls	YES	YES	YES	YES	YES
Municipality FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Robust standard errors in parentheses are clustered at the municipality level
*** p<0.01, ** p<0.05, * p<0.1

Note: Difference in discontinuity regressions with artificial cutoff for Law 215/2012 implementation. In this case, the sample includes only municipalities with less than 5000 inhabitants and the cutoff has been moved to 4000 inhabitants. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. Interaction terms indicate the post-quota share of elected female councillors, taken with respect to the overall distribution of treated cities' councils. The dependent variable is yearly per capita expenditures on day care-related categories. Expenditures taken during election years are attributed to the council governing up to the elections' month. The bandwidth around the artificial 4000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

Tab. 15: **Placebo test with artificial population cutoff - treatment municipalities**

VARIABLES	Baseline interaction Daycarepc	25th percentile Daycarepc	median Daycarepc	75th percentile Daycarepc	90th percentile Daycarepc
TreatPost	-3.127	0.004	-1.421	-0.800	-0.438
	(2.104)	(1.210)	(0.985)	(1.169)	(1.145)
TreatPost*Fem. post quota	7.332				
	(5.795)				
TreatPost*Twenty-fifth p.		-2.168			
		(1.339)			
Median			-3.840***		
			(1.454)		
TreatPost*Median			1.746		
			(1.096)		
Seventy-fifth p.				-1.651	
				(1.537)	
TreatPost*Seventy-fifth p.				1.178	
				(1.220)	
Ninetieth p.					-0.816
					(0.845)
TreatPost*Ninetieth p.					-0.277
					(1.496)
Observations	968	993	993	993	993
R-squared	0.037	0.037	0.038	0.035	0.033
Number of municipalities	243	249	249	249	249
Mayor controls	YES	YES	YES	YES	YES
Municipality FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Robust standard errors in parentheses are clustered at the municipality level
*** p<0.01, ** p<0.05, * p<0.1

Note: Difference in discontinuity regressions with artificial cutoff for Law 215/2012 implementation. In this case, the sample includes only municipalities with more than 5000 inhabitants and the cutoff has been moved to 7000 inhabitants. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. Interaction terms indicate the post-quota share of elected female councillors, taken with respect to the overall distribution of treated cities' councils. The dependent variable is yearly per capita expenditures on day care-related categories. Expenditures taken during election years are attributed to the council governing up to the elections' month. The bandwidth around the artificial 7000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. The dummy for the twenty-fifth percentile has been dropped due to multicollinearity. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

Results from Tables 14 and 15 present no evidence of significant heterogeneous effect with respect to these “fake” population thresholds, supporting our claim that the results we observe are due to the new legislation introduced.

In the second placebo test, instead of an artificial population threshold, we move the implementation of the law to year 2015, thus considering as “post treatment” period the years 2016, 2017 and 2018. Tab.16 shows this second placebo's results.

Tab. 16: **Placebo test with artificial year of law's introduction**

VARIABLES	Baseline interaction Daycarepc	25th percentile Daycarepc	median Daycarepc	75th percentile Daycarepc	90th percentile Daycarepc
TreatPost	0.734	0.013	0.300	0.557	0.451
	(3.454)	(1.329)	(1.339)	(1.259)	(1.270)
TreatPost*Fem. post quota	-7.892				
	(8.598)				
Twenty-fifth p.		-0.442			
		(1.388)			
TreatPost*Twenty-fifth p.		1.002			
		(1.738)			
Median			1.150		
			(1.178)		
TreatPost*Median			-0.452		
			(0.971)		
Seventy-fifth p.				1.300	
				(1.075)	
TreatPost*Seventy-fifth p.				-1.316	
				(0.984)	
Ninetieth p.					1.279
					(1.071)
TreatPost*Ninetieth p.					-1.187
					(1.025)
Observations	926	2,441	2,441	2,441	2,441
R-squared	0.065	0.031	0.031	0.033	0.032
Number of municipalities	515	528	528	528	528
Mayor controls	YES	YES	YES	YES	YES
Municipality FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES

Robust standard errors in parentheses are clustered at the municipality level
*** p<0.01, ** p<0.05, * p<0.1

Note: Difference in discontinuity regressions with artificial year of introduction for Law 215/2012. In this case, we consider as councils governing under the new legislation those elected after 2015. TreatPost corresponds to the treatment coefficient, the interaction between the dummy identifying the treatment group and the dummy for the post-treatment period. Interaction terms indicate the post-quota share of elected female councillors, taken with respect to the overall distribution of treated cities' councils. The dependent variable is yearly per capita expenditures on day care related categories. Expenditures taken on election years are attributed to the council governing up to the elections' month. The bandwidth around the 5000 inhabitants' threshold is computed following one of the procedures described by Calonico et al. (2020), through the minimization of the Mean Squared Error. Year and Time fixed effects are included in each regression. Robust standard errors in parentheses are clustered at the municipality level.

Also this second placebo presents non-significant coefficients for the interaction terms and the treatment variable.

Marriage patterns and the gender gap in labor force participation: evidence from Italy

Giovanni Righetto[1]

April 20, 2023

Abstract

The Italian rate of gender participation gap, defined as the differential between female and male rates of labor force participation, was 18.2% in 2020, the second highest among EU countries. In this paper, we present evidence highlighting a new possible determinant of this unbalance in the labor force: endogamy intensity. We define endogamy as "marriage within the community", and we argue that it helps preserve and reinforce social norms stigmatizing working women, along with reducing the probability of divorce, which in turn disincentivizes women's participation in the labor force. We proxy the endogamy rate of a community by the degree of concentration of its surnames' distribution, and we provide evidence that a more intense custom of endogamy contributed to enlarging gender participation gaps across Italian municipalities in 2001. In order to deal with endogeneity issues, we make use of an instrumental variable strategy, by instrumenting the endogamy measure of a municipality by the degree of ruggedness of its territory: the asperity of a municipality's surface indeed contributes to its geographical isolation, thus incentivizing in-marriage. In our main 2SLS result, a standard deviation increase in our proxy of endogamy is linked to roughly a 0.3 standard deviation increase in the gender participation gap of 2001. In addition, we provide evidence supporting our main hypothesis, documenting how higher rates of in-marriage are linked to the preservation of social norms and to greater marriage stability, with a lower probability of divorce.

JEL classification numbers: A13, J21, Z13
Keywords: Endogamy - Female labor - Italy - Social norms

---

[1] PhD Student in Economics, University of Bologna, E-mail: giovanni.righetto3@unibo.it.
I would like to thank Paolo Vanin, Maria Bigoni, Vincenzo Scrutinio and Francesca Barigozzi for the helpful comments and suggestions that helped improve my paper. Moreover, I would like to express my gratitude to Paolo Buonanno and Erik Kimbrough for granting me access to their data on Italian surnames and Consanguinity. I am also grateful to the participants in all conferences and seminars I attended for all the comments and constructive feedback. In addition, I acknowledge the financial support of the University of Bologna.

1 Introduction

When compared to other EU countries, Italy is still lagging in the integration of women into the labor market. According to Istat, the gender participation gap, defined as the difference between ratios of men and women of working age that participate in the labor force, was 18,2% in 2020, the second highest value in the European Union (after Malta). This gap has been declining in the last decades, with a decrease of 16,6% since 1971 (always according to Istat data), but it still remains a complicated and yet unsolved issue characterizing this country.

The factors that led the peninsula to this situation of deep gender inequality are complex and multifaceted, from discouraging characteristics of the labor demand (such as the gender wage gap) to cultural reasons involving strict and long-standing social norms. We focus here on explaining Italian contemporaneous rates of the gender participation gap by looking at a new possible determinant: endogamy rates. We define endogamy as "marriage within the community", so a marriage between two individuals that belong to the same social group, live in the same location, and share the same traditions and customs. We argue that endogamy generates two mechanisms limiting a community's female labor supply. From a first perspective, endogamy increases communities' social isolation, which helps preserve and reinforce traditional social norms and a more conservative role of the woman inside the household. In addition, endogamous unions are characterized by higher costs of dissolution with respect to exogamous ones, given the stronger social ties of the former communities. As a consequence, the probability of divorce is lower and this decreases incentives for wives to look for jobs, given the reduced need to have an "outside option" providing economic stability in case the marriage ends[2]. Keeping in mind these two mechanisms, our prediction is that if a custom of in-marriage is persistent in an Italian municipality (which is identified as a community in our analysis), it will exhibit lower contemporaneous levels of female labor

[2] see Fernández and Wong (2014), Gray (1998) for a more complete explanation of the link between marital instability and female labor supply.

force participation and larger gender participation gaps with respect to more "exogamous" municipalities. When we focus on how to technically evaluate this prediction, we encounter two identification issues: first, how to measure the intensity of endogamous marriages within a community, and second, to what extent we are able to detect a causal effect of endogamy on the gender participation gap. We decide to measure the past intensity of endogamy within a community through the diversity of its residents' surnames, exploiting the indexes by Buonanno and Vanin (2017). To address endogeneity concerns, namely omitted variable bias and reverse causality, we make use of an instrumental variable strategy: municipalities' intensity of in-marriage is instrumented by the degree of ruggedness of its territory. Here, the underlying logic is that the geographic fragmentation of a municipality's surface contributes to its isolation and augments the frequency of in-marriage, as it is confirmed by our first-stage regressions. The main concern of using this instrument is the validity of the exclusion restriction, since it has been demonstrated that ruggedness has direct and negative effects on economic activity (Nunn and Puga (2012)) and it could therefore generate a scarcity of jobs. To support the validity of our instrumental variable strategy, we extensively discuss the inclusion of controls in our specification preventing ruggedness from directly impacting our outcome of interest. In addition, in Appendix D we present a placebo analysis supporting the validity of the exclusion restriction assumption. However, even after controlling for observables and performing this placebo, ruggedness could still be suspected to be an imperfect instrument: this is why in Appendix E we show that the effect of endogamy on the gender participation gap is robust even if we relax the exclusion restriction assumption, following the approach suggested by Nevo and Rosen (2012)[3].

Our main results show how the past intensity of endogamy within a community contributes to enlarging gender participation gaps across Italian municipalities in 2001. These results, confirmed by a series of robustness checks, shed light on a new perspective linking marriage patterns and female labor supply, providing a new explanation for the gender-unbalanced

[3] In few words, the approach consists in allowing some degree of correlation between the instrument and the error term in the IV framework.

Italian labor force.

The paper is organized as follows: in Section 2 we discuss the related literature, with the contribution given by our work. In Section 3 we illustrate the problematic integration of women into the Italian labor force and explain how endogamy could have contributed to generating the current large gender participation gap in the peninsula. Section 4 describes our data and identification strategy, with a comprehensive discussion of the instrumental variable validity. Section 5 focuses on the main empirical results with the link between endogamy and the gender participation gap, and Section 6 discusses the channels that could have generated these results. We then perform a series of robustness checks in Section 7, confirming the soundness of the main results. Finally, section 8 concludes.

2 Related literature

When we focus on the determinants of female labor force participation, the current literature offers a multifaceted picture, identifying factors such as the characteristics of the labor market (Olivetti and Petrongolo (2016), Attanasio et al. (2008)), going through the availability of childcare (Attanasio et al. (2008), Carta and Rizzica (2018)) but also medical advances such as the invention of the contraceptive pill (Bailey (2006), Goldin and Katz (2002)) or in general improved maternal health conditions (Albanesi and Olivetti (2016)). However, these factors offer just a partial picture, given that the literature explained how women's decision not to work is frequently dictated by social norms and how their role inside the society is perceived (Fernandez and Fogli (2009), Fernández (2013), Bertrand et al. (2015)). These norms are enforced by the community in which an individual lives and impose a standardized model of behavior for a social category. In the case of women, this model usually includes the chores of housekeeping and childcare: if a woman does not adapt to this "archetype", she suffers a cost in terms of social stigma by the community (Akerlof and Kranton (2000)) or even inner sense of guilt (Barigozzi et al. (2018)). The influence of neighbors has been proven to be fundamental in the transmission of norms to young women (Maurin and Moschion (2009), Cavapozzi et al. (2021)), who frequently base their decision to work or not also on what they learn from their peers and parents. The literature presents evidence proving that the origins of gender roles could be determined by ancient agricultural practices: specifically, descendants of societies that traditionally used the plough exhibit lower rates of female participation into work, politics and entrepreneurship today (Alesina et al. (2013)). In the case of our project, we contribute to the understanding of why specific social norms survive in some places while they disappear in others, identifying a new possible determinant of their persistence. In addition to the studies on female labor, our research aims to contribute also to the literature about the effects of endogamy. This literature has been emphasizing how endogamy, and in particular consanguineous marriages, can be detrimental to the birth of inclusive institutions, contributing to generating more hierarchical societies (Greif and Tabellini (2017)). In line with our theoretical framework, in-marriage rates raise individuals' sense of belonging to their kinship or community, discouraging impartial cooperation and penalizing interaction with strangers (Akbari et al. (2019)). In addition, higher consanguineous marriage rates have been linked to lower current levels of generalized trust, impersonal cooperation and individualism, consequently hindering economic cooperation and development (Schulz et al. (2019)). Moreover, endogamy has been linked to higher levels of corruption (Akbari et al. (2019)) and lower women's enfranchisement in India (Bahrami-Rad (2021)). However, the connection between the intensity of in-marriage within a community and the characteristics of its labor force still constitutes a gap in this growing literature: this work aims at providing evidence of the link between endogamy and the gender participation gap.

3 Female labor in Italy and the role of endogamy

Given its strong and deeply-rooted Catholic origins, Italy has, more than other developed countries, maintained a more traditional familiar model, in which the mother is cherished as a symbol of matronly warmth and nurturing. The idea of the Italian mother has "turned into an international stereotype - a strong, capable woman who spends her days cooking for her children" (Kovick (2021)). Speaking of female empowerment in Italy, the evolution of the Italian law system suggests deep-rooted negative prejudices about enfranchised women (Passaniti (2011)): for instance, until 1919 the Italian Civil Code imposed on wives wishing to find an occupation the need to present a written marital authorization. In addition, until 1975 the Civil Code stated that "the husband is the head of the family, and the wife must support his decisions" (art.144). It was not until 1945 that all women were granted the right to vote in national elections and, while the right to divorce was not introduced until 1970, abortion was not legally possible until 1978. All these anecdotes suggest the existence of specific social norms with a precise role of the woman inside the households, incompatible with high-profile career jobs. It seems that in Italy those social norms still survive in some places but were overcome in others: especially in the Southern Italian regions, the role of women is still that of raising children and taking care of house chores. In fact, according to data from Swimez ("Associazione per lo sviluppo dell'industria nel Mezzogiorno"), the percentage of children aged 0-3 that benefited from public childcare was 5% for Southern regions, while it was 17,8% for Center-Northern Italian regions (SVIMEZ). Furthermore, the average per capita public spending for social services (such as disabled people assistance) was three times higher for Northern regions with respect to Southern ones (155 against 52 euros for the averages at the municipality level). Clearly, those data reflect also the well-known underdevelopment and low quality of institutions in Southern regions (Putnam et al. (1994)), but there is also a cultural component that contributes to maintaining the central role of women in taking care of the weaker members of the family. Thus, we may wonder what is preserving such a traditional model inside those households: this project offers a new explanation, identifying the intensity of endogamy within a municipality as a possible determinant of the conservation and enforcement of traditional social norms, consequently lowering local levels of female labor supply. As we anticipated previously, we define endogamy as "marriage within the community", so a marriage between two people that belong to the same social group, live in the same location and share the same traditions and customs, for instance speaking the same dialect. An example of an endogamous union is represented by a consanguineous marriage, that is matrimony between two people from the same family: however, here we intend endogamy in a broader sense, referring to people united by a common heritage of customs and usages and who live spatially close to each other. It is obviously complex to give a precise definition of community and to identify its social borders: therefore, in our analysis we identify communities with the 8000 Italian municipalities[4]. We argue that endogamy creates a closed system that helps preserve and reinforce traditional social norms and a more conservative role of the woman inside the household. Indeed, when a community is characterized by a high intensity of in-marriage and there is no influx of new individuals bringing different social norms, costumes and traditions are more easily maintained, and consequently traditional social norms are somehow sheltered. As Akbari et al. (2019) demonstrate, communities with higher degrees of endogamy tend to preserve a more traditional and kinship-based structure, discouraging cooperation with non-members and therefore the influx of new social norms. In a similar way, we can think of a parallelism with the Indian caste-based system, which is based on endogamous unions to preserve traditions and customs, with a gender-based system of punishment for those who marry outside the caste (Bidner and Eswaran (2015)). Moreover, endogamy has been linked to more unequal gender norms and a higher acceptance of domestic violence across African communities (Alesina et al. (2021). Here, our point is that Italian communities that experienced higher rates of endogamy maintained a more conservative view of the woman inside the family, hence keeping a larger gender participation gap with respect to more exogamous communities [5]. In our analysis, the past levels of in-marriage within a municipality proxy both the degree of conservation of the traditional gender roles in the

[4] The approximation of identifying communities with municipalities could be controversial, since municipalities with larger population might clearly include multiple communities. In Section 7 we thus repeat our analysis but on a subsample of municipalities with less than 5000 inhabitants, that are more likely to be associated with distinct communities.

[5] We focus on the gender gap and not only on the female labor supply since the former variable is a more reasonable indicator of the gender inequality brought by traditional social norms

community and the intensity of enforcement of the social norms that a woman living in the community must observe. The more those gender norms are maintained and enforced, the costlier it is for a woman to default, given the dynamics of social stigma and sense of guilt that we mentioned previously, and the more likely it is that she aligns with the traditional model of housewife. Thus, we argue that a stronger custom of in-marriage across Italian municipalities is associated with the preservation of gender norms stigmatizing working women[6]

In addition to the maintenance and enforcement of social norms, endogamy creates another indirect mechanism reducing female labor supply, that is through the reduction of the probability of divorce. According to pre-existing evidence, divorce rates have been demonstrated to be lower for endogamous unions (Houseworth and Chiswick (2020), Davenport (2016)). This link is not exclusive to developing countries, but it seems to be present also in developed ones such as Netherlands or Sweden, with marriages between locals and foreigners being more likely to end up in divorce (Smith et al. (2012)), and with higher risks of separation for spouses with larger cultural distances (Dribe and Lundh (2012)). We argue that endogamous unions are characterized by higher costs of dissolution with respect to exogamous ones, because of the heavier social stigma on couples that separate. This is because divorcing implies breaking the ties between the couple and its community and exiting from the closed system that endogamy preserves. A community that has been marrying "inside" for decades strengthens the social bonding between its members, who feel a powerful sense of affinity between themselves. Divorcing from a community member would fracture the community ties, and thus it is stigmatized by all other members of the social group who feel "betrayed". Therefore, given this higher cost of separation for endogamous couples, we argue that the probability of divorce is lower with respect to non-endogamous ones. As a consequence, the wife is less incentivized to look for an outside option to provide economic stability in case the marriage ends, given the higher solidity of endogamous marriages. In fact, there is evidence

[6] It might be argued that this mechanism may also contribute to maintaining a high level of female labor supply. In fact, since endogamy preserves the traditional social norms of society, if the community's norm dictates that women should participate in the labor force, a larger custom of endogamy may help reinforce and maintain this norm. However, as anecdotal evidence suggests, in Italy conservative social norms stigmatize working women and thus this is the custom most likely to be preserved by endogamy.

that a higher probability of divorce and marital instability can generate an increase in female labor force participation (Fernández and Wong (2014), Gray (1998)). Hence, we argue that if a custom of in-marriage is persistent in a community, women will be less incentivized to participate in the labor force also by reason of the lower probability of divorce. We analyze the channels linking endogamy and the gender participation gap in Section 6.

4 Data and identification strategy

To evaluate the major argument of the paper, we make use of a model with the following functional form:

$$Partgap_i = \alpha + \beta Endogamy_i + \gamma X_i + \epsilon_i$$

With:

- $Partgap_i$ corresponding to the rate of gender participation gap in municipality i, that is the difference between rates of men and women aged 15-64 who joined the labor force (in 2001);
- $Endogamy_i$ measuring the intensity of in-marriage within the municipality i;
- X_i set of geographic and demographic controls for municipality i

The core of this project is developed around the estimation of β, and the assessment of the effect of in-marriage rates on our outcome. In this section, we first describe the characteristics of our outcome and treatment, and then we proceed by explaining more in detail the identification strategy of the project.

Starting from our outcome, official statistics available from ISTAT indicate that the Italian gender participation gap was 18.2% in the second quarter of 2020, the second highest among the EU countries[7]. In Fig.1 we can see on the left a graph showing our outcome, the

[7] We choose to focus on the gender participation gap and not exclusively on female labor force participation because the main interest of this project relies on understanding the gender inequality characterizing the Italian labor force. Indeed, low rates of female labor supply alone do not identify a gender-unbalanced labor force, while a large gender participation gap does.

differential between participation rates of men and women in 2001, also known as "gender participation gap". To give an idea of the low involvement of women in the Italian labor force, on the right we present also the "Female Labor Force Participation" (percentages of women aged 15-64 participating in the labor force).

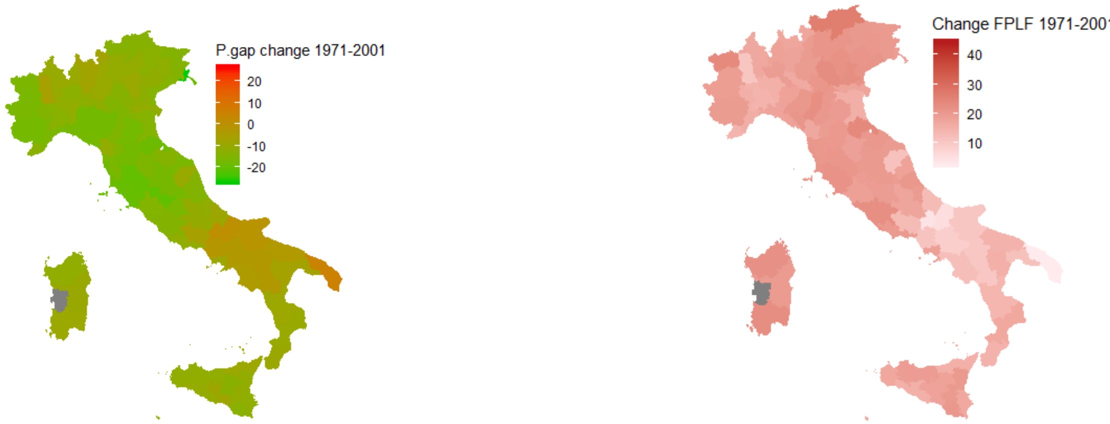
Fig. 1: **Labor force in Italy - 2001**

(a) Gender Participation Gap (b) Female Labor Force Participation

Notes: Fig. 1a shows the gender participation gap, namely the difference between rates of working-age men (15-64 years old) participating in the labor force and working-age women participating in the labor force. Fig. 1b shows exclusively the rates of working-age women participating in the labor force. Data are according to the 2001 census and are aggregated at the province level to provide an easier eye-catcher of the characteristics of our outcome. Data include all 107 Italian provinces in 2001.

Especially the southern Italian regions, historically characterized by lower institutional quality (Putnam et al. (1994)), have lagged in reducing this kind of gender inequality. The gender participation gap reaches impressive peaks of over 30% in the southern provinces of Foggia and Caltanissetta, that present the worst gap among the 110 Italian provinces. To grasp the idea of the persistence of this kind of inequality across Italian provinces, in Figure 2 we show how these two variables changed between 1971 and 2001:

Fig. 2: **Variations Part. gap, Female Labor Force Participation**



(a) Δ Gender Gap 1971-2001 (b) Δ FLFP 1971-2001

Notes: Fig. 2a shows variations in gender participation gaps between 1971 and 2001, with greener areas representing a reduction of the gap and reddish areas standing for an increase. Fig. 2b shows variations in rates of working-age women participating in the labor force between 1971 and 2001. Data are according to the censuses of 1971 and 2001 and are aggregated at the province level. Data include all 107 Italian provinces in 2001.

On the left, we present the changes in the gender participation gap for those 30 years, with the green areas indicating a decrease while the red areas represent an increase in this gap. We can observe again smaller reductions in southern Italy, with some provinces even exhibiting an increase in the gender participation gap. On the right, we can see the change in rates of women participating in the labor force between 1971 and 2001, and we can notice a positive trend everywhere, with higher variations in the Center-North provinces. As we can notice, there is a high persistence of the low rates of female labor force participation in some provinces: we argue that the survival of traditional gender roles is a key factor in explaining this persistence. In line with what we presented in the previous sections, our point is that the frequency of in-marriage has been a determinant of the survival of conservative social norms, leading to the permanence of the scarce levels of female labor supply in some parts of the peninsula. Indeed, we could think of the variable endogamy as a proxy of the strength of maintenance and enforcement of conservative gender roles.

When we aim to identify the relationship between endogamy and the gender participation gap, the first question we need to address is how to measure these variables. While for rates of men and women joining the labor force we have official data from the census of 2001, determining the frequency of in-marriage is trickier. We decided to proxy the intensity of endogamy at the municipality level with the surnames' distribution, from the work of Buonanno and Vanin (2017). For each municipality, they computed how frequent a specific surname is within the resident population, by looking at the percentages of residents with the same surname. The data source exploited to construct surnames' distributions was the national telephone directory, which virtually covered all households for the two years of focus, namely 1993 and 2004[8]. In particular, they offered two different, and somehow opposite, measures of surnames' concentration:

1. *Entropy* (available for 1993). This measure is directly proportional to the diversity of surnames in a specific municipality, and has the following specification:

   $$Entropy = -\sum_{i=1}^{S}(p_i log p_i)$$

   With S total number of surnames in the municipality of interest and p_i the percentage of people with the surname i. Therefore, the more variable surnames are in a municipality, the higher is the value of *Entropy*.

2. *First share* (available for 1993 and 2004). This variable quantifies the percentage of residents with the most common surname in the municipality, thus it estimates the extent of concentration in surnames' distribution.

Those two measures are opposite in the sense that *Entropy* quantifies the spread in surnames' distribution, while $Firstshare$ its concentration. Moreover, their main difference is that *Entropy* takes into account all surnames in a specific municipality, while $Firstshare$ only the most frequent one. The link with endogamy is simple: the diversity of surnames reflects a community's recent history of migration and in-marriage. When a community is more

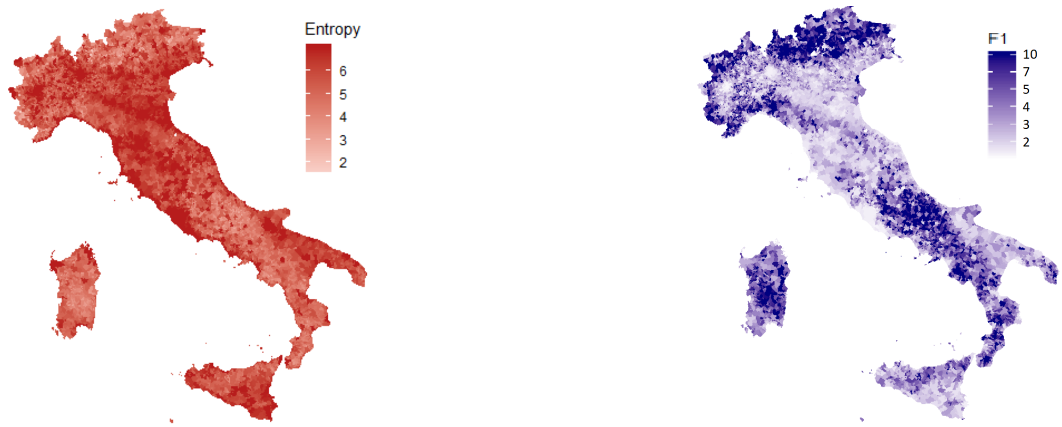
[8] We chose to use data from the 2001 census and not the most recent one from 2011 because of the higher temporal proximity to our treatment variables. Indeed, using data from 2011 might increase the worry of possible migration waves affecting the results of our analysis.

socially isolated, its members marry between each other more frequently and thus the distribution of surnames shrinks [9]. Surnames' diversity can tell many things about a community, from genetic characteristics to the degree of inbreeding (Crow and Mange (1965)). In particular, since surnames among patrilinear societies are transmitted from father to children, the distribution of surnames is similar to that of any social or cultural trait (Cavalli-Sforza et al. (2004), Darlu et al. (2012), Cavalli-Sforza and Feldman (1981)), and at the same time to the spread of the neutral alleles of a gene inherited exclusively through the Y-chromosome (Yasuda and Furusho (1971), Yasuda et al. (1974)). In our specific case, we argue that these two measures of surnames distribution capture the past intensity of endogamy within a community: a municipality that exhibited a tendency towards in-marriage would be characterized by a smaller *Entropy* and a larger $Firstshare$. Indeed, two individuals having the same surname are more likely to belong to the same community or social group: a higher homogeneity of surnames within a city signals that people have been marrying inside the community more frequently, reducing surnames' distribution. Therefore, in our empirical analysis we consider endogamous a marriage whose progeny is not expanding the surnames' distribution of a municipality, while a non-endogamous marriage is potentially enlarging this distribution[10]. In Fig. 3 we can see the distribution of those two measures across Italy, with *Entropy* in 1993 on the left and $Firstshare$ in 2004 on the right.

[9] Since the unification of Italy in 1861, laws concerning the inheritance of surnames have remained unchanged, with sons and daughters inheriting the father's surname.

[10] An example of a non-endogamous marriage in our case is that of a woman from a municipality marrying an individual from another municipality, whose surname is not present in the woman's municipality.

Fig. 3: **Entropy in 1993 and First share in 2004**



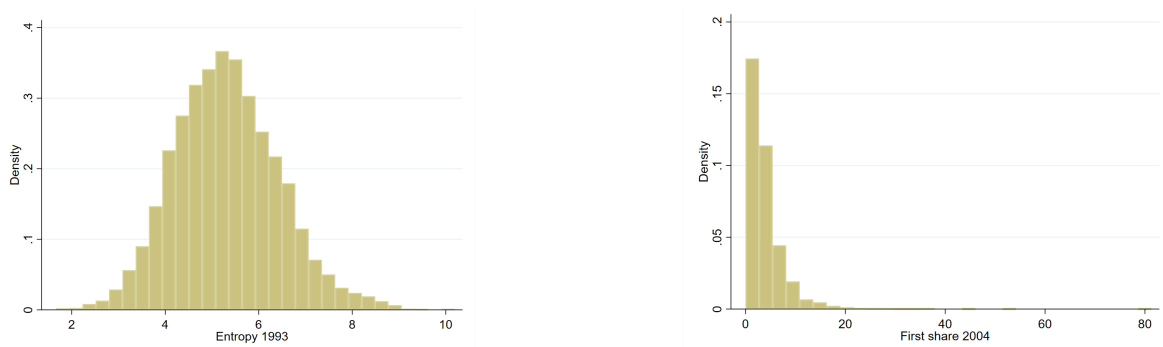
(a) Entropy 1993

(b) First share 2004

Notes: Fig. 3a shows the distribution of our variable Entropy across Italian municipalities in 1993. Here, reddish areas indicate municipalities with more variable surnames distributions. Fig. 3b illustrates the distribution of the variable First share in 2004, always at the municipality level. In this case, the bluer areas have more concentrated surnames distributions. Data from Buonanno and Vanin (2017) include all 8090 Italian municipalities in the case of First Share, and 8074 in the case of Entropy.

In Fig. 4, we present the density functions of those two variables, together with some summary statistics in Tab. 1.

Fig. 4: **Entropy and First share density functions**



(a) Entropy 1993

(b) First share 2004

Tab. 1: **Summary statistics for Entropy 1993 and First share in 2004**

Variable	Obs	Mean	Std.Dev.	Min	Max
Entropy 1993	8,074	5.31	1.12	1.66	10.19
First Share 2004	8,090	3.93	3.78	0	81.32

Notes: Fig. 4a shows the density function of the variable Entropy in 1993, while Fig. 4b illustrates the density function of First share in 2004. Tab. 1 describes some summary statistics for the variables Entropy in 1993 and First share in 2004. Number of observations, mean, standard deviation, minimum and maximum values are reported. Data from Buonanno and Vanin (2017) include all 8090 Italian municipalities in the case of First Share, and 8074 in the case of Entropy.

As we can see, their densities are extremely different, as *Entropy* resembles a normal distribution and $Firstshare$ is way more right-skewed. This is not particularly surprising as $Firstshare$ takes into account only the most common surname in the municipality, while *Entropy* all of them. The year of reference of those two variables, as we pointed out, is 1993 for *Entropy* and 2004 for $Firstshare$: we can assume that they proxy the intensity of endogamy rates of the previous decades. As we explained previously, we can interpret the intensity of endogamy as a proxy for the conservation and enforcement of traditional social norms. One of the most worrying confounders for the use of these proxies for endogamy is that the diversity of surnames is also a consequence of migration waves: in Appendix A we discuss further this confounder and argue why it should not bias our results.

4.1 Sources of endogeneity and identification strategy

Our identification strategy relies on the use of instrumental variables in order to prove that endogamous marriages generate larger gender participation gaps. In fact, there might be two possible sources of endogeneity when we assess the relationship between these two variables through simple OLS:

1. Omitted variable: as Cavalli-Sforza et al. (2004) argue, there is a large variety of elements influencing the frequency of in-marriage, namely the financial stimulus not to disperse the familiar patrimony, literacy rates, birth rates, industrialization, the impor-

tance of family values, population size, size of the extended family, ruralization. Since some of these elements might impact differently men's and women's labor supplies, we could have an endogeneity problem. While we can control for some of these factors, such as literacy rates, others (e.g. importance of family values) are difficult to measure and include in our regressions, creating an omitted variable bias in OLS' estimated coefficients.

2. Reverse causality: the effect that we theorized might be reversed, since having a high share of non-working women in a community might increase its endogamy rate. As a matter of fact, not having a job reduces the number of social exchanges outside an individual's community: even working part-time as a company's secretary expands the frequency of "outer" interactions that a woman can have with respect to a housewife. As a consequence, it might be more likely that a non-working woman ends up marrying one of the members of her community simply because she has contact exclusively with them.

Therefore, when we estimate the relationship between endogamy and rates of gender participation gap through simple OLS, the coefficients that we obtain may be biased[11].

To reduce the first of the two concerns we listed, we include in our specification a set of controls at the municipality level, namely:

- **Geographic controls**
- **Demographic controls** [12]
- **"Sistema Locale del Lavoro" fixed effects**. Italy can be virtually divided into 686 "Sistema Locale del Lavoro" ("Local Labor Markets"), which are territorial circumscriptions that take into account how people commute and move daily to go to

[11] In the case of reverse causality, the direction of the bias should be positive: more endogamy generates less working women (thus, larger gender participation gap) who are more likely to marry inside the community, thus increasing the endogamy level. Therefore, the OLS coefficient should be downward biased in the case of *Entropy* and upward biased in the case of *First share*. On the other hand, the direction of the bias might be both upward-biased or downward-biased by the omitted variables.

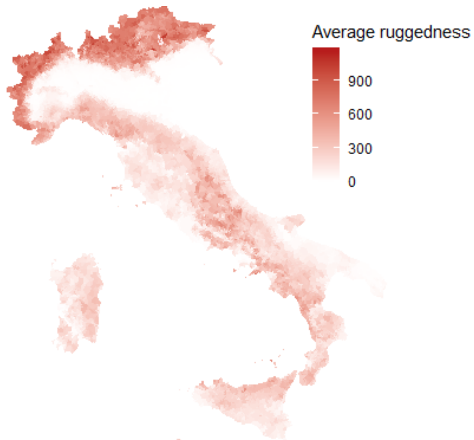
[12] See Appendix B for the entire list of controls

their workplace. The average local labor market circumscription in Italy comprises 7 municipalities and has a surface extension of 335 square kilometers.

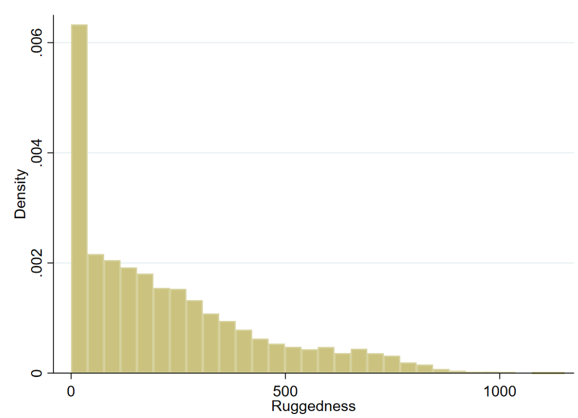
The inclusion of SLL fixed effects aims to capture possible unobserved differences in terms of characteristics of the labor market between these territorial circumscriptions. However, even with the inclusion of these controls we cannot exclude the possibility of other potentially relevant omitted variables correlated with both the endogamy proxies and rates of gender participation gap, and reverse causality might still bias our results. Using a measure of endogamy from 1993 (Entropy) reduces the concern of reverse causality: even if the treatment and the outcome are two variables characterized by sluggishness, the time distance in this case (8 years) should reduce the worry of possible effects of the gender gap on endogamy. Nevertheless, some reverse causality effects might still persist: this is why we recur to an instrumental variable strategy in order to assess our causal relationship of interest. Specifically, we instrument endogamy rates at the municipality level with the average degree of ruggedness of the municipality's territory. Ruggedness is a measure of irregularity of the land that is available at a disaggregated level again from the same dataset of Buonanno and Vanin (2017), who constructed it from the Global Land One-km Base Elevation Project (GLOBE). Ruggedness weights lands' average differences in elevation, thus giving a sense of how arduous the transition through a territory is[13]. In Fig.5 we present a map of how the average ruggedness index varies across Italian municipalities, together with its summary statistics and density function:

[13]To construct this measure, the Italian territory was divided into cells at 10-minute spatial resolution (approximately 1 squared kilometer of surface). Then, it was computed the difference in elevation between the centroid of each cell and each of the centroids in each of its eight confining directions (north, northeast, east, southeast, south, southwest, west, and northwest). Then, each cell's ruggedness measure is given by the squared root of the sum of the squared differences in elevation between the central point and the eight adjacent centroids. Finally, each municipality's ruggedness index is the average ruggedness level of its cells.

Fig. 5: **Ruggedness' distribution and density function**



(a) Average ruggedness (m)



(b) Ruggedness density function

Tab. 2: **Summary statistics for average ruggedness**

Variable	Obs	Mean	Std.Dev.	Min	Max
Avg Ruggedness (m)	8,074	224.27	215.76	0	1151.44

Notes: Fig. 5a shows the distribution of ruggedness across Italian municipalities, while Fig. 5b illustrates its density function. Tab. 2 describes some summary statistics for ruggedness. Number of observations, mean, standard deviation, minimum and maximum values are reported. Ruggedness is measured in meters. Data from Buonanno and Vanin (2017) include 8074 out of 8090 Italian municipalities.

The relationship between ruggedness and endogamy is straightforward: more rugged lands are harder to traverse and live in, thus communities living in such areas are more geographically isolated. Therefore, we argue that this kind of isolation incentives marriage within the community, given the geographical obstacles in finding outer partners. Our first stage has the following specification:

$$Endogamy_i = \delta + \theta Ruggedness_i + \lambda X_i + u_i$$

With *Ruggedness* the average ruggedness level for municipality i. When we implement our 2SLS regression, the coefficient obtained captures the so-called Local Average Treatment

Effect (LATE). In other words, it estimates the effect of the treatment "Endogamy" on the gender participation gap for those municipalities for which treatment status, i.e. the rate of in-marriage, has been changed by the instrument *Ruggedness*.

4.2 Validity of the instrumental variable

Following Angrist and Imbens (1995), for our identification strategy to capture the correct LATE-effect of endogamy on the gender participation gap, five assumptions concerning the instrument must hold. In this section, we discuss exclusively the most problematic one, namely the exclusion restriction validity, while we leave the discussion on the other four assumptions to Appendix C. For the exclusion restriction to hold, our instrument, ruggedness, should not have a direct impact on the gender participation gap, the outcome. However, there is evidence that ruggedness directly influences economic activity (Nunn and Puga (2012)): rugged territories are harder to cultivate, obstacle transportation of goods, and make it more difficult to build infrastructures, thus creating a scarcity of jobs. This direct effect on economic activity might impact labor market characteristics, and therefore demand or supply of female labor. To address this concern, we focus on the main channels through which ruggedness affects the economic system: agriculture, road system, and labor market characteristics. In order to prevent ruggedness from directly impacting our outcome through either one of these channels, we include in our analysis controls at the municipality level for all of them, as we discuss next.

- **Agriculture**: Ruggedness' effect in terms of deterrence of agricultural activities is captured by "suitability for agriculture" control (available from Istat, referred to 2001), the percentage of land possible to cultivate. We argue that this control is the most exogenous one among all variables describing agricultural activities since it directly measures a characteristic of the territory[14]. Moreover, we know from a recent paper

[14] Conversely, variables such as the number of agricultural cooperatives or percentages of residents employed in the agricultural sectors are more endogenous to our dependent variable, and may depend on other unobservables (e.g. the existence of cooperatives depends on the level of civicness of a community (Putnam

by Boone and Wilse-Samson (2021) that more rugged areas are less suitable for mechanization and are characterized by less capital-intensive production. In Appendix G, we discuss in detail this further concern and show how ruggedness is not influencing individuals' occupational choices across Italian municipalities.

- **Road connection**: We include a "Roman roads" dummy in our analysis (indicating the historical presence of a main road connecting the municipality during the Roman empire) which should prevent ruggedness to influence labor market's characteristics by impeding the creation of connections and transportation of goods to a municipality. Indeed, the road system created during the Roman empire constituted the main skeleton of what is Italy's contemporaneous transportation network (Buonanno and Vanin (2017)). Given the historical reference of this variable, we can argue that it is less influenced by the most recent heterogeneous levels of economic development across Italy with respect to contemporaneous measures of municipalities' road connections, and therefore it is less subject to endogeneity concerns. Thus, if ruggedness influenced female or male labor supply by reducing the availability of transports and connections, this heterogeneous effect across municipalities should be captured by our Roman Roads dummy[15].

- **Other labor market characteristics**: Ruggedness could generate a scarcity of jobs for the unwillingness of companies to do business in those irregular territories, generating high unemployment rates that might discourage women from entering the labor force. Furthermore, given the scarcity of jobs, we may think of other possible dynamics in rugged territories, such as employers deciding to offer jobs with a gender bias, and preferring to fill vacancies with men. In line with the conservation of traditional gender roles, we may think of the preservation of prejudices about productivity levels

et al. (1994)), which might also have an impact on female labor supply.)

[15] Moreover, to further reduce the worry of ruggedness affecting the degree of individuals' mobility, we include a control for the percentage of commuters over the resident population (namely, the percentage of residents moving daily out of the municipality for work reasons).

of men and women in the more "endogamous" communities. To reduce concerns about these other possible direct effects of ruggedness, we include controls for the gender unemployment gap (difference between unemployment rates of men and women) and for the percentage of graduated women. Last, we include "Sistema locale del lavoro" (local labor market) fixed effects, which should capture other possible and unobserved consequences of higher ruggedness levels on the labor market. In other words, the effects of endogamy on the gender gap that we present in Section 5 are observed within the same local labor market.

The inclusion of all these controls should prevent ruggedness from directly impacting the gender participation gap by generating adverse conditions in the labor market[16]. After their inclusion in our regression, we can assume the effect of ruggedness on the gender participation gap to be exclusively through the increase of endogamy within a community[17]. Moreover, to provide an additional robustness check supporting the validity of the exclusion restriction, we perform a placebo test in Appendix D, demonstrating that on subsamples of municipalities for which the endogamy level is stable, a variation in the level of ruggedness is not linked to alterations in the gender participation gap. Furthermore, in Appendix E we also relax the exclusion restriction assumption by following the approach developed by Nevo and Rosen (2012)[18]: results in terms of significance level and direction of coefficients are not affected. The validity of this and of the other four assumptions is fundamental in order to assign to 2SLS' estimated coefficients of $Entropy$ and $FirstShare$ the LATE interpretation and for the soundness of our identification strategy. It is important to remember that the 2SLS-LATE framework identifies an effect valid only for a subsample of observations, the so-called compliers: in other words, the subsample of municipalities for which the treatment effect is

[16] Since we may speculate that more isolated communities have an older population with lower involvement in the labor market or education level, we control for the age structure of the municipality

[17] Regarding the potential bias coming from the importance of family values, we discuss it in Appendix H, where we show how ruggedness is not linked to two different proxies of family values. Therefore, the exclusion restriction is preserved.

[18] Substantially, this approach replaces the exclusion restriction assumption in the IV framework with a less restrictive condition allowing some degree of correlation between the instrument and the endogenous variable: even in this case, the sign and magnitude of our treatment effect are not heavily affected.

influenced by the instrument (Angrist and Imbens (1995)). In Section 7 we discuss more extensively which kinds of cities could be the drivers of our results

5 Main results

Moving to our empirical results, we first present the baseline OLS results, with the gender participation gap in 2001 as the dependent variable and $Entropy/Firstshare$ as the main explanatory variables, including the set of controls we listed. Given the endogeneity concerns that we explained previously, we then move to the 2SLS regressions, instrumenting $Entropy/Firstshare$ with average $Ruggedness$. We present the OLS results in Tab.3.

Tab. 3: **Effect of endogamy on gender participation gap - OLS**

	Dependent variable:					
	Participation gap 2001					
	OLS	OLS	OLS	OLS	OLS	OLS
Entropy 1993	-0.575***		-0.829***		-0.227*	
	(0.196)		(0.221)		(0.118)	
First share 2004		0.187***		0.240***		0.037*
		(0.046)		(0.045)		(0.022)
Geographic controls	No	No	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	No	Yes	Yes
SLL FE	No	No	No	No	Yes	Yes
Observations	7,726	7,726	7,726	7,726	7,726	7,726
R-squared	0.015	0.017	0.045	0.046	0.550	0.549

Note: Standard OLS regressions. The dependent variable is the gender participation gap in 2001 (male labor force participation - female labor force participation), while the explanatory variables of interest are Entropy 1993 in the first case and First share 2004 in the second one. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level.
*** p<0.01, ** p<0.05, * p<0.1

As we can observe, the coefficients of both our proxies for endogamy are highly significant in all specifications, and their direction is as we theorized: municipalities with higher rates of in-marriage exhibit larger gender participation gaps. In fact, $Entropy$'s coefficients are negative, while $Firstshare$'s ones are positive in each specification, even with the inclusion of the large set of controls that we outlined previously. Coefficients for both endogamy proxies are characterized by a degree of volatility, especially after the inclusion of demographic controls and SLL fixed effects: this is due to the fact that some of the controls are linked to both labor market characteristics and marriage patterns[19]. Given the concerns for omitted variable bias and possible reverse causality, we turn next to 2SLS regression. We begin by presenting the first stage results on Tab. 4:

Tab. 4: **First stage 2SLS**

	First stage results					
Dependent Variable	**Entropy**	**F.share**	**Entropy**	**F.share**	**Entropy**	**F.share**
Avg ruggedness	-0.196***	0.716***	-0.166***	0.462***	-0.086***	0.254***
	(0.020)	(0.089)	(0.020)	(0.060)	(0.012)	(0.072)
Geographic controls	No	No	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	No	Yes	Yes
SLL FE	No	No	No	No	Yes	Yes
Observations	7,726	7,726	7,726	7,726	7,726	7,726
Cragg-Donald F test	1246	809.8	422.3	291.9	151.2	65.4

Note: First stage regression in the 2SLS setting. The dependent variable is Entropy 1993 in the odd columns and First share in the even ones, while the explanatory variable of interest is average ruggedness. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Coefficients and standard errors have been multiplied by 100, to allow an easier understanding of the relationship between the instrument and our endogenous variables. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

As we can observe from first-stage regression results, the effect of *Ruggedness* on our proxies for endogamy is coherent with what we theorized previously: a higher level of irregularity of the municipality's surface contributes on average to increasing the social closure of its community. As a consequence, the effect of average *Ruggedness* is positive for $Firstshare$ and negative for *Entropy*, in both cases shrinking the concentration of surnames. Next, we move to second stage results on Tab. 5, with the gender participation gap as our dependent variable:

[19] For instance, the variables most responsible for the change in magnitude of *Entropy* and $FirstShare$ coefficients are the percentage of women graduated and the percentage of women with the diploma. Clearly, the decision to invest in education has potential effects on both the career and marriage decision of women. Also the extension of surface and the suitability for agriculture are responsible for a large change in the magnitude of our baseline coefficients, as these two variables are likewise potentially influencing both marriages and labor market patterns across communities.

Tab. 5: **Effect of endogamy on gender participation gap - 2SLS**

	Dependent variable:					
	Participation gap 2001					
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Entropy 1993	-0.582		-1.749***		-1.488**	
	(0.562)		(0.620)		(0.716)	
First share 2004		0.160		0.628***		0.505**
		(0.150)		(0.227)		(0.231)
Geographic controls	No	No	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	No	Yes	Yes
SLL Fixed Effects	No	No	No	No	Yes	Yes
Observations	7,726	7,726	7,726	7,726	7,726	7,726
R-squared	0.015	0.016	0.022	0.191	0.532	0.501

Note: Second stage regression in the 2SLS setting. The dependent variable is gender participation gap in 2001 (male labor force participation - female labor force participation), while the explanatory variables of interest are Entropy 1993 in the first case and First share 2004 in the second one. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

As we can see, the effect of both *Entropy* and $Firstshare$ is significant and coherent to our main argumentation that endogamy is positively linked to the gender gap in every specification apart from the first two columns. Coefficients present the same sign as OLS' ones but are larger in magnitude: the direction of the bias overcome by the instrumental variable strategy seems to be negative in the case of *First share* and positive in the case of *Entropy.* This is the opposite result that we would have expected to be generated by reverse causality: it might be due to the fact that there are some relevant omitted variables negatively linked to endogamy and positively linked to the gender participation gap, generating a bias in the opposite direction of the reverse causality's one[20]. However, the 2SLS coefficients still confirm our prediction about the positive relationship between in-marriage and the gender participation gap.

To give an idea of the magnitude of our 2SLS coefficients, if we standardized the variables we would see that:

[20] For instance, immigration of men inside the municipality for job-related motives could have enlarged both the variety of surnames (thus, decreasing endogamy) and the gender participation gap.

- A standard deviation increase in *Entropy* is linked to a 0.32 standard deviation decrease in the gender participation gap

- A standard deviation increase in *Firstshare* corresponds to a 0.35 standard deviation increase in the gender participation gap

The standard deviation of the gender gap is 5.31, thus a 0.3 standard deviation change corresponds to roughly a 1.6% variation in this variable. To understand the size of potential shocks in endogamy, consider that the standard deviations of our two endogamy proxies are 3.51 for *FirstShare* and 1.10 for *Entropy*, while the two averages are respectively 3.82 and 5.35[21]. The gender participation gap reduction in the period 1971-2001 was on average 9.4% across Italian provinces: the potential effect of variations in endogamy on the gap reduction is therefore sizeable. In the period 2001-2011 the average reduction was 3.87%: therefore, a standard deviation change in endogamy has a potential effect that corresponds to almost half of 10 years' reduction in the gender gap in participation in the labor force. If the LATE assumptions we discussed in Section 4 and Appendix C hold, we can state that higher endogamy rates raised the gender participation gap for the Italian municipalities for which ruggedness increased the frequency of in-marriage. We might even interpret these results in the following way: if we made parallelism between surnames and social norms, we provided evidence that not only the kind of culture matters in determining the gender gaps, but also its degree of homogeneity. The inclusion of all the controls we listed previously, as well as 686 "Sistemi Locali del Lavoro" fixed effects[22], does not decrease the magnitude or significance of *Entropy* and *FirstShare*'s point estimates [23][24]. Given that also in this 2SLS

[21] As we explain in Section 4, these two variables have extremely different density functions, being *FirstShare* more right-skewed.

[22] Thus, we are comparing municipalities within the same local labor market

[23] It seems that the controls most responsible for the increase in absolute value of *Entropy/First Share* coefficients are average elevation level, suitability for agriculture, gender unemployment gap and Roman Roads. As we discussed in the previous section, those variables are the controls that we implemented to prevent *Ruggedness* from having additional effects on our outcome, so they are more likely to remove any ulterior biases from the 2SLS coefficients.

[24] Even if contexts and variables are different, we can compare our work's results to similar ones in terms of magnitude: for instance, studying the effects of culture on second-generation immigrants in the US, Fernandez and Fogli (2009) find that a standard deviation increase in labor force participation in 1950 is associated with approximately a 7.5 percent increase in hours worked per week by women in 1970. Again

framework, coefficients are characterized by some volatility which might be caused by potentially bad controls, in Appendix I we offer a more "conservative" specification, by including also the most exogenous among the controls: results are still confirming the positive effect of endogamy intensity on the gender gap.

Even if we justified the choice to use the gender participation gap, and not female labor force participation, as an outcome for its stricter link with gender inequality, in Appendix M we show how larger endogamy is also causing lower participation of women in the labor force across Italian municipalities. In addition, coherently with our theoretical framework, in appendices J-K we show that municipalities with higher degrees of endogamy present also larger gender education gaps (difference between rates of men and women holding at least a diploma) and lower provision of public childcare (defined as the number of kindergarten places available every 100 children aged 0-2 years old). These results contribute to underlining how in those municipalities the role of woman is still central inside the house, leaving less time to invest in education or high-profile career jobs.

Since all the 2SLS results could be affected by the spatial autocorrelation issues described by Kelly (2019), in appendix F we correct our standard errors for interdependence between neighboring municipalities: our main coefficients remain significant at the 10% level.

6 Discussion and channels' assessment

As we discussed in the previous section, the influence of endogamy rates on the wideness of the participation gap is noteworthy and deserves a further and in-depth analysis.

Indeed, our next step focuses on discussing the potential channels through which in-marriage discourages women from entering the labor force. As we pointed out in the introduction, our argument is that there might be two mechanisms through which higher rates of endogamy

for the US, Fernández and Wong (2014) found that an increase in divorce risk was responsible for over 42% of the LFP increase between the 1935 and the 1955 cohorts during the ages of 25-40 for married college women and 49% of the LFP increase for married high-school women. Moreover, studying the effects of having had a working mother because of World War II, Fernández et al. (2004) found that a 10 percent higher mobilization rate was associated with 3.3 additional weeks worked by women 45–50 years old in 1980.

produce larger gender participation gaps: first, by preserving traditional social norms and thus the older gender roles, and secondly, by lowering the divorce probability for those women who married inside the community. Thus, we want to provide additional empirical evidence supporting the validity of these channels and assess whether data can confirm our predictions. We proceed as follows:

1. To verify the existence of the social norms' channel, we move back our analysis to a period in which divorces were not legal or anyway extremely unlikely: using data from a national survey of 1972 (Barnes (1984)), we assess the link between endogamy and female labor force participation. Indeed, divorce was introduced in Italy only in 1970 and it was subject to heavy criticism by the Catholic Church and the most traditional parties, leading to a (lost) referendum for divorce abrogation in 1974: therefore, we can assume that in 1972 the probability of a couple divorcing was negligible in the whole peninsula, as actual divorce numbers confirm[25]. Thus, when we assess the impact of endogamy rates on female labor supply in 1972, we would consider exclusively the social norms' effect and not the divorce channel.

2. To evaluate the probability of divorce channel, we compare municipalities' percentages of female divorced residents in 2001 and estimate whether rates of endogamy significantly increased marital stability across Italian municipalities.

6.1 Social norms' channel - 1972 survey

The survey was conducted in May and June 1972 (after the general elections on the 7th of May) on a sample of 1841 individuals which is representative at the national level.[26] Data were collected through face-to-face interviews conducted by Samuel Barnes and Giacomo Sani, two researchers from the "Inter-university Consortium for Political and Social Research". The interviews focused on respondents' political interest, behavior and attitudes,

[25] In 1972 there were 24000 divorced residents in Italy, corresponding to the 0,04% of the population.

[26] Individuals come from all 20 Italian regions. However, the sample is not representative at regional or municipality levels.

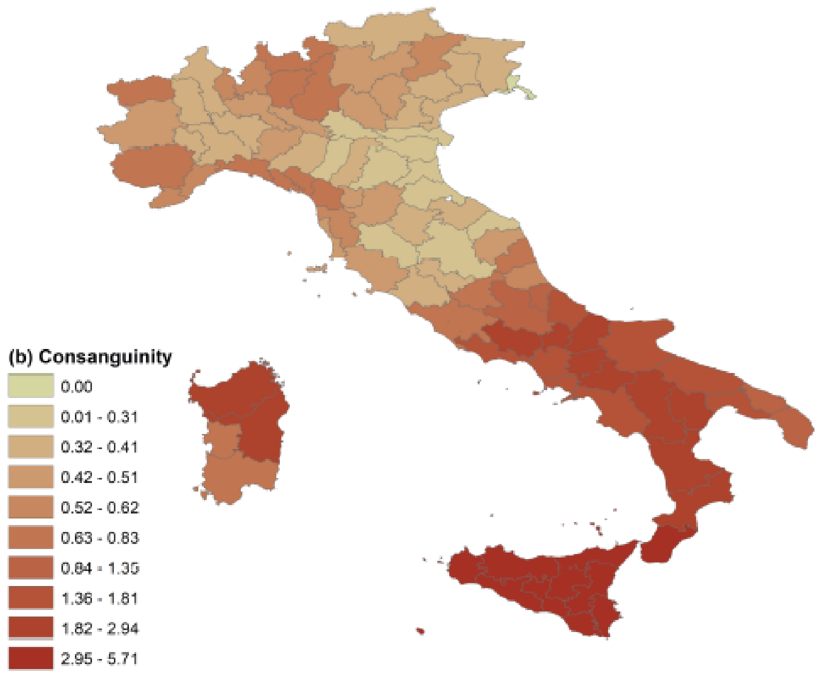
their party identification and organizational memberships, trust in government, reaction to the multi-party system, and views on left-right political differences. Demographic information about respondents included age, occupation, full-time work status, profession and political beliefs of the father. We focus on the 934 female respondents: the main dependent variable of this analysis is a dummy equal to 1 if the respondent is a housewife[27]. Among the 934 women in the sample, 57% declared to belong to this category (thus, not participating in the labor force). Since we have only information about the respondents' province of residence and not about the municipality, we would need a measure of the intensity of endogamy available for this level of disaggregation (NUTS 3 level). For this purpose, we choose the indicator of rates of consanguineous marriages in the period 1945-1964, available from the work of Cavalli-Sforza et al. (2004). This variable tells us about the percentages of marriages that were either between first cousins or between uncle/aunt and nephew/niece[28]. Those data were collected from the Vatican Archives: as a matter of fact, in Italy it is still necessary to have written authorization from the Catholic Church in the case that two individuals from the same family want to celebrate a religious marriage [29]. In Fig. 6, we can see a map of the frequency of consanguineous unions in Italy for the cited period.

[27] The fact that a woman is a housewife might be dictated by a large variety of factors, and it might not have been her decision. We try to exclude the possible confounding factors by controlling for observables, and again by implementing our instrumental variable strategy.

[28] Ideally, we would have liked to aggregate $Entropy/FirstShare$ at the province level. Our worry, in that case, is that a time gap of 30-40 years between this measure and our outcome would not allow us to correctly identify the effect of endogamy on female labor supply, since possible migration waves or heterogeneous fertility/mortality trends across cities could bias our results. However, if we nevertheless average the two endogamy proxies by province and look at their correlation coefficients with Consanguinity, we obtain a correlation score of -0.2428 with Entropy and 0.1212 with First Share. This additional piece of evidence confirms that all these three indicators are measuring the tendency to marry within the community.

[29] Data on the written authorizations from the Vatican Archives were publicly available only for the period 1945-1964.

Fig. 6: **Consanguinity rates 1945-1964**



Note: Rates of marriages celebrated either between first cousins or between uncle/aunt and niece/nephew over total marriages between 1945 and 1964. Cavalli-Sforza et al. (2004) collected those data from the Vatican Archives, which contain information about almost the totality of consanguineous marriages celebrated in Italy during that period. The image is taken from Akbari et al. (2019) and data include all 92 Italian provinces in 1964.

Consanguinity rates are another convenient measure of endogamy intensity at the local level, proxying the extent to which marriage patterns in the provinces were circumscribed to local communities. This is our explanatory variable of interest, and we instrument it with the province's ruggedness level, as in our main regression. In substance, we want to evaluate whether the intensity of consanguineous marriages in the province of residence of the respondent is able to explain the probability that she does not work, i.e. that she is a housewife. We include also individual-level controls [30], together with province-level ones [31] to reduce the concern of omitted variable bias [32]. On Tab.6, we can see both OLS and 2SLS

[30] Size of municipality of residence, educational level, age, dummy for catholic, dummy for married

[31] Distance from the coast, elevation, suitability for agriculture, population in 1971, unemployment rate, ratio of civil versus religious marriages, surface size, population density

[32] We could consider some of the individual-level controls as bad controls, e.g. consanguinity might affect education level. However, our main results are stable when we include only pure demographics, as we see from Tab. 6.

results.

Tab. 6: **Consanguinity rates and probability of being a housewife**

	Dependent variable:					
	Housewife					
	OLS	*2SLS*	*OLS*	*2SLS*	*OLS*	*2SLS*
Consanguinity rates	0.046***	−0.104	0.080***	0.260**	0.043***	0.189*
	(0.013)	(0.065)	(0.022)	(0.109)	(0.016)	(0.098)
Province controls	NO	NO	YES	YES	YES	YES
Individual controls	NO	NO	NO	NO	YES	YES
Observations	934	934	934	934	920	918
First stage F test		42.110		40.139		42.009

Note: OLS and 2SLS regressions with the dummy "Housewife" as dependent variable. The explanatory variable of interest is the consanguinity rate of the respondent's province of residence between 1945 and 1964. In the 2SLS setting, consanguinity rates are instrumented with the average ruggedness level of the province. Geographic and demographic controls at the province level are included in the specification. Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

As we can observe, consanguinity rates are significantly increasing the probability of being a housewife in almost every specification, apart from the first one in the 2SLS framework. We argue that in this case, endogamy decreases female labor supply only through the preservation of social norms concerning the inappropriateness of working women. Indeed, the "probability of divorce" channel was practically non-existing in 1972: in this way, we are presenting supporting evidence to validate a first channel between endogamy and the gender participation gap.

An additional piece of evidence supporting this channel is included in appendix L: in this case, we evaluate how consanguinity rates influenced the divorce referendum results. Indeed, if we follow our theoretical mechanisms, the more endogamous communities should have been more averse to the introduction of divorce, since it represented a threat to traditional Catholic values and marriages' stability. Indeed, as we show in the appendix, Italians living in provinces with higher consanguinity rates were more inclined to vote for divorce abrogation.

6.2 Probability of divorce channel

Beyond the preservation and enforcement of gender norms, a custom of in-marriage decreases the probability of divorce among community couples, thus reducing incentives for wives to look for jobs. As a next step, we would like to test whether higher endogamous rates within a community can generate stronger marital stability, thus another channel explaining our main result. We choose to perform a 2SLS regression with the usual specification but using "Percentage of divorced female residents" as the dependent variable, defined as the percentage of resident women aged more than 15 that appear as being divorced in their legal status. This variable, obtained from the 2001 census, is a valid measure of marital instability: clearly, if more divorced women are resident in a municipality, it means that marriages are on average less stable and couples break up more frequently [33]. On Tab. 7, we can observe 2SLS results exposing the relationship between endogamy rates and divorced women.

Tab. 7: **Endogamy and marital stability**

	Dependent variable:					
	Female divorced residents 2001					
	2SLS	*2SLS*	*2SLS*	*2SLS*	*2SLS*	*2SLS*
Entropy 1993	0.080*		0.401***		0.179**	
	(0.042)		(0.071)		(0.087)	
F.Share 2004		-0.022*		-0.119***		-0.065*
		(0.012)		(0.023)		(0.033)
Geographic controls	No	No	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	No	Yes	Yes
SLL Fixed Effects	No	No	No	No	Yes	Yes
Observations	7,712	7,712	7,712	7,712	7,712	7,712
First stage F test	1249.02	815.80	394.91	182.73	237.76	52.31

Note: 2SLS regressions with the percentage of female divorced residents in 2001 (namely percentage of the female population resident in the municipality with "divorced" as legal status) as the dependent variable. The explanatory variables of interest are Entropy 1993 in the odd columns and First share 2004 in the even ones, both instrumented with the average ruggedness level of the municipality. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values, 378 observations have been dropped, out of the total 8090 municipalities. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Results from Tab. 7 show how communities with higher endogamy rates had on average fewer female divorced residents in 2001. Hence, data seem to confirm the existence of this

[33] Heterogeneous frequencies of re-marriages within the population could produce measurement error since a divorced and re-married woman figures as "married" in the census. Unfortunately, we do not have data at the municipality level for second and third marriages, which were however quite uncommon before 2001, with frequencies around 6% of total marriages celebrated every year.

second channel, with municipalities more prone to in-marriage exhibiting a lower probability of divorce, which in turn might have led to lower female labor force participation and a higher gender participation gap in 2001.

7 Robustness checks

7.1 Main result in subsamples

To provide some robustness to our main result, we check whether it is confirmed by an analysis across subsamples. To do so, we first evaluate the relationship between endogamy and the gender participation gap in subsamples including only Southern, Center or Northern municipalities, and afterward we verify the strength of our main result across subsamples based on municipalities' size and geographical characteristics. All these analyses together also provide an assessment of the extent to which our results can be considered valid for Italy as a whole, and not driven exclusively by particular subsamples of municipalities. In Tab.8 we show the results from performing our main 2SLS regression on macro-areas subsamples (North-Center-South), together with the whole sample results (included to allow comparisons).

Tab. 8: **2SLS results over macro-regional subsamples**

	Dependent variable: Participation gap 2001							
SUBSAMPLE	North	North	Center	Center	South	South	Whole	Whole
Entropy 1993	-1.830***		-2.403**		-0.352		-1.488**	
	(0.668)		(1.055)		(1.095)		(0.716)	
F.Share 2004		0.951***		0.554*		0.135		0.505**
		(0.320)		(0.293)		(0.418)		(0.231)
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SLL Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,878	3,878	1,329	1,329	2,519	2,519	7,726	7,726
Mean dep. variable	22.73	22.73	22.82	22.82	24.70	24.70	23.39	23.39
1st Stage CD Wald F-test	108.32	24.09	48.37	60.32	85.12	27.32	151.22	65.49
1st Stage KP Wald F-test	88.87	7.78	35.14	21.24	63.84	23.30	55.40	12.30

Note: 2SLS regressions over macro-regional subsamples: the first two columns include only Northern municipalities, the third and fourth columns include municipalities from Center Italy, while the fifth and sixth columns include Southern ones. The last two columns present the whole sample results, to allow an easier comparison. The dependent variable is the gender participation gap in 2001, while explanatory variables of interest are Entropy 1993 in the odd columns and First share 2004 in the even ones, both instrumented with the average ruggedness level of the municipality. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level.
*** p<0.01, ** p<0.05, * p<0.1

Looking at Tab. 8, we can clearly observe that our main result holds in significance level on two out of three subsamples of observations. The coefficients of the Southern subsample, while always keeping the same sign (*Entropy*-negative and $FirstShare$-positive) and similar magnitude with respect to the full sample's results, are less precise and not significant in explaining variations in gender participation gap[34]. With respect to the full sample's results, both magnitudes of coefficients and their standard errors are larger for the Northern and Center subsamples: thus, the results are less precise but still confirm the positive effect of endogamy on the outcome. It seems that Center-North municipalities are the main drivers of our results in terms of significance, even if these cities have slightly lower levels of par-

[34] The fact that these coefficients are not significant seems to be due to a lack of precision. Indeed, when we perform the same regressions on a subsample including both Center and Southern Italian municipalities, both *Entropy* and $FirstShare$ coefficients become significant at the 1% level in explaining the gender participation gap.

ticipation gap with respect to the South, as we can observe from the table. This could be explained by the fact that having a higher and more variable ruggedness in the North allows our instrument to capture more variation in endogamy, making its statistical power stronger for this subsample of municipalities. However, the first stage F tests confirm that our instrument is informative across all our three subsamples. Thus, even if ruggedness exhibits lower variation in the South and Center of Italy, it is still able to capture enough variations in endogamy to allow the implementation of our identification strategy. What we find by looking at these geographical subsamples is still in line with our major argument: higher levels of endogamy generate larger gender participation gaps across municipalities.

Furthermore, we might be concerned that our main result is not a direct consequence of endogamy, but it is driven exclusively by smaller Italian villages. In fact, one could argue that in a small village or town, both marriage and labor markets offer scarcer opportunities with respect to larger cities: as a consequence, one could observe both higher rates of in-marriage and a larger gender participation gap in villages, but this would not imply a causal link between them. Thus, on Tab. 9 we perform once again our main regression on two subsamples, the first time excluding small villages (identified as those with a population lower than 5000) and the second time focusing exclusively on them.

Tab. 9: **2SLS results, cities and villages**

	Dependent variable:			
	Participation gap 2001			
Population	$> 5k$	$> 5k$	$< 5k$	$< 5k$
Entropy 1993	-1.716*		-3.633**	
	(0.936)		(1.843)	
F.Share 2004		0.802		0.787**
		(0.502)		(0.378)
Geographic controls	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
SLL Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,329	2,329	5,397	5,397

Note: 2SLS regressions over subsamples with different municipality's size. In the first two columns, municipalities' population is higher than 5000 inhabitants, while in the last two the population is lower than 5000 inhabitants. The dependent variable is the gender participation gap in 2001, while explanatory variables of interest are Entropy 1993 in the odd columns and First share 2004 in the even ones, both instrumented with the average ruggedness level of the municipality. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1

Our main result holds in significance and sign, however we can notice how the magnitude of *Entropy*'s effect is larger when we consider only villages. We have that for cities with more than 5000 inhabitants the effect is not significant when we use *FirstShare* as a proxy[35], but it is still in line in terms of magnitude and sign with the coefficient from the last column. Therefore, this robustness check confirms how our main result is not driven exclusively by small Italian villages: the link between endogamy and the gender participation gap regards also bigger cities.

An additional concern is that our main results are driven by cities in highly mountainous areas, and thus are not informative about Italy as a whole. To exclude this possibility, we perform our 2SLS regressions by focusing only on cities located in areas with an altitude lower than 300m, which constitute 50% of our sample[36]. This subsample, therefore, excludes areas in highly mountainous areas, located in the Alps or Apennines. Tab.10 shows these results, comparing them to the full sample ones.

[35] P-value is 1.6.

[36] The average altitude for the Alps is 1300m, while for the Apennines is 800m, but this classification might change with respect to the areas considered.

Tab. 10: **2SLS regressions - non mountainous areas**

	Dependent variable:			
	Participation gap 2001			
Altitude level	$< 300m$	$< 300m$	Full sample	Full sample
Entropy 1993	-2.474**		-1.488**	
	(1.246)		(0.716)	
First Share 2004		0.606**		0.505**
		(0.293)		(0.231)
Observations	4,007	4,007	7,726	7,726
R-squared	0.614	0.623	0.532	0.501
Cragg Donald Wald F test	33.40	57.75	151.2	65.50
KP Wald F test:	12.75	7.752	55.40	12.30
Geographic controls	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
SLL Fixed Effects	Yes	Yes	Yes	Yes

Notes: Second stage regression in the 2SLS setting, subsamples divided with respect to the altitude level of the municipality's surface. The dependent variable is gender participation gap in 2001 (male labor force participation - female labor force participation), while the explanatory variables of interest are Entropy 1993 in the first case and First share 2004 in the second one. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1

As we can see from Tab.10, the coefficients remain coherent in sign to our main results. Also coefficients' magnitudes are in line, and the significance level is preserved. Moreover, the instrument is informative also for municipalities in low-altitude areas, with high values of the first stage F tests. Thus, we demonstrated that our results are not driven exclusively by municipalities in Alps or Apennines, since results remain strong and significant also for areas in non-mountainous areas.

8 Conclusion

A persistent custom of endogamy, namely a community's inclination towards in-marriage, can produce significant socio-economic consequences, such as higher corruption (Akbari et al. (2019)), lower impartial cooperation levels (Schulz et al. (2019)) and lower women enfranchisement (Bahrami-Rad (2021)). We look here at the effects of endogamy on the gender participation gap in Italy, a country historically plagued by poor levels of women's participation in the labor force. We argue that one of the main sources of this kind of gender inequality, together with other factors, can be identified with persistent social norms stigmatizing working women. An important element that contributes to preserving and enforcing these kinds of norms within a community is its past intensity of in-marriage: by marrying "inside", a community shields its traditional gender roles and protects them from the influence of outer social norms. In addition, endogamous unions are characterized by higher marital stability with respect to exogamous ones, which translates into lower divorce rates: this gives wives less incentive to enter the labor market, since they do not feel the need to have an "outside option" in the case that the marriage ends. To prove the existence of a causal link between in-marriage rates and the gender participation gap across Italian municipalities, we made use of an IV strategy, by instrumenting endogamy intensity (proxied by the concentration of surnames within a municipality) with the level of ruggedness of the municipality's territory. We showed that the more intense in-marriage had been within a municipality, the larger its gender participation gap was in 2001. Moreover, we provided evidence that another proxy for endogamy, the province of residence's rate of consanguineous marriages, is increasing the probability of women not joining the labor force, according to a representative sample of Italian women from 1972. Furthermore, we demonstrated that the intensity of in-marriage across Italian municipalities is associated with lower female divorced residents in 2001, confirming the existence of a causal link between endogamy and marriage stability. The robustness checks that we conducted demonstrate that our results are not exclusive to specific areas of Italy, such as North, villages or the Alps/Appennines, but can be generalized to the whole country.

The analysis that we performed through this research sheds light on a so far neglected dynamic contributing to generating and maintaining the large gender participation gap in Italy. The correct identification of the channels through which endogamy influences this gender inequality is fundamental for two reasons: first, to produce additional evidence supporting our theoretical idea, and second, to provide possible policy suggestions to overcome this vicious circle. Indeed, the high persistence of conservative social norms in determining gender roles and women's decision to enter or not the labor force has been widely established by the literature (Fernandez and Fogli (2009), Fernández (2013), Bertrand et al. (2015)): proving that endogamy contributes to social norms' persistence might help policymakers in taking effective measures in order to change them.

References

Mahsa Akbari, Duman Bahrami-Rad, and Erik O Kimbrough. Kinship, fractionalization and corruption. *Journal of Economic Behavior & Organization*, 166:493–528, 2019.

George A Akerlof and Rachel E Kranton. Economics and identity. *The quarterly journal of economics*, 115(3):715–753, 2000.

Stefania Albanesi and Claudia Olivetti. Gender roles and medical progress. *Journal of Political Economy*, 124(3):650–695, 2016.

Alberto Alesina, Paola Giuliano, and Nathan Nunn. On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2):469–530, 2013.

Alberto Alesina, Yann Algan, Pierre Cahuc, and Paola Giuliano. Family values and the regulation of labor. *Journal of the European Economic Association*, 13(4):599–630, 2015.

Alberto Alesina, Benedetta Brioschi, and Eliana La Ferrara. Violence against women: a cross-cultural analysis for africa. *Economica*, 88(349):70–104, 2021.

Joshua D Angrist and Guido W Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430):431–442, 1995.

Orazio Attanasio, Hamish Low, and Virginia Sánchez-Marcos. Explaining changes in female labor supply in a life-cycle model. *American Economic Review*, 98(4):1517–52, 2008.

Duman Bahrami-Rad. Keeping it in the family: Female inheritance, inmarriage, and the status of women. *Journal of Development Economics*, 153:102714, 2021.

Martha J Bailey. More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply. *The quarterly journal of economics*, 121(1):289–320, 2006.

Francesca Barigozzi, Helmuth Cremer, and Kerstin Roeder. Women's career choices, social norms and child care policies. *Journal of Public Economics*, 168:162–173, 2018.

Samuel Henry Barnes. *Italian Mass Election Survey, 1972*. Inter-university Consortium for Political and Social Research, 1984.

Marianne Bertrand, Emir Kamenica, and Jessica Pan. Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614, 2015.

Chris Bidner and Mukesh Eswaran. A gender-based theory of the origin of the caste system of india. *Journal of Development Economics*, 114:142–158, 2015.

Christopher DA Boone and Laurence Wilse-Samson. Structural change and internal labor migration: Evidence from the great depression. *The Review of Economics and Statistics*, pages 1–54, 2021.

Paolo Buonanno and Paolo Vanin. Social closure, surnames and crime. *Journal of Economic Behavior & Organization*, 137:160–175, 2017.

Francesca Carta and Lucia Rizzica. Early kindergarten, maternal labor supply and children's outcomes: evidence from italy. *Journal of Public Economics*, 158:79–102, 2018.

Luigi Luca Cavalli-Sforza and Marcus W Feldman. *Cultural transmission and evolution: A quantitative approach*. Princeton University Press, 1981.

Luigi Luca Cavalli-Sforza, Antonio Moroni, and Gianna Zei. *Consanguinity, inbreeding, and genetic drift in Italy*, volume 39. Princeton University Press, 2004.

Danilo Cavapozzi, Marco Francesconi, and Cheti Nicoletti. The impact of gender role norms on mothers' labor supply. *Journal of Economic Behavior & Organization*, 186:113–134, 2021.

Fabrizio Colella, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. Inference with arbitrary clustering. 2019.

James F Crow and Arthur P Mange. Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugenics Quarterly*, 12(4):199–203, 1965.

Pierre Darlu, Gerrit Bloothooft, Alessio Boattini, Leendert Brouwer, Matthijs Brouwer, Guy Brunet, Pascal Chareille, James Cheshire, Richard Coates, Kathrin Dräger, et al. The family name as socio-cultural feature and genetic metaphor: From concepts to methods. *Human Biology*, 84(2):169–214, 2012.

Lauren D Davenport. Beyond black and white: Biracial attitudes in contemporary us politics. *American Political Science Review*, 110(1):52–67, 2016.

Martin Dribe and Christer Lundh. Intermarriage, value context and union dissolution: Sweden 1990–2005. *European Journal of Population/Revue européenne de Démographie*, 28 (2):139–158, 2012.

Raquel Fernández. Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review*, 103(1):472–500, 2013.

Raquel Fernandez and Alessandra Fogli. Culture: An empirical investigation of beliefs, work, and fertility. *American economic journal: Macroeconomics*, 1(1):146–77, 2009.

Raquel Fernández and Joyce Cheng Wong. Divorce risk, wages and working wives: A quantitative life-cycle analysis of female labour force participation. *The Economic Journal*, 124 (576):319–358, 2014.

Raquel Fernández, Alessandra Fogli, and Claudia Olivetti. Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, 119(4): 1249–1299, 2004.

Claudia Goldin and Lawrence F Katz. The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of political Economy*, 110(4):730–770, 2002.

Jeffrey S Gray. Divorce-law changes, household bargaining, and married women's labor supply. *The American Economic Review*, 88(3):628–642, 1998.

Avner Greif and Guido Tabellini. The clan and the corporation: Sustaining cooperation in china and europe. *Journal of Comparative Economics*, 45(1):1–35, 2017.

Christina A Houseworth and Barry R Chiswick. Divorce among european and mexican immigrants in the us. *Review of Economics of the Household*, 18(1):1–25, 2020.

Morgan Kelly. The standard errors of persistence. 2019.

Margaret Kovick. Sexism in italy: a cultural phenomenon. *Wanted in Rome*, 2021.

Eric Maurin and Julie Moschion. The social multiplier and labor market participation of mothers. *American Economic Journal: Applied Economics*, 1(1):251–72, 2009.

Aviv Nevo and Adam M Rosen. Identification with imperfect instruments. *Review of Economics and Statistics*, 94(3):659–671, 2012.

Nathan Nunn and Diego Puga. Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36, 2012.

Claudia Olivetti and Barbara Petrongolo. The evolution of gender gaps in industrialized countries. *Annual review of Economics*, 8:405–434, 2016.

Paolo Passaniti. *Diritto di famiglia e ordine sociale: il percorso storico della società coniugale in Italia*. A. Giuffrè, 2011.

Robert D Putnam, Robert Leonardi, and Raffaella Y Nanetti. *Making democracy work: Civic traditions in modern Italy*. Princeton university press, 1994.

Jonathan F Schulz, Duman Bahrami-Rad, Jonathan P Beauchamp, and Joseph Henrich. The church, intensive kinship, and global psychological variation. *Science*, 366(6466), 2019.

Sanne Smith, Ineke Maas, and Frank Van Tubergen. Irreconcilable differences? ethnic intermarriage and divorce in the netherlands, 1995–2008. *Social science research*, 41(5): 1126–1137, 2012.

Douglas Staiger and James H Stock. Instrumental variables regression with weak instruments. Technical report, National Bureau of Economic Research, 1994.

Gruppo SVIMEZ. Rapporto 2009 sull'economia del mezzogiorno.

Nobuyoshi Yasuda, Luigi Luca Cavalli-Sforza, Maurice Skolnick, and Antonio Moroni. The evolution of surnames: an analysis of their distribution and extinction. *Theoretical population biology*, 5(1):123–142, 1974.

Norikazu Yasuda and Toshiyuki Furusho. Random and nonrandom inbreeding revealed from isonymy study. i. small cities of japan. *American Journal of Human Genetics*, 23(3):303, 1971.

Appendices

A Migration and surnames

The use of our two measures of surnames distribution as a proxy for the intensity of in-marriage might be biased by migration waves: a more spread surnames distribution could just indicate that individuals tend to move more intensively to that specific municipality. At the same time, out-migration might shrink the distribution of surnames in a municipality, increasing $FirstShare$ and decreasing $Entropy$: thus, if we want to use these two proxies for endogamy, we need to discuss the recent Italian history of migration. Italy was interested by different migrating waves after the Second World War, in particular from the South to the North, from the whole peninsula to foreign countries and from villages to cities. According to Istat data, those waves stopped after the economic boom, at the end of the 60s. The period between 1970 and 2000, on the other hand, registered an average of zero net migration with respect to foreign countries, and also the North-South waves decreased significantly. Afterward, at the beginning of the 2000s, Italy started experiencing influxes of migrants, especially from Eastern Europe and North Africa, which increased constantly in the following years. However, our period of focus (1993-2004) was not preceded by important waves of migration, and this decreases the probability for our measures of surnames to be biased by these factors. However, we take some further steps in order to reduce the migration concern. A first countermeasure is to include as a control in our main regressions the percentage of commuters in 2001[37], which should capture the heterogeneous tendency across municipalities' inhabitants to migrate. In addition, we perform in Section 7.1 two different robustness analyses. As we said, the most important migration waves of the recent Italian past were either from the South to the North or from small villages to big cities. Thus, migration should have increased $Entropy$ and decreased $FirstShare$ for Northern and bigger cities: this is indeed what we observe from our data. In Section 7.1 we conduct two kinds of robustness checks, performing our main regressions on subsamples of North/Center/South Italian municipalities and big/small cities. Our results are stable across all these different subsamples but the southern one, which however has coefficients with coherent signs to the other regressions but less precise in their standard deviations. This means that even if we focus exclusively on those places that people tend to leave or tend to migrate to, the effect of endogamy on the gender participation gap is still present. For instance, if we move our focus to small cities with populations lower than 5k inhabitants, which are characterized by more frequent out-migration, the cities with a higher tendency of in-marriage on average present larger gender participation gaps. We still cannot completely exclude migration to be

[37]Commuters are defined as working-age residents who move daily out of the municipality for work reasons

a source of measurement error, but all these countermeasures give us more confidence that what we are observing in our main results is mostly an effect of marriage, and not migration, patterns.

B Controls' description

For each of our OLS and 2SLS regressions we included the following list of controls at the municipality level:

Tab. 11: **Geographic controls**

Variable	Description	Source
River	Dummy for the presence of a river in the municipality's surface	Buonanno and Vanin (2017)
Lake	Dummy for the presence of a lake in the municipality's surface	Buonanno and Vanin (2017)
Elevation	Average elevation level of the municipality's surface (m)	Buonanno and Vanin (2017)
Sea distance	Average distance from the sea (m)	Buonanno and Vanin (2017)
Surface	Extension of the municipality's surface (km^2)	Buonanno and Vanin (2017)
Suit. Agriculture	Percentage of the municipality's surface possible to cultivate	http://dati.istat.it/

Notes: Description of geographic controls

Tab. 12: **Demographic controls**

Variable	Description	Source
Population	Total number of residents in the municipality	Buonanno and Vanin (2017)
Childcare availability	Number of kindergarten spots available every 100 children aged 0-2 years	http://dati.istat.it/
Gender unemployment gap	Difference in unemployment rates of men and women	http://dati.istat.it/
Gender wage gap	Difference in average hourly gross wage between men and women (private sector)	http://dati.istat.it/
Graduated women	Percentage of women holding a degree	http://dati.istat.it/
Roman Roads dummy	Presence of a major Road connecting the municipality during the Roman empire	Buonanno and Vanin (2017)
Women with a diploma	Percentage of women holding a diploma	http://dati.istat.it/
Population's age structure	Percentage of population over 65 over percentage of population under 15	http://dati.istat.it/
Percentage of commuters	Percentage of residents moving out of the municipality daily for work reasons	http://dati.istat.it/
Women with a diploma	Percentage of women holding a diploma	http://dati.istat.it/

Notes: Description of demographic controls

We acknowledge that variables such as the gender unemployment gap and gender wage gap could be bad controls. However, our results are robust in their sign, magnitude and significance level to the exclusion of these two controls.

C Validity assumptions for the instrumental variable

Following Staiger and Stock (1994), a total of 5 assumptions have to hold for an instrumental variable to be considered valid. While we previously discussed one of them, we now assess the validity of the four remaining assumptions:

1. **SUTVA**: for this assumption to hold, we would need levels of endogamy and gender gap of one municipality to not influence or be influenced by neighboring municipalities. This is unlikely to hold if close municipalities interact and influence each other's social norms and gender roles. However, we argue that for those municipalities whose treatment status is more affected by the instrument, in other words the ones living in places with high terrain ruggedness, the degree of reciprocal influence is reduced. Indeed, given the geographical isolation created by ruggedness, municipalities living in such areas are also less likely to influence each other's social norms, thus we can argue that this assumption holds for them, which are our likely compliers. [38]

2. **Random assignment**: this assumption requires the instrument to be assigned randomly, with each municipality's surface having the same probability of having a certain ruggedness level. We can assume this to hold, since ruggedness is a geographic characteristic of the terrain that is hard to be modified artificially.

3. **Monotonicity**: for this assumption to hold, we would need municipalities with higher ruggedness levels to have at least the same rates of endogamy with respect to municipalities with lower ruggedness, after controlling for other possible determinants. We argue that there are no reasons for which this assumption should not hold, due to the geographic isolation that ruggedness creates for a community: as a matter of fact, a higher irregularity of a municipality's surface increases the cost bear by a resident in case he/she searched for a partner outside his/her community.

[38]However, for municipalities living in non-rugged areas it is harder to assume that they did not influence each others' cultural heritage and customs

4. **Non-zero first stage**: after the inclusion of the controls we listed previously, our instrument should still be able to explain enough variation in the endogenous variables through which we proxy endogamy. This assumption is satisfied: here we present First-Stage's Cragg-Donald Wald F tests for the most complete specifications for both $Entropy$ and $FirstShare$:

   Cragg-Donald Wald F statistic (Entropy) : 151.22

   Cragg-Donald Wald F statistic (First share): 65.49

   As we can see, our F tests are well beyond the rule of thumb limit of 10 indicated by Staiger and Stock (1994), thus we can state that weak identification is not a problem.[39]

D Placebo test for instrument's validity

The greatest concern regarding the use of ruggedness as an instrument is that this geographic characteristic has been proven to influence economic development (Nunn and Puga (2012)). Indeed, rugged surfaces make it more difficult to cultivate, build infrastructures and transport goods, generating a scarcity of jobs. Thus, there might be consequences in terms of reducing the labor supply, since a shortage of jobs might discourage the working-age population from entering the labor force. This could invalidate our instrument: a direct effect of ruggedness on our outcome would in fact violate the exclusion restriction assumption that we discussed previously. As a first step to reduce this concern, we included a battery of controls aimed at capturing other possible effects of ruggedness on the gender participation gap. But even after the inclusion of those controls, one could still suspect ruggedness to have an additional unobserved impact on our dependent variable. To address this concern, we perform the following placebo test:

1. We first divide observations into groups for which endogamy levels are somehow stable (namely, deciles of $Entropy/Firstshare$ density functions)

[39] Other tests confirm that weak identification is not a problem even when we account for heteroskedasticity: Kleibergen-Paap rk Wald F statistic is 55.40 for Entropy and 12.30 for First Share.

2. We then perform separate regressions of the Participation gap on ruggedness plus a set of controls for each group of observations with stable endogamy levels

3. If ruggedness does not show correlations with the gender participation gap in each of the regressions, then the possibility that it has additional gender-biased effects on the labor force is lower.

The rationale of this placebo is the following: in each of these subsamples endogamy is stable, but ruggedness may vary. If variations in ruggedness are not linked to variations in the outcome, we can be more confident that our instrument does not violate the exclusion restriction.

We start by dividing the municipalities into 10 groups based on the deciles of the *Entropy*'s density function. In other words, we have 10 clusters with stable values of *Entropy*, as shown in Tab. 13:

Tab. 13: **Subsamples based on deciles of *Entropy***

Variable	Mean	Std. Dev.	Min.	Max.	N.obs.
entropy1	3.47	0.40	1.66	3.94	808
entropy2	4.16	0.12	3.94	4.36	807
entropy3	4.53	0.09	4.36	4.69	808
entropy4	4.84	0.08	4.69	4.99	807
entropy5	5.13	0.07	4.92	5.27	807
entropy6	5.40	0.08	5.27	5.55	808
entropy7	5.69	0.08	5.55	5.85	807
entropy8	6.04	0.11	5.85	6.23	808
entropy9	6.48	0.14	6.24	6.74	807
entropy10	7.40	0.59	6.74	10.19	807

Note: Summary statistics for 10 subsamples of municipalities, divided on the basis of the deciles of Entropy.

In each of these groups, we can argue the endogamy level is somehow stable, especially for the deciles from 2 to 9. We do the same with the other proxy for endogamy, creating 10 groups of observations based on the deciles of $FirstShare$. Tab. 14 describes these ten subsamples.

Tab. 14: **Subsamples based on deciles of $FirstShare$**

Variable	Mean	Std. Dev.	Min.	Max.	N
f.share1	0.68	0.21	0	0.96	809
f.share2	1.16	0.11	0.96	1.37	809
f.share3	1.60	0.13	1.37	1.83	809
f.share4	2.06	0.13	1.83	2.30	809
f.share5	2.58	0.16	2.30	2.86	809
f.share6	3.2	0.19	2.86	3.53	809
f.share7	3.96	0.26	3.54	4.42	809
f.share8	5.01	0.37	4.43	5.70	809
f.share9	6.71	0.65	5.70	8.01	809
f.share10	12.39	5.66	8.02	81.31	809

Note: Summary statistics for 10 subsamples of municipalities, divided on the basis of the deciles of First Share 2004.

Differently from the deciles of *Entropy*, for $FirstShare$ values are more stable in the first deciles because of the different distribution of this variable's density function. If we perform our main 2SLS regression on each of these 20 subsamples, the first stage regression exhibits a problem of weak instrument: the Wald F statistic is extremely low for all 20 cases, as we show on Tab.15. The low magnitude of these F-tests indicates that in each of the 20 subsamples Ruggedness, our instrument, is not able to capture enough variations in the endogenous variables (*Entropy/First Share*), and this is because our endogamy proxies are relatively stable within each decile.

Tab. 15: **Within-decile first stage F-tests**

Entropy	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Wald F statistic	0.028	0.852	0.133	0.344	0.000	1.607	0.677	0.584	0.101	3.952
FirstShare	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Wald F statistic	0.154	0.816	4.214	0.010	0.411	2.281	0.073	0.000	1.937	0.343
Dem. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geo. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
SLL FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	772	787	793	783	777	775	768	763	743	704

Note: Kleibergen-paap Wald F statistics for 2SLS regressions over 20 subsamples of municipalities based on the deciles of *Entropy* (first 10 regressions) and *First Share* (second 10 regressions). In each case, either *Entropy* or *First Share* is instrumented with *Ruggedness*. Specifications include geographic and demographic controls, together with "Sistema Locale del Lavoro" fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Next, we perform a separate OLS regression of our outcome on ruggedness for each of these 20 subsamples. On Tab. 16 we show the results:

Tab. 16: *Entropy* **and** *FirstShare* **deciles - Part. gap regressions**

	Dependent variable:									
	Participation gap 2001									
Entropy	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Ruggedness	0.002	-0.000	0.004	0.002	-0.001	0.001	0.002	0.006**	0.001	0.001
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
FirstShare	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Ruggedness	0.002	-0.000	0.004*	0.002	-0.001	0.001	0.002	0.005**	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Dem. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geo. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
SLL FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	772	787	793	783	777	775	768	763	743	704

Note: Standard OLS regressions for 20 distinct subsamples of municipalities. In each case, the dependent variable is the gender participation gap of those specific municipalities. Specifications include geographic and demographic controls, together with "Sistema Locale del Lavoro" fixed effects and *Entropy 1993* as control for the first battery of regressions, *First Share 2004* for the second one. The reduced form's coefficient for *Ruggedness* in our main 2SLS results is **0.001**, and it is significant at the 1% level. Due to missing values in the control variables, 425 out of 8090 municipalities have been dropped from the sample. Robust standard errors are clustered at the province level.
*** p<0.01, ** p<0.05, * p<0.1

In almost every case, ruggedness does not seem to have any effect on the gender participation gap other than the ones we controlled for [40]. Moreover, if we perform the same

[40] We also included *Entropy* and *FirstShare* in the regressions as controls, to exclude the possibility of presenting simply a reduced form of 2SLS

regressions but with Female Participation in the Labor Force as the dependent variable, we obtain similar results, as Tab. 17 shows.

Tab. 17: *Entropy* **and** *FirstShare* **deciles - Female Labor Force Participation**

	Dependent variable:									
	Female Labor Force Participation 2001									
Entropy	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Ruggedness	-0.005	-0.001	-0.006**	-0.001	-0.001	-0.003	-0.001	-0.003	-0.004*	-0.001
	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
FirstShare	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Ruggedness	-0.006	-0.001	-0.006*	-0.001	-0.001	-0.002	-0.002	-0.005***	-0.003	-0.000
	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Dem. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geo. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
SLL FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	772	787	793	783	777	775	768	763	743	704

Note: Standard OLS regressions for 20 distinct subsamples of municipalities. The dependent variable is female labor force participation of those specific municipalities, according to the census of 2001. All specifications include geographic and demographic controls, together with "Sistema Locale del Lavoro" fixed effects and either *Entropy 1993* (for the first battery of regressions) or *First Share 2004* (for the second battery) as controls. The reduced form's coefficient for *Ruggedness* in the 2SLS results with FLFP as dependent variable is **-0.007**, and it is significant at the 1% level. Due to missing values in the control variables, 425 out of 8090 municipalities have been dropped from the sample. Robust standard errors are clustered at the province level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The effect is significant in 4 specifications out of 20, thus it seems that there are just a few groups for which ruggedness is still able to significantly explain variations in FLFP. Therefore, there might be some additional channel reducing female labor supply that we did not consider in our battery of controls but these are biasing only a few groups of observations, namely the third decile, the eighth and the ninth only with one of the two proxies: this small effect should not be able to largely bias our results. Note that ruggedness moves together with endogamy levels in terms of mean values: cities with larger endogamy (low Entropy – high First Share) have on average higher ruggedness. In terms of variability, ruggedness' Standard Deviation is larger in the cities with higher endogamy: however, even with this larger variability, ruggedness exhibits no correlation with the Participation Gap in those deciles with higher endogamy levels, as we see on Tab.16. As a robustness check, we performed the same set of regressions with male labor force participation as dependent variable: results are in line with Tab. 17, with ruggedness significantly correlated to the outcome in only one out of 20 subsamples (see Tab.18).

Therefore, this analysis supports the soundness of the most problematic validity assumption for our instrument, which is the exclusion restriction. We provided evidence that, for subsamples of municipalities with stable levels of endogamy, variations in ruggedness are not linked to variations in the gender participation gap, after the inclusion of our battery of controls. Given the absence of additional unaccounted effects of ruggedness on our outcome, we can be more confident about the unbiasedness of the 2SLS coefficients describing the relationship between endogamy and the gender participation gap.

Tab. 18: ***Entropy*** **and** ***First Share*** **deciles - Male Labor Force Participation**

	Dependent variable:									
	Male Labor Force Participation 2001									
Entropy	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Ruggedness	-0.004	-0.003	-0.003	-0.001	-0.002	-0.001	0.000	0.000	-0.003*	0.000
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
First Share	**Decile1**	**Decile2**	**Decile3**	**Decile4**	**Decile5**	**Decile6**	**Decile7**	**Decile8**	**Decile9**	**Decile10**
Ruggedness	-0.004	-0.001	-0.002	0.001	-0.003	-0.000	0.001	0.000	-0.001	0.001
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Dem. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geo. Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
SLL FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	772	787	793	783	777	775	768	763	743	704

Note: Standard OLS regressions for 10 distinct subsamples of municipalities. In the first set of regressions, the dependent variable is the female labor force participation of those specific municipalities, while in the second set the dependent variable is male labor force participation. All specifications include geographic and demographic controls, together with "Sistema Locale del Lavoro" fixed effects and either *Entropy 1993* (for the first battery of regressions) or *First Share 2004* (for the second battery) as controls. Due to missing values inthe control variables, 425 out of 8090 municipalities have been dropped from the sample. Robust standard errors are clustered at the province level.
*** p<0.01, ** p<0.05, * p<0.1

E Relaxing the exclusion restriction assumption

Even after our comprehensive description of the included controls and the placebo analysis performed, one might still be worried about potential biases coming from unaccounted additional effects of ruggedness on the gender participation gap. In this section, we relax the exclusion restriction assumption and show how, even if we allow some correlation between the instrument and the error term, endogamy is still significantly enlarging the gap between men and women participating in the labor force. We do so by following the "Imperfect Instrumental Variable" approach proposed by Nevo and Rosen (2012). This method replaces the exclusion restriction assumption with an assumption about the sign of the correlation between the instrument and the unobservables. To be more precise, the traditional IV assumption of the instrument not being correlated with the unobserved error term is replaced with two assumptions:

1. The correlation between the instrument and the unobserved error term has the same direction as the correlation between the original endogenous regressor and the error term

2. The instrument is less correlated with the error term than the original endogenous variable

Given these two assumptions, confidence intervals and bounds for the endogenous variable's coefficient can be produced if the conditional correlation between the IV and the endogenous variable is negative.

To focus on our specific case, in the 2SLS framework we have the following two equations:

$$Partgap_i = \alpha + \beta Endogamy_i + \gamma X_i + \epsilon_i$$

$$Endogamy_i = \delta + \theta Ruggedness_i + \lambda X_i + u_i$$

The main issue with using Ruggedness as an instrument is that it could have additional unaccounted effects on the outcome, which are included in the error term:

$$\epsilon_i = \phi Ruggedness_i + \nu_i$$

The exclusion restriction imposes that $\phi = 0$, thus the unobserved error term must be uncorrelated with the instrument. The approach proposed by Nevo and Rosen (2012) allows instead to compute bounds for β even if some correlation is present. To make this possible, the conditional correlation between the IV and the endogenous variable must be negative: thus, we just focus on *Entropy* as a proxy for Endogamy, given its negative correlation with the instrument. Then, two assumptions must be met:

1. The correlation between Ruggedness and ϵ has the same direction as the correlation between *Entropy* and ϵ. We know from Section 5 that there seem to be some relevant omitted variables positively linked to *Entropy*, which influence the gender participation gap. It seems reasonable to argue that there are also additional variables positively linked to Ruggedness, which influence our outcome variable: an example of this could be the relative intensity of male versus female migration, which in turn can have an effect on the gender gap.

2. Ruggedness is less correlated with the error term than *Entropy*. This is also plausible: after the inclusion of the controls we listed on the possible additional effects of ruggedness on the gender gap (Roman roads, suitability for agriculture, age structure of the population,...), it seems likely that the correlation between ruggedness and potential elements included in the error term is lower than the correlation between our endogamy proxy and the error term.

Thus, all the assumptions seem to be satisfied, and we can relax the exclusion restriction assumption. The procedure proposed by Nevo and Rosen (2012) allows us to compute new bounds for *Entropy*'s coefficients, which we show in table 19, together with new confidence intervals.

Tab. 19: **Nevo and Rosen (2012)'s imperfect IV bounds**

	Dependent variable:			
	Participation gap 2001			
	Lower Bound(CI)	LB(Estimator)	UB(Estimator)	Upper Bound(CI)
Entropy 1993	[-2.594	(-1.488	-0.394)	-0.185]

Note: Upper and lower bounds for *Entropy*'s coefficient, computed following the Nevo and Rosen (2012) procedure, with the gender participation gap in 2001 as dependent variable. Also bounds for the 95% confidence interval are reported. *Entropy* is instrumented by the average ruggedness level, but in this case allowing for some correlation between the instrument and the unobserved error term. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

As we can see from the estimator's bounds and its confidence intervals, the effect of *Entropy* is still negative and significantly decreasing the gender participation gap. Both the upper and lower confidence interval's bounds for *Entropy*'s parameter are lower than zero, confirming the negative direction of the relationship with the outcome variable. If we take the average of the estimator's bounds, we have a value of 0.94: therefore, the magnitude of *Entropy*'s effect, in this case, seems to be reduced with respect to standard 2SLS regressions. Indeed, in our preferred specification, its coefficient was -1.5: as a consequence, we are induced to think that if we relax the exclusion restriction assumption, the magnitude of endogamy's effect on the gender participation gap is reduced. This is what we would have expected, given that in this case we are allowing for some direct effect of Ruggedness on the outcome variable, and in the standard 2SLS case, these effects were most likely biasing *Entropy*'s coefficient.

F Correction for spatial autocorrelation

Our results could be affected by the spatial autocorrelation issues described by Kelly (2019). Indeed, there could be significant spatial interdependence between local social norms and gender roles of neighboring cities, and simply clustering the standard errors at the province level could not be enough in order to address this issue. This is why we conduct a further robustness check, allowing for arbitrary dependence of the standard errors across neighboring observations. To do so, we exploit the estimator developed by Colella et al. (2019): in table 20 we present the main results with the corrected standard errors.

Tab. 20: **2SLS regressions - correction for spatial autocorrelation**

	Dependent variable:	
	Part. gap 2001	
	2SLS	2SLS
Entropy 1993		-1.565*
		(0.928)
First share 2004	0.459*	
	(0.263)	
Geographic controls	Yes	Yes
Demographic controls	Yes	Yes
SLL FE	Yes	Yes
Observations	7,605	7,605
R-squared	0.513	0.531

Note: Second stage regression in the 2SLS setting. Standard errors are corrected for spatial autocorrelation, following the procedure by Colella et al. (2019) The dependent variable is gender participation gap in 2001 (male labor force participation - female labor force participation), while the explanatory variables of interest are Entropy 1993 in the first case and First share 2004 in the second one. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 485 out of 8090 Italian municipalities. Compared to the main 2SLS regressions, 121 municipalities were dropped due to missing information on latitude and longitude.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

As we expected, the new standard errors reduce the significance level of our treatment variables. The fact that we observe a big increase in standard errors should not be surprising, as this kind of correction can generate substantially larger standard errors with respect to the standard clustered ones adopted in Tab.5. However, both *Entropy* and $FirstShare$'s coefficients remain significant at the 10% level in explaining variations of the gender participation gap. Thus, even after correcting for spatial autocorrelation, we have evidence proving that endogamy is causing larger gender participation gaps.

G Ruggedness and labor market characteristics

A recent paper by Boone and Wilse-Samson (2021) discusses the possible effects of ruggedness on labor market characteristics. According to these authors, the effects are concentrated on agriculture: more rugged areas are less suitable for mechanization and are characterized by less capital-intensive production. More sloped territories are unsuitable, for instance, for tractor use, and still need animals or humans in order to produce agricultural output. Therefore, we might think that since a more human-intensive production characterizes rugged territories, this might also affect the occupational choices of individuals, moving them to different sectors of the job market. This in turn could change the gender composition of the labor force, and thus constitute a direct channel through which ruggedness influences our outcome, violating the exclusion restriction assumption. To evaluate this possibility, we perform a series of OLS regressions with the percentages of employed in Agriculture, Industry and the Tertiary sector as main outcomes (number of occupied in the sector over the total number of occupied) and check their correlations with Ruggedness: this analysis should expose any potential link between the instrument and the degree of industrialization or ruralization of the labor market. The results of these additional regressions are shown in Tab.21.

Tab. 21: **Effect of Ruggedness on labor market structure - OLS**

Dep. Variable	**Perc_agriculture**	**Perc_agriculture**	**Perc_industry**	**Perc_industry**	**Perc_tertiary**	**Perc_tertiary**
Ruggedness	-0.003	-0.001	-0.002	-0.001	0.005**	0.002
	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)
Observations	7,621	7,621	7,621	7,621	7,621	7,621
R-squared	0.004	0.663	0.001	0.793	0.008	0.744
Geographic controls	NO	YES	NO	YES	NO	YES
Demographic controls	NO	YES	NO	YES	NO	YES
SLL FE	NO	YES	NO	YES	NO	YES

Note: Standard OLS regressions. The dependent variables are the percentages of occupied in the three major labor market sectors. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

We can observe from Tab.21 that Ruggedness is uncorrelated with the percentages of occupied in the Agricultural, Industrial or Tertiary sector. There is exclusively a significant correlation with the percentage of individuals in the Tertiary sector, but it becomes insignificant after we include the controls: this means that, within the same Local Labor Market and with the same conditions in terms of the controlled characteristics, variations in Ruggedness are not significantly linked to variations in individuals occupied in the Tertiary sector. Overall, Tab.21 suggests that Ruggedness is not influencing the structure of the labor market at the local level. We can add a further piece of evidence on this point, which is the correlation between Ruggedness and income per capita. Essentially, if there was a variation in the degree of industrialization/ruralization within the same local labor market which was correlated with Ruggedness, this should emerge from a regression with income per capita as the dependent variable and Ruggedness as one of the predictors. As we can observe from Tab.22, there is no correlation.

Tab. 22: **Effect of Ruggedness on income per capita - OLS**

	Dependent variable:		
	Income per capita		
	OLS	*OLS*	*OLS*
Ruggedness	17.662***	2.747	4.983
	(4.400)	(5.649)	(12.408)
Observations	7,726	7,726	7,726
R-squared	0.002	0.008	0.058
Geographic controls	NO	YES	YES
Demographic controls	NO	NO	YES
SLL FE	NO	NO	YES

Note: Standard OLS regressions. The dependent variable is the declared income per capita of the municipality. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level. *** p<0.01, ** p<0.05, * p<0.1

In Tab.22 we can observe how Ruggedness is not correlated to the income per capita across municipalities when we net out the effects of other geographical variables, and also when we include local labor market fixed effects and demographic variables. Thus, it seems that there are no additional effects, which we did not control for, of the asperity of the surface on the characteristics of the labor market in terms of industrialization or ruralization, and also on the level of economic development of the municipality. Therefore, we can be confident that this potential confounder, which would violate our exclusion restriction assumption, is not a source of bias for our results.

H Ruggedness and the importance of family values

The heterogeneous importance of family values can create a strong incentive for endogamy (Cavalli-Sforza et al. (2004)), but also influence the structure of the labor market (Alesina et al. (2015)) and thus potentially the gender gap in participation to the labor force. To avoid the bias coming from this unobservable, we instrument endogamy with ruggedness: this strategy is valid exclusively if the instrument is uncorrelated with the importance of family values. We can offer two pieces of evidence supporting this point, by looking at the correlations between ruggedness and two different proxies of family values. The first one is a proxy utilized by Akbari et al. (2019), which is the percentage of individuals aged 18-35 who still live with their parents. Unfortunately, we have this proxy only at the province (NUTS 2) level, thus we have a sample with 101 observations. We use this variable, related to the period 2002-2009, as the dependent variable and ruggedness plus geographic[41] and demographic[42] controls as regressors. Tab.23 shows the results.

Tab. 23: **Effect of Ruggedness on Family Ties - OLS**

	Dependent variable:		
	Family Ties proxy		
	OLS	*OLS*	*OLS*
Ruggedness	-0.411	0.461	0.222
	(0.374)	(0.386)	(0.200)
Observations	101	101	101
R-squared	0.021	0.378	0.709
Geographic controls	NO	YES	YES
Demographic controls	NO	NO	YES

Note: Standard OLS regressions. The dependent variable is the average percentage of individuals aged 18-35 that live with their parents in the time interval 2002-2009, taken at the province level. Geographic and demographic controls at the province level are included in the specification. Robust standard errors in parentheses are clustered at the regional level. *** p<0.01, ** p<0.05, * p<0.1

As we can see from Tab.23, there is no correlation at the province level between ruggedness and the percentages of young individuals who live with their parents. In addition to this, we present ulterior evidence by looking at the correlation between ruggedness and voting outcomes across Italian municipalities. The proxy of family values that we utilize in this case is the percentages of “Yes” to the abrogation of divorce in the 1974 Referendum. Divorce was introduced in Italy in 1970, and the most traditional Italian parties, together with the Catholic Church, fought for its abrogation, as it was a threat to the traditional model of family. Therefore, we can think that the percentages of “Yes” to the abrogation of divorce proxy the importance placed on family values by a municipality’s population. Thus, we regress these percentages on ruggedness plus geographic controls and local labor market fixed effects[43]. Tab.24 shows these results.

[41] Distance from the coast, suitability for agriculture, mean temperature and mean precipitation.
[42] Dummy south and population
[43] The demographic controls that we include in our main results are not available for 1970

Tab. 24: **Effect of Ruggedness on divorce referendum - OLS**

	Dependent variable:		
	%Yes to divorce abrogation		
	OLS	*OLS*	*OLS*
Ruggedness	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Observations	7,726	7,726	7,726
R-squared	0.010	0.086	0.617
Geographic controls	NO	YES	YES
SLL FE	NO	NO	YES

Note: Standard OLS regressions. The dependent variable is the percentage of votes in favor of the abrogation of divorce in the 1974 referendum. Geographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level.

Tab.24 shows clearly that ruggedness is not linked to the percentages of people that voted in favor of the abrogation of divorce in 1974: coefficients are always not significant and with a magnitude close to zero, therefore also this second proxy of family values is not correlated with ruggedness. In conclusion, the evidence we presented in this Section supports our claim that this unobservable is not biasing our results and that the exclusion restriction is preserved.

I 2SLS regression - conservative specification

The paper's main results, illustrated in Tab.5, are characterized by a degree of volatility after the inclusion of geographic and demographic controls: it might be argued that some of these controls might be endogenous and then be bad controls. We explain in Pag. 22 of the paper the variables that are most responsible for the instability of the coefficients, which are mostly the demographic controls. As for the geographical controls, they are unlikely to be endogenous, but, nonetheless, they cause the baseline coefficients to vary. This is explained by the fact that variables such as the extension of the surface or the suitability of agriculture are correlated with Ruggedness, therefore their inclusion reduces the variability in endogamy that the instrument captures. However, we can show a more "conservative" specification of our model, where we include only the most exogenous controls, namely:

- Dummy river
- Dummy lake
- Distance from the sea
- Roman Roads dummy
- Local Labor Market fixed effects

This "conservative" specification should overcome the concern about bad controls: in Tab.25 we present the baseline regressions with no covariates, the full controls ones (same specification of the main results in tab.5) and the "conservative" specification.

Tab. 25: **Effect of Endogamy on the Gender Participation Gap - 2SLS**

	Dependent variable:					
	Participation gap 2001					
Specification	Baseline	Baseline	Full	Full	Conservative	Conservative
Entropy 1993	-0.582		-1.488**		-0.720*	
	(0.562)		(0.716)		(0.412)	
First Share 2004		0.160		0.505**		0.229*
		(0.150)		(0.231)		(0.128)
Observations	7,726	7,726	7,726	7,726	7,726	7,726
R-squared	0.015	0.016	0.532	0.501	0.505	0.496
Geographic controls	NO	NO	YES	YES	YES*	YES*
Demographic controls	NO	NO	YES	YES	NO	NO
SLL FE	NO	NO	YES	YES	YES	YES

Note: Second stage regression in the 2SLS setting. The dependent variable is gender participation gap in 2001 (male labor force participation - female labor force participation), while the explanatory variables of interest are Entropy 1993 in the first case and First share 2004 in the second one. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
* Only the most exogenous Geographical controls have been included in the Conservative specifications

As we can notice from Tab.25, even in the most conservative specification, there is a significant effect of endogamy on the gender participation gap. Thus, if we net out the effects of the most exogenous geographic characteristics and within the same local labor market, variations in endogamy rates influence the size of the gender participation gap across Italian municipalities. The effect is still quite sizeable: in this conservative specification, a standard deviation increase in $Entropy$ generates a 0.15 standard deviation decrease in the gender gap, while a standard deviation increase in $FirstShare$ corresponds to a 0.16 standard deviation increase in the outcome.

J Endogamy and the gender education gap

The preservation of gender roles generated by endogamy could produce other kinds of gender gaps: here we provide additional evidence supporting this claim. We start from the gender education gap, defined as the percentage of men minus the percentage of women resident in the municipality holding at least a diploma (thus, with at least 13 years of education), as indicated by the census of 2001. The reason why we should observe an effect on the education gap is coherent with what we found in our main results: since women in more endogamous communities are less likely to participate in the labor force, it is possible that they invest less in their education. To assess whether this is true, we present in Tab. 26 the 2SLS results showing the effect of endogamy on this gender gap across Italian municipalities in 2001.

Tab. 26: **Endogamy effect on gender education gap - 2SLS**

	Dependent variable:					
	Education gap 2001					
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Entropy 1993	-0.760*		-1.041*		-1.167***	
	(0.402)		(0.550)		(0.452)	
F.Share 2004		0.208**		0.371**		0.396***
		(0.104)		(0.181)		(0.143)
Geographic controls	No	No	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	No	Yes	Yes
SLL Fixed Effects	No	No	No	No	Yes	Yes
Observations	7,726	7,726	7,726	7,726	7,726	7,726

Note: 2SLS regressions with gender education gap (rate of resident male population with at least a degree - rate of resident female population with at least a degree) in 2001 as dependent variable. The explanatory variables of interest are *Entropy 1993* in the odd columns and *First share 2004* in the even ones, in both cases instrumented with the average ruggedness level. Specifications include geographic and demographic controls, together with "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors are clustered at the province level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

We can see how higher rates of endogamy, proxied by our usual two surnames' concentration measures, are linked to larger gaps in education between men and women. We can postulate that this is also due to the higher persistence of the norm stigmatizing working women, which discourages greater investments in education on behalf of the female population.

K Endogamy and the provision of childcare

As we explained in the third section, the conservative familiar model that endogamy contributes to preserving involves some standardized chores that have to be carried out by the wife: on top of those chores, there is childcaring. In traditional families, nurturing and feeding babies during their growth process is carried out by mothers, with no need for kindergartens. With lower demand for external childcare, it is likely that local administrations will not use public money to offer this kind of service. Thus, in municipalities with higher endogamy levels we would expect a lower offer of public childcare since this activity is performed by mothers or other female family members (e.g. grandmothers). To verify whether this is true, we perform 2SLS regressions with "childcare availability" as the dependent variable, defined as "number of kindergarten spots available every 100 children aged 0-2 years"[44]. On Tab. 27 we can see the results.

Tab. 27: **Endogamy effect on childcare availability - 2SLS**

	Dependent variable:					
	Childcare availability					
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Entropy 1993	6.783***		6.783***		12.191***	
	(1.919)		(2.271)		(3.458)	
F.Share 2004		-1.860***		-2.436**		-4.169***
		(0.622)		(0.949)		(1.222)
Geographic controls	No	No	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	No	Yes	Yes
SLL Fixed Effects	No	No	No	No	Yes	Yes
Observations	7,726	7,726	7,726	7,726	7,726	7,726

Note: 2SLS regressions with "childcare availability" (number of spots in public childcare every 100 children aged 0-2 years old) in 2001 as dependent variable. The explanatory variables of interest are *Entropy 1993* in the odd columns and *First share 2004* in the even ones, in both cases instrumented with the average ruggedness level. Specifications include geographic and demographic controls, together with "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors are clustered at the province level.
*** p<0.01, ** p<0.05, * p<0.1

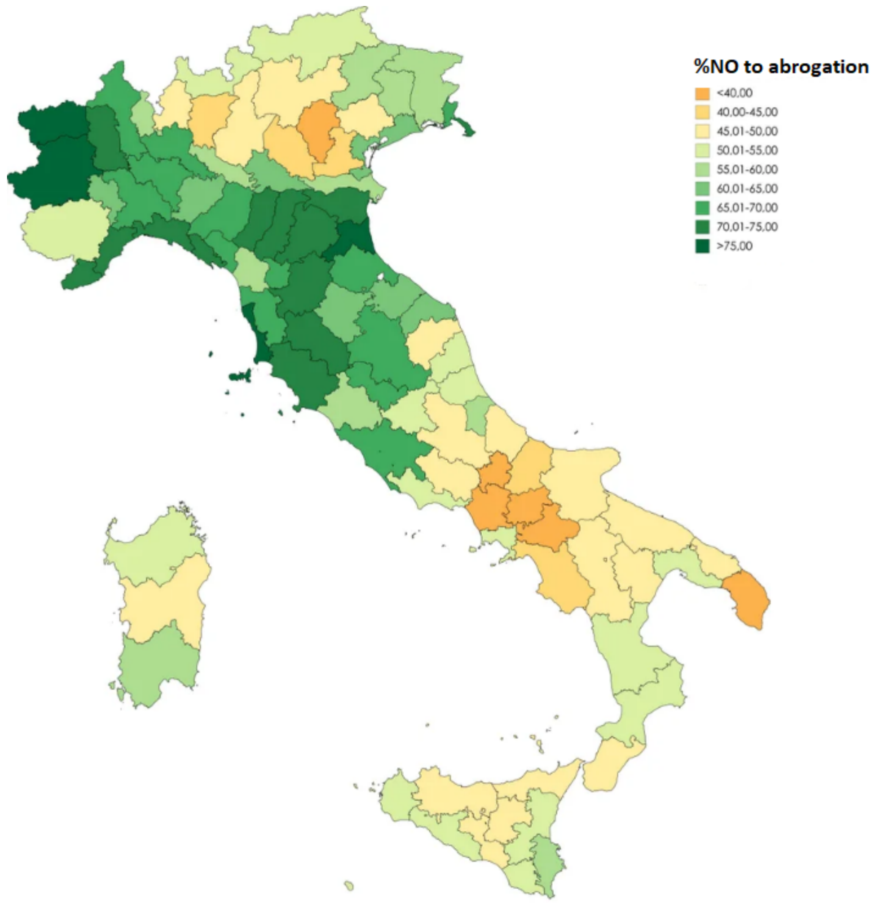
As we can observe, the more a municipality experienced in-marriage in the past, the less childcare is provided by formal institutions. We claim that the preservation of the woman's role as child-rearer in those municipalities is responsible for this negative relationship, since the need for kindergartens and other external places providing such services is reduced.

[44] On this kind of local service, each municipality's local council can autonomously decide the amount of public funds to invest. Thus, the extent to which childcare is provided is decided at the local level.

L Endogamy and divorce referendum

Divorce was introduced in Italy on the 1st of December 1970, with the law "Fortuna Baslini". It was followed by a huge amount of controversy and criticism especially on behalf of the Catholic Church and of the most popular party at the time, Democrazia Cristiana ("Christian Democrats"). Indeed, the Catholic doctrine states the indissolubility of matrimony: divorce was thus clearly perceived as a threat to Catholic values and to the traditional model of family. This controversy lead to an abrogative referendum held on the 12th/13th of May 1974, which gave the possibility to Italians to directly express their position on this matter. The turnout was high, with 87.7% of Italian voters casting their preference either for the "Yes" or "No" position (thus, for either divorce abrogation or maintenance). The "No" position won with 59.1% of preferences, with the majority of Italian supporting the decision of maintaining the divorce possibility. In Fig. 7, we can observe referendum results, with the percentages of "No" to abrogation across Italian provinces.

Fig. 7: **% NO to divorce abrogation**



Note: Percentage of voters expressing their preference for rejecting divorce abrogation in the referendum of 12th/13th May 1974. Data include all 92 Italian provinces in 1974.

Given the mechanisms discussed in Section 3, our claim is that the more endogamous communities are better able to preserve traditional social norms and gender roles. As a consequence, we would be inclined to think that these kinds of communities were more in favor of divorce abrogation, perceived as a threat to the traditional familiar model. We can evaluate this further prediction empirically: on Tab.28 we show both OLS and 2SLS results with "Percentage of No to abrogation" as the dependent variable and Consanguinity rates (both base variable and instrumented with ruggedness) as the main explanatory variables.

Tab. 28: **Consanguinity rates and divorce abrogation**

	Dependent variable:	
	% NO to abrogation	
	OLS	2SLS
Consanguinity rates	-3.31***	-13.61**
	(1.01)	(6.44)
Demographic controls	YES	YES
Geographic controls	YES	YES
Regional FE	YES	YES
Observations	92	92

Note: OLS and 2SLS regressions with percentages of "No" to divorce abrogation in the 1974 referendum as dependent variable. The explanatory variable of interest is the rate of Consanguineous marriages at the province level in the period 1945-1964, which is instrumented by the average ruggedness level in the 2SLS setting. Specifications include geographic and demographic controls, together with regional fixed effects. Units of observations are the 92 Italian provinces in 1964. Robust standard errors are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

According to Tab.28 results, voters living in provinces with more frequent consanguineous marriages were less likely to vote for divorce maintenance (and thus, more in favor of divorce abrogation). This additional piece of evidence is in line with our theoretical mechanism: the more intense the custom of endogamy within a community, the more it will stick to its traditional values and social norms. At the same time, in-marriage strengthens the social bonding between community members, and as a consequence divorces are more stigmatized in such communities, as referendum results confirm.

M Endogamy's effect on female labor force participation

We chose to use the gender participation gap as the main outcome for our main analysis because it proxies the extent of gender unbalance in the labor force at the local level. On the contrary, female labor force participation alone is not a measure of gender disparity. However, in Tab.29 we also show our 2SLS results with female labor force participation in 2001 as the main outcome. As we can see, larger endogamy is also causing lower levels of female labor supply across Italian municipalities.

Tab. 29: **Effect of endogamy on female labor force participation - 2SLS**

	Dependent variable:					
	Female labor force participation 2001					
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Entropy 1993	2.752**		3.962***		5.757***	
	(1.246)		(0.993)		(1.083)	
First share 2004		-0.755**		-1.423***		-1.953***
		(0.354)		(0.396)		(0.474)
Geographic controls	No	No	Yes	Yes	Yes	Yes
Demographic controls	No	No	No	No	Yes	Yes
SLL Fixed Effects	No	No	No	No	Yes	Yes
Observations	7,726	7,726	7,726	7,726	7,726	7,726
R-squared	0.143	0.099	0.368	0.265	0.638	0.464

Note: Second stage regression in the 2SLS setting. The dependent variable is female labor force participation in 2001, while the explanatory variables of interest are Entropy 1993 in the first case and First share 2004 in the second one. Geographic and demographic controls at the municipality level are included in the specification, as well as "Sistema Locale del Lavoro" fixed effects. Due to missing values in the control variables, we dropped from the sample 364 out of 8090 Italian municipalities. Robust standard errors in parentheses are clustered at the province level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Information and quality of politicians: is transparency helping voters?

Giovanni Righetto[1]

April 20, 2023

Abstract

In this paper, we study how voting choices are affected by giving voters more personal information on candidates right before elections. Specifically, we exploit the introduction of the "Spazzacorrotti" law in Italy in January 2019, which imposed to candidates at local elections to publish their CVs and criminal records at least 45 days before elections. By means of a Difference in Discontinuity strategy, we find no effects on elected candidates' age, gender, educational level, or ideology. Moreover, we present anecdotal evidence that candidates with a criminal record received fewer votes on average, but only in the case of local media exposure on this subject. We interpret these findings as a signal that voting choices are potentially influenced more by the criminal past of candidates with respect to other personal characteristics of politicians.

JEL classification numbers: D72, D80, H73
Keywords: Voters, Information, Local Elections, Criminal Records

[1] PhD Student in Economics, University of Bologna, E-mail: giovanni.righetto3@unibo.it.
I would like to thank Paolo Vanin, my PhD supervisor. I am also grateful to the participants in all conferences and seminars I attended for all the comments and constructive feedback. In addition, I acknowledge the financial support of the University of Bologna.

1 Introduction

The political selection process can work in many different directions and is a product of multiple mechanisms. Clearly, party leaders and electoral rules are fundamental in deciding who runs at elections and how the political support is translated into effective political power. Still, a major role in determining which candidates are governing is played by voters, who decide who are the best candidates to rule for the following years. However, it can happen that voters are uninformed or unaware of the characteristics of the candidates when they cast their vote at the polls (Djankov et al. (2010), Ferraz and Finan (2008) Larreguy et al. (2020)). Therefore, giving voters more information is a fundamental step in enhancing the democratic process of political selection and helping them choose the best candidate available. In this paper, we assess the effects of a new law, called "Spazzacorrotti" (literally "corrupts-wiper") introduced in Italy in 2019. Thanks to this law, new requirements were imposed on candidates at local elections and more information was given to voters in order to help them through the voting process. More in particular, candidates at local elections of municipalities with more than 15000 inhabitants were required to publish on the municipality's website their Curriculum Vitae and criminal records at least 45 days before elections are held. This novelty gave voters more information on the personal history and skills of candidates at local elections, possibly helping them in the screening process. Therefore, we investigate whether voters in the cities subject to the law changed their choices at the polls and selected different candidates. By the means of a Difference in Discontinuity design, we assess the effects of the Spazzacorrotti law on a bunch of characteristics of elected officials, namely education, age, gender, and ideology. We do not find any significant effect on none of the above outcomes, signaling that voters' choices were not affected by the new information published, at least in terms of these variables. Moreover, we deepen our analysis with a case study on two Southern cities, namely Lecce and Vibo Valentia. In both these cities, candidates were required to publish their criminal records (together with their CVs), in compliance with the new law. While in both cities there were candidates that had received final sentences, local media gave publicity to candidates' records only in Lecce. While we do not make causal claims on this effect, we find a negative correlation between having a “dirty” criminal record and the number of votes received only in Lecce, where local media diffused information about the candidates’ past. Candidates with a criminal past received fewer votes on average only in Lecce, even if this effect was weaker for right-wing candidates. Overall, we interpret our results as useful to understand which information is more valued by voters when they are asked to choose candidates. Indeed, our evidence tells us that while characteristics such as age and educational level do not capture voters’ attention, criminal records seem to have a much broader influence on voting choices.

The paper is structured as follows: Section 2 presents the related literature, and Section 3 describes more in detail the characteristics of the new law. Section 4 introduces the data used, and Section 5 explains in detail the identification strategy. Section 6 is dedicated to the empirical results and Section 7 to their discussion. Finally, Section 8 concludes.

2 Literature contribution

In this work, we mainly refer to the political economy literature on transparency and voting choices. Informed voters are fundamental to enhancing political selection at the polling booth, as well as making elected politicians accountable for their conduct. As an example of this, it was demonstrated how in Brazil releasing corruption audits on the misbehavior of local mayors had a significant impact on the incumbent’s performance in the following elections (Ferraz and Finan (2008)), with the probability of being re-elected negatively influenced by the degree of misconduct (i.e. the amount of embezzled public funds) followed while in charge. Moreover, having been audited in the past has been shown to reduce corruption levels by 8 percent on average (Avis et al. (2018)), possibly by increasing the non-electoral costs of engaging in corruption. Indeed, there is evidence that voters update their beliefs once given additional information about the valence or ideology of candidates, and they change their decisions accordingly (Kendall et al. (2015)). This can lead to decreased support for the worse candidates, but also to lower political participation by citizens, as knowing that politicians are corrupt can make voters withdraw from the political process. At the same time, also politicians respond to an increase in transparency about their achievements: In India, giving citizens report cards on Delhi local councilors’ policy decisions had the effect of increasing the pro-poor level of expenditures (Banerjee et al. (2020)). However, the process of delivering information to voters frequently needs to be mediated by local media: an example in this sense is provided by the case of audited Mexican mayors, who were punished at the pooling booth only in the case whether local media had reported the amount of diverted public funds (**?**). Moreover, cross-country evidence shows that the degree of financial and conflict disclosure by politicians is positively related to government quality and negatively to corruption level (Djankov et al. (2010)). In this regard, also the timing of information disclosure is fundamental: as Bobonis et al. (2016) show, information on malfeasance by local politicians must be disseminated when it's most relevant to voters to help them select better candidates. These authors study the case of audits for Puerto Rico elected officials: releasing information about audits of local mayors was proven to reduce their corruption levels only when audits were performed and published shortly before elections. Therefore, it is clear that an increase in information among voters is able to enhance politicians' accountability, helped by timely and efficient dissemination by local media. Our first contribution in this sense is showing the consequences of increased transparency on both curricula and criminal records of candidates for local elections in a developed country's context, and highlighting how consequences in electoral outcomes can be different from developing countries' settings. Moreover, the structure of the Spazzacorrotti law allows the implementation of a neat identification strategy, which was not allowed in other settings: the transparency requirements are indeed binding only for local elections of cities above a 15000 inhabitants threshold, and this grants the possibility to exploit a Difference in Discontinuity design (Grembi et al. (2016)). This is the second contribution of our work, as previous papers exploited less neat or more general identification strategies. The third and last one of our contributions is to offer evidence on what kind of information is more valued by citizens, whether the level of education, ideology, or criminal records of candidates. In this sense, we document how politicians' CV seems to be valued less with respect to their criminal records in terms of directing voting choices.

3 The Spazzacorrotti law and its implications for candidates

The law 9 January 2019 n.3, also known as "Legge Spazzacorrotti" (literally "Corrupts-wiper law") was introduced in Italy in January 2019 by the coalition government between The League and the Five Star Movement. It was designed by Justice Minister Alfredo Bonafede, a member of the Five Star Movement, in order to enhance the prevention and repression of crimes against the public administration. Italy is a country plagued by high levels of corruption with respect to other European countries, with an estimated 60 billion euros of yearly costs connected to corruption (total costs in Europe amount to 120 billion), as described by the EU report on corruption in 2021. Always according to this report, 97% of Italians think that corruption is widespread in Italy (EU average is 76%) and 90% of Italians claim that corruption is the easiest way to obtain some public services (EU average 69%). Moreover, in the 2018 legislature, 30 national politicians have been investigated for corruption-related motives. Therefore, public opinion has demanded that the governing class address this overwhelming issue, especially in the last decades. The novelties introduced by the law were diverse and multifaceted and interested many different sectors of the Italian public administration. First of all, the available tools in the hands of law enforcers were extended with the introduction of the "undercover agent" and electronic devices for wiretapping, both with the goal of discovering potential malfeasance among public administrators. In addition to that, sanctions against corrupt public servants were hardened, with larger fines and more years of detention for embezzlement, the introduction of "Daspo" (i.e. interdiction from public offices) for corrupt public servants, up to the reduction of prison benefits for inmates condemned for crimes against the public administration. Moreover, one of the most controversial parts was the reduction of the prescription terms not only for crimes against the public administration but for all kinds of processes. In Italy, trials follow three grades of sentences, and the prescription terms normally start from the date of the malfeasance. The Spazzacorrotti law suspended prescription terms from the date of the first sentence until the date of the enforceability of the sentence: as a consequence, the prescription could not happen within the dates of either the second or the third-degree sentence. As we mentioned previously, this novelty was followed by a huge amount of controversy since extending prescription terms reduced incentives for terminating the trials earlier, taking also into account that criminal proceedings in Italy last much longer with respect to comparable EU countries[2]. Last, new transparency requirements for parties were introduced. Every donation larger than 500 euros had to be reported and published online on the party's website. Moreover, every candidate for local elections of cities with more than 15000 inhabitants had to publish online, on the respective municipality's website, their curriculum vitae and criminal record at least 45 days before elections are held[3]. Regarding this last requirement, fines for non-complying parties were particularly high, ranging from a minimum of 12000 to a maximum of 120000 euros. Local newspapers showed interest in this novelty for local candidates, both in the new transparency requirements and in the content of published material, in particular criminal records[4]. In this paper, we move our focus on this last part of the new law and assess the effects of this increased transparency between voters and candidates on the characteristics of politicians elected and on electoral outcomes in general. We might expect that the increased transparency between voters and candidates raises the quality of elected officials, for instance by selecting more educated candidates. In addition, our prior is that candidates with criminal records could be punished at the polls, together with their parties and coalitions, as happened in other contexts (Ferraz and Finan (2008), **?**).

3.1 Conceptual framework: electoral consequences of disclosure

This paper focuses on the electoral consequences of increasing transparency between voters and candidates: we discuss here why we should expect any consequence, and in which direction. The political selection process that brings citizens to be elected is composed by three parts (Djankov et al. (2010)), which we argue could all be affected by the Spazzacorrotti law. The first part is the selection into politics, in which a citizen decides to enter into a political career. In this case, more transparency requirements about individuals' previous

[2] Criminal proceedings take on average 1589 days, according to a 2019 dossier published by the Italian newspaper Il Sole 24 Ore, the longest time among EU countries.

[3] Parties were held responsible for collecting this information and sending it to the local administrators in order to be published online.

[4] https://www.ilcittadinomb.it/news/cronaca/elezioni-a-desio-scoppia-la-polemica-per-i-quattro-candidati-di-forza-italia-con-la-fedina-penale-sporca/

jobs or criminal records, might make some of the less educated or more "dirty" candidates self-exclude from the political career since they might not want to disclose their personal information. The second part of the selection process is the lists' composure, in which party leaders decide who are the most suitable candidates to be elected. In this phase, party leaders might anticipate an electoral backlash against "bad" candidates due to the disclosure of curricula and criminal records, and select the most educated and "clean" ones. The last step of the political selection process is composed of voters' choices, which are influenced by their available information or biases (Kendall et al. (2015)). In this phase, granting voters more information could translate into voting "better" candidates. Thus, the main outcome for which we expect an effect of the law is the educational level of elected candidates since all the three mentioned phases should perform a selection that excludes the least educated ones after the increase in transparency. In addition, candidates with criminal records should decrease in the municipalities interested by the new law. Moreover, we might also expect an effect on gender due to the publishing of criminal records. Indeed, there is evidence that female politicians are less likely to incur in criminal activities (Brollo and Troiano (2016)): given that the new law should reduce the number of running candidates with criminal records, who are most likely to be men, the share of female candidates might increase. Party leaders or voters could not want a "dirty" candidate elected, but these candidates might also decide not to run given that they would have to make public their criminal record. Another potential channel in this direction is that publishing curricula can possibly move the attention of voters to another dimension with respect to observable characteristics such as gender (which can be inferred from the first name). As existing evidence demonstrates, voters are frequently gender-biased in their choices, also in the more developed countries (Casas-Arce and Saiz (2015). In this sense, the new law could reduce voters' gender bias, since voters' decisions might be dictated by the information contained in CVs, and not by their personal biases. Therefore, we might also expect the Spazzacorrotti law to have an effect on gender, with more women elected. On the other hand, the potential effect on age is ambiguous: we might speculate that younger candidates are less likely to have criminal charges, thus being favoured by the new transparency requirements. However, older candidates have on average richer curricula, thus they could take advantage of publishing them. Finally, we expect the effect to be mediated by the presence of local media, as we assess in Sections 6.1 and 7. In fact, for the Spazzacorrotti law to be effective it is important to deliver the information contained in CVs/criminal records to voters before they make their decisions: some authors have clearly outlined the role of local media in this sense (Ferraz and Finan (2008), Snyder Jr and Strömberg (2010), Larreguy et al. (2020)), and we expect it to be fundamental also in the Italian context.

4 Data

In order to understand the effects of the Spazzacorrotti law on elected politicians' characteristics, we collected data from multiple sources. The first dataset, named "Anagrafe degli amministratori locali", comes from the Italian Ministry of Interior and contains yearly information on elected public officials at the local level, namely each city's mayor, councillors and aldermen. Each municipality has a mayor assisted by a local council ("Consiglio comunale"), which owns the legislative power, and by an executive committee ("Giunta comunale") owning the executive power. In the Italian system, citizen over 18 years old can cast their vote for the mayor and for either one or two local councillors[5], since the electoral system prescribes semi-open lists. Each mayoral candidate can be backed by either one or more lists of council candidates, and the assignment of seats in the local councils is decided by different systems depending on the municipality's population. Under 15000 inhabitants, the mayoral candidate receiving the relative majority of votes obtains two-thirds of the seats that are allocated to his or her councillors. The remaining third of seats are allocated to other mayoral candidates via a proportional system. Seats are then attributed to councillors according to their vote ranking, which is relative to each party[6]. Above this threshold of 15000 inhabitants, the electoral system allocates council seats proportionally to candidates' lists, with the condition that at least one candidate obtained a relative majority of at least 50% of votes. When this last condition is not met, the two most-voted candidates proceed to a second ballot with another round of votes. Then, seats are allocated following a d'Hondt method.

[5] Two councillors can be chosen with the condition of having a different gender since 2013

[6] A total of 12 councillors are elected through this system.

For each politician, this first dataset contains information on their age, gender, previous occupation, educational level, and party affiliation. It is important to point out that we exclusively have information on elected officials, and not on all candidates in local elections since the Anagrafe degli amministratori local does not provide information on losing candidates. In addition, we dropped from the sample all cities below 5000 inhabitants since they are not subject to the Spazzacorrotti law and they are characterized by different electoral rules with respect to the cities above 15000 inhabitants: we kept the cities between 5000 and 15000 inhabitants to constitute our control group of elected officials. In addition, we dropped the cities belonging to special Italian regions (Sicily, Sardinia, Valle d'Aosta, Friuli, Trentino Alto Adige) since they can be characterized by different electoral rules. Moreover, we dropped from the sample politicians not elected by voters (e.g. chosen Commissioners), who might be put in charge following particular events such as the local council's dissolution. We collected data from 2002 until 2022, ending with 211916 elected officials. Tab.1 describes the characteristics of these politicians.

Tab. 1: **Summary statistics for municipalities' politicians**

Variable	Obs	Mean	Std.Dev.	Min	Max
Years of ed.	195025	14.55044	3.465206	0	22
Female	211916	.2443185	.4296834	0	1
Age	211916	45.505	11.57983	18	94
High skilled	211916	.2674173	.4426129	0	1
Right wing	211916	.1648483	.3710445	0	1
Left wing	211916	.1634846	.3698081	0	1

Notes: Summary statistics for sampled municipalities' local politicians, namely age, years of education, gender, high skill (defined with respect to their current or previous occupation), left and right-wing. Sampled municipalities are 2427 out of 8100 Italian municipalities, observed over the period between 2002 and 2022.

As we can notice from Tab.1, most of these politicians are middle aged-men, not extremely educated[7]. The fact that almost 70% of elected officials are belonging to parties not possible to locate on the left-right axis should not be surprising, since in Italy the "civic list", local

[7] The high school diploma in Italy is obtained after at least 13 years of education. This is the highest educational achievement for most of our sampled politicians.

parties with no ideological orientation, are largely diffused at local elections. We will use these data to understand the broad effects of the Spazzacorrotti law on the characteristics of elected politicians, in the first part of our analysis.

In addition, we perform an additional evaluation of the law's consequences by looking at the criminal records of candidates from two comparable cities: Lecce and Vibo Valentia. Unfortunately, criminal records of candidates for all Italian cities are not aggregated on the Ministry of Interior website but are published on each municipality's website for their respective candidates. Thus, we decided to collect all records for candidates of these two cities, since they are comparable in terms of electoral rules, local council size, and Macro-region (South of Italy). Both these two municipalities held elections in 2019, and they were subject to the Spazzacorrotti law with the new transparency requirements in force. In Lecce, the center-left coalition of Carlo Maria Salvemini was confirmed, while in Vibo the center-right coalition of Maria Limardo won. We collected data on criminal records for 835 candidates in Lecce and 444 in Vibo Valentia: among these, 10 candidates in Lecce and 5 in Vibo Valentia had received a final sentence, while other candidates had no criminal records (i.e. their criminal record certificate was empty). In Section 6.1, we explain more in detail the goal and implementation of this later part of our analysis, where we assess the electoral consequences of having a "dirty" criminal record.

5 Identification strategy

In this section, we focus exclusively on the broad effects of Spazzacorrotti on the characteristics of elected officials, while we leave the analysis of criminal records to Section 6.1. We adopt a Difference in Discontinuity strategy (Grembi et al. (2016)). The characteristics of the law allow for the evaluation of the treatment effect at the threshold of 15000 inhabitants, comparing elections that happened before and after the introduction of the "Spazzacorrotti". In other words, we will analyze the characteristics of elected politicians from municipalities close to the 15000 inhabitants' cutoff, before and after the new transparency requirements came into force. Unfortunately, at the same threshold we have another discontinuity, which is the change of the electoral system from a single-round to a double-round majoritarian

system. These two systems can produce different outcomes in terms of elected officials' characteristics: for instance, it was demonstrated how the runoff system is able to limit the presence of extremist candidates within the council (Bordignon et al. (2016)). This additional discontinuity prevents us from evaluating Spazzacorrotti's effects by means of a simple regression discontinuity design. Moreover, at different cutoffs we have also additional discontinuities, such as the size of the council at 10000 inhabitants, salaries of local politicians at 5000 inhabitants, etc. These other discontinuities could be ulterior confounders for our analysis in the case that we adopted a Difference in Difference approach, by including all sampled municipalities. On the other hand, a Difference in Discontinuity design, by focusing on a restricted group of municipalities around the 15000 inhabitants' threshold, is able to provide the best comparison between treated and control groups of elected officials.
Our model takes the following functional form:

$$Characteristic_{it} = \alpha + \beta * Treatment_i + \gamma * PostLaw_{it} + \delta * TreatPost_{it} + \zeta * normPop_i * (\eta * Treatment_i + \theta * PostLaw_{it} + \iota * Treatment_i * PostLaw_{it}) + Year_t + City_i + \epsilon_{it}$$

With:

- $Characteristic$ the characteristic of politician i elected at time t, namely either education, age, ideology or sex;
- $Treatment$ dummy for politician elected in a municipality with more than 15000 inhabitants;
- $PostLaw$ dummy for politician elected after the Spazzacorrotti law entered into force;
- $TreatPost$ interaction between dummies $Treatment$ and $PostLaw$;
- $normPop$ normalized population, equal to population-15000;
- $Year$ year fixed effects;
- $City$ municipality fixed effects.

The coefficient δ identifies the treatment effect, quantifying the consequences of the introduction of the law on treated municipalities' candidates around the threshold of 15000 inhabitants. As we are not implementing a simple Difference in Difference, we take into account only municipalities belonging to a bandwidth $[normPop - h; normPop + h]$, with h computed following Calonico et al. (2020): the bandwidth considered is the standard MSE-optimal one. As Gelman and Imbens (2019) advise, we perform only a linear fit, since higher order polynomials are less precise in identifying our treatment effect.

According to Grembi et al. (2016), our framework needs to comply with three validity assumptions in order for this strategy to correctly identify our treatment effect. The assumptions are the following:

1. All potential outcomes are continuous in the running variable at the threshold
2. The effect of the confounding policy is constant over time
3. The effect of the treatment at the cutoff does not depend on the confounding policy

The first condition is that at our threshold of 15000 inhabitants, all potential outcomes are continuous in the running variable, which is the normalized population [8]. In order to verify compliance with this condition, we perform a set of pre-treatment Regression Discontinuity designs, that should reveal whether there are unbalances in a series of municipalities' characteristics at our cutoff. The variables we use as outcomes, both time-variant and time-invariant, are the following[9]:

- South dummy
- Average ruggedness
- Dummy river
- Dummy lake
- Surface extension

[8] Normalized population here is defined as the municipality's population - 15000.

[9] The variable are taken from the website of the Istat, the Italian Institute of Statistics

- Average altitude level
- Unemployment rate (2011)
- Population density (2011)
- Female labor force participation (2011)
- Gender gap in participation to the labor force (2011)
- Percentage of land suitable for agriculture
- Percentage of population ove 60 years old (2011)

In the following tables, we can observe the coefficients of these regression discontinuity designs.

Tab. 2: **RDD with time-invariant characteristics**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	South	Ruggedness	River	Lake	Surface (km^2)	Altitude (m)	Sea distance (km)	Suit. Agri.
RD_Estimate	0.029	5.043	0.038	-0.012	20.542*	-23.723	-8.384	0.017
	(0.108)	(24.628)	(0.092)	(0.081)	(11.583)	(29.700)	(12.428)	(0.046)
Observations	2,356	2,356	2,356	2,356	2,356	2,356	2,356	2,356

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Regression discontinuity designs with municipalities' time-invariant covariates as main outcomes. The running variable is the normalized population, with 15000 inhabitants (corresponding to zero in the normalized population) as the discontinuity cutoff. Outcomes include South (dummy for municipality located in the south), average Ruggedness, River (dummy for river presence), Lake (dummy for lake presence), distance from the sea, surface extension, mean altitude level and percentage of surface that is suitable for cultivating. Regressions are performed over the pre-treatment period, in other words over the period before the Spazzacorrotti Law came into force. Sampled municipalities are 2307 out of 8100 Italian municipalities.

Tab. 3: **RDD with time-variant characteristics**

VARIABLES	(1) Un. rate	(2) Pop. Density	(3) FPLF	(4) Part. Gap	(5) Pop. over 60
RD_Estimate	0.271	-177.262	-1.265	0.621	0.117
	(1.029)	(148.047)	(1.627)	(1.066)	(0.888)
Observations	2,356	2,350	2,356	2,350	2,307

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Regression discontinuity designs with municipalities' time-variant covariates as main outcomes. The running variable is the normalized population, with 15000 inhabitants (corresponding to zero in the normalized population) as the discontinuity cutoff. Outcomes include Unemployment rate, population density, female labor force participation, gender participation gap, percentage of population over 60 years old. Regressions are performed over the pre-treatment period, in other words over the period before the Spazzacorrotti Law came into force. Sampled municipalities are 2307 out of 8100 Italian municipalities.

As we can observe from Tables 2 and 3, the only significant jump in covariates is the extension of the surface, which appears to be slightly larger above 15000 inhabitants. This unique discontinuity should not bias our results since this particular characteristic cannot influence electoral outcomes, and the variable is only marginally significant (p-value of 0.076). Moving to the second assumption, we need to make sure that the effect of the confounding policy, which is the change in the electoral system, is constant over time. We know that having a majoritarian with runoff can limit the presence of extremist politicians among the elected councillors (Bordignon et al. (2016)). To check whether this effect is constant over time, we look at percentages of elected officials from extremist parties in treated and control municipalities, over the observed time period before the introduction of the Spazzacorrotti Law. We perform a panel regression with the following specification:

$$MeanExtremists_{it} = \alpha + \beta * Treatment_i * Year_t + City_i + \epsilon_{it}$$

With:

- $MeanExtremists_{it}$ percentage of elected councillors from extremist parties in municipality i at time t ;
- $Treatment$ dummy for a municipality with more than 15000 inhabitants;

- $Year$ year fixed effects;
- $City$ municipality fixed effects.

The specification is similar to a Dynamic Difference in Difference (Vannutelli (2021)), but in this case we just observe the difference in the probability of electing extremist councillors between treated and control municipalities over time. In this way, we can observe whether there is a change over the different electoral years in the effect that Bordignon et al. (2016) observe. The unit of observation, in this case, is the municipality, and the specification includes both municipality and time fixed effects. We define extremist parties the following:

- Lega Nord ("Northern League", later changed in "The League")
- Fratelli d'Italia ("Brothers of Italy")
- Alleanza Nazionale ("National Alliance")
- Casapound
- Forza Nuova ("New Force")
- Partito Comunista Italiano ("Italian Communist Party") and its derivatives (e.g. Rifondazione Comunista)

While the first 5 parties are on the far right political spectrum, the last one (Partito Comunista) is located on the far left. In Tab.4 we show the empirical results of this test.

Tab. 4: **Probability of electing councillors from extremist parties**

VARIABLES	(1) Mean Extremists
Treated	0.080***
	(0.009)
Treated*2003.Year	-0.007
	(0.015)
Treated*2004.Year	0.010
	(0.011)
Treated*2005.Year	-0.004
	(0.014)
Treated*2006.Year	-0.001
	(0.011)
Treated*2007.Year	0.001
	(0.008)
Treated*2008.Year	-0.101***
	(0.011)
Treated*2009.Year	-0.096***
	(0.010)
Treated*2010.Year	-0.101***
	(0.010)
Treated*2011.Year	-0.083***
	(0.009)
Treated*2012.Year	-0.079***
	(0.008)
Treated*2013.Year	-0.081***
	(0.013)
Treated*2014.Year	-0.077***
	(0.010)
Treated*2015.Year	-0.078***
	(0.010)
Treated*2016.Year	-0.069***
	(0.009)
Treated*2017.Year	-0.056***
	(0.011)
Treated*2018.Year	-0.078***
	(0.015)
Observations	6,822
Number of municipality1	2,216
R-squared	0.314

Robust standard errors in parentheses are clustered at the municipality level
*** p<0.01, ** p<0.05, * p<0.1

Notes: Two-way fixed effects regressions. The outcome variable is the percentage of elected councillors from extremist parties. The variable treated identifies municipalities with more than 15000 inhabitants, those with a double-round majoritarian electoral system. Regressions are performed over the pre-treatment period, in other words over the period before the Spazzacorrotti Law came into force. Sampled municipalities are 2232 out of 8100 Italian municipalities.

We observe that there is a decreased probability of electing councillors from extremist parties since 2008, while before this year there is no significant difference in probability between treated and control municipalities. However, we can see that from 2008 until 2018 the magnitude of this differential probability has remained in line over the years, as well as the respective coefficients' standard errors. Thus, we can state that over the 11 years before the introduction of the Spazzacorrotti Law, the effect of the confounding policy, the different electoral system, has remained constant over time. As a consequence, to respect this parallel trend, we drop from our sample the observation years before 2008[10]. Note that we are not comparing the same municipalities over the years, but different electoral years: local elections are held every 5 years in the Italian system, thus the cities we observe in one year are different from the following ones, for instance. This is reinforcing the validity of the second assumption: even comparing different groups of municipalities over different years, the effect of the confounder is stable in the 11 years before the introduction of Spazzacorrotti.

The third validity assumption concerns the interaction between the confounding policy, the change in the electoral system, and Spazzacorrotti's effects. For our identification strategy to identify the correct treatment effect, we need it to not be biased by different electoral rules: in other words, increasing transparency between voters and politicians should have the same effects in terms of electoral outcome with a single round and a double round majoritarian system. This assumption is complicated to confirm empirically, but we argue that the electoral system should not bias voters' choices in this particular case. Bordignon et al. (2016) describe how, with a double round majoritarian system, moderate parties can avoid large coalitions with extremist parties in order to win, and this limits extremists' presence in local councils. Thus, the different mechanism triggered by this change in the electoral system concerns exclusively local parties' coalition strategies, but not voters' choices. In other words, an increase in transparency between voters and candidates should have the same effect on voters' choices in electoral systems with a single or double-round majoritarian system. Also, Bordignon et al. (2016) find that a runoff system influences policy outcomes, with more moderate and less volatile policies implemented. Again, the effect of the different electoral systems is on the policy outcomes but not on voters' choices. Thus, we have no reason to believe that the consequences of the introduction of the Spazzacorrotti Law could be different with a single round with respect to a double-round majoritarian system. Still, if this third validity assumption did not hold, it would mean that we are observing an average treatment effect on the treated, which is an effect valid only for those municipalities above the cutoff. On the other hand, if the third validity assumption held (and, as we argue, this is the case), we would observe an average treatment effect, valid for all our sampled mu-

[10] Keeping in the sample the years from 2002 to 2008 does not change our main results, however.

nicipalities. Therefore the third validity assumption has only consequences in terms of the external validity of our results, and not in terms of internal validity for the treated sample.

6 Empirical results

In this Section, we explore the consequences of the Spazzacorrotti law in terms of elected officials' characteristics. In line with the literature, our main prior is that, by having the chance to look at candidates' CVs, voters choose more qualitative politicians. For this purpose, we proxy the quality of politicians with their years of education, following the literature (Gagliarducci and Nannicini (2013), Galasso and Nannicini (2011)). We also explore other outcomes, such as elected officials' age, ideology and gender. We show our results in two ways: first, with pooled cross-sections of elected politicians and then with panel regressions of municipalities. In the first case, the unit of observation is the elected politician, with their characteristics as the primary outcomes of the regressions. On the other hand, in the second part of our analysis we have municipalities as units of observation, and we average the characteristics of elected councillors in order to use them as regressions' outcomes. In Tab.5 we show our first set of regressions, with pooled cross-sections of elected councillors.

Tab. 5: **Difference in Discontinuity regressions: politicians' characteristics**

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Years Ed.	Female	Age	Right wing	Left wing
TreatPost	0.066	0.003	-0.541	0.030	-0.018
	(0.246)	(0.025)	(0.792)	(0.051)	(0.036)
Year FE	YES	YES	YES	YES	YES
Observations	21,275	23,931	23,931	23,931	23,931
R-squared	0.008	0.045	0.007	0.060	0.128

Robust standard errors clustered at the municipality level
*** p<0.01, ** p<0.05, * p<0.1

Notes: Difference in Discontinuity regressions. We identify the Spazzacorrotti's effect on elected politicians' educational level, gender, age and ideological orientation. The variable Treatpost identifies the treatment effect, captured by the interaction between the variables Treatment and PostLaw. The dataset includes repeated cross-sections of elected councillors from the electoral year 2008 until 2022. The bandwidth has been chosen by the minimization of the MSE, as suggested by Calonico et al. (2020). Sampled municipalities are 2216 out of 8100 Italian municipalities.

As we can see, the Law had negligible effects on elected politicians' characteristics, namely educational level, age and gender. Also coefficients of ideological sides are not significant, indicating that the law seems to not have favored either left or right-wing candidates. In order to better understand the law's effects in terms of candidates' selection, we look at panel regressions. As we anticipated previously, in this second part of the analysis we change the unit of observation, from individual candidates to municipalities' local councils. Therefore, for each city in this case we have only one observation per electoral year. The advantage of this second analysis is that we can perform a panel Difference in Discontinuity, and thus look at how differences at the threshold of 15000 inhabitants evolved over time. Here, we introduce municipality-fixed effects: in other words, we observe a treatment effect that is within-municipality. In Tab.6 we show the results.

Tab. 6: **Difference in Discontinuity panel regressions: politicians' characteristics**

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Mean Years.Ed.	Mean Female	Mean Age	Mean RW	Mean LW
TreatPost	0.020	0.020	-0.901	0.069	-0.025
	(0.353)	(0.034)	(1.040)	(0.054)	(0.037)
Municipality FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Observations	864	867	867	867	867
R-squared	0.126	0.530	0.171	0.238	0.174
Number of municipalities	357	357	357	357	357

Robust standard errors clustered at the municipality level

*** p<0.01, ** p<0.05, * p<0.1

Notes: Difference in Discontinuity panel regressions. We identify Spazzacorrotti's effects on the elected councillors for each treated municipality belonging to the bandwidth. The variable Treatpost identifies the treatment effect, captured by the interaction between the variables Treatment and PostLaw. The dataset is a panel including data on sampled municipalities from the electoral year 2008 until 2022. The bandwidth has been chosen by the minimization of the MSE, as suggested by Calonico et al. (2020). Sampled municipalities are 2216 out of 8100 Italian municipalities, of which only 412 belong to our optimal bandwidth.

Also from the panel regressions in Tab.6, we do not see significant effects on elected councillors' education, age, gender, or ideology. Thus, even if we look at Spazzacorrotti's effects within each municipality, no change seems to have happened in terms of elected officials' characteristics. In other words, voters' choices were not affected by the availability of new information, at least in terms of the outcomes we focus on in this Section. Many reasons might explain this result, from voters being inattentive to this novelty to local media not helping to spread information: in Section 7 we discuss them. In the next subsection, we present an additional analysis on criminal records, to understand whether voters' choices were influenced by the availability of this other information.

6.1 Criminal records' analysis: the case of Lecce and Vibo Valentia

In this subsection, we move the focus of our analysis from CVs' to criminal records' content. Criminal records of candidates, together with their CVs, must be published on the municipality's website, at least 45 days before local elections. We performed webscraping in order to collect criminal records for all candidates at local elections of two comparable cities in 2019: Lecce and Vibo Valentia. Since collecting data for all treated municipalities' criminal records would have been an almost impossible task, we decided to focus on these two for various reasons. Indeed, these two cities are both from the south of Italy, and were both subject to Spazzacorrotti's new transparency requirements, having respectively 94.989 and 33.742 inhabitants. The two cities held elections on the same day, the 26th of May 2019: Lecce re-elected the center-left coalition supporting Carlo Maria Salvemini and Vibo Valentia elected the center-right coalition supporting the mayoral candidate Maria Limardo. The downloaded data contains criminal records from 835 candidates in Lecce and 444 in Vibo Valentia: as we anticipated previously, among these just a tiny fraction had actually received a final sentence. To be more precise, 10 candidates in Lecce and 5 in Vibo Valentia had a "dirty" criminal record. The kinds of sentences received by these candidates were of various types:

- Embezzlement (1 candidate)
- Bankruptcy (1 candidate)
- Slander (1 candidate)
- Drug dealing (1 candidate)
- Animal cruelty (1 candidate)

- Defamation (2 candidates)
- Fiscal evasion (1 candidate)
- Invasion of premises (2 candidates)
- Threatening (1 candidate)
- Fencing (1 candidate)
- Fraud (2 candidates)
- Violation of the obligations of family support (1 candidate)

The goal of this section is to analyze what were the consequences in terms of electoral outcomes of publishing this kind of information: were these candidates punished at the polls or did voters not give importance to their shady past? There is an important difference in the electoral context of these two cities, and it concerns local media: in Lecce, more than one local newspaper published articles regarding the content of candidates' criminal records, while in Vibo Valentia no publicity was given to the available information by local media. The "Corriere Salentino"[11] and the "Quotidiano di Puglia"[12], two local newspapers from Apulia, both published articles indicating the names of all 10 candidates in Lecce with existing criminal records, specifying the kind of sentence and the party. These two newspapers do not have any ideological tendency, and the articles were limited to reporting facts from the published records, without any opinion: indeed, candidates who had been convicted were from all different lists apart from the Five Star Movement. We can exploit this difference in media exposure to implement our identification strategy. First, we show the effect of having received a final sentence on votes received by the candidate in question by pooling data from Lecce and Vibo Valentia. Then, we perform a Difference in Difference by looking at the effect of having a "dirty" criminal record, with the local media talking about it, on votes. In other words, we have the following specification:

[11] https://www.corrieresalentino.it/2019/05/dieci-condannati-nelle-liste-leccesi-la-legge-spazzacorrotti-fa-fuori-un-candidato-di-centrodestra/

[12] https://www.quotidianodipuglia.it/lecce/elezioni$_{t}rasparenti_{d}ieci_{c}andidati_{c}on_{c}ondanne-4499568.html$

$$Votes_i = \alpha + \beta * CriminalRecord_i + \gamma * MediaExposure_i + \delta * (CriminalRecord_i * MediaExposure_i) + List_l + Coalition_c + \epsilon_{i,l,c}$$

With:

- $Votes_i$ votes received by the candidate i;
- $CriminalRecord_i$ dummy for candidate i having received a final sentence which appears in the criminal record;
- $MediaExposure_i$ dummy for having local newspapers talking about criminal records before elections;
- $List_l$ list fixed effects;
- $Coalition_c$ coalition fixed effects.

Therefore, the first difference would be having or not having a criminal record, while the second difference would be the presence of media exposure talking about it. The key identification assumption here is that media exposure is not endogenous to other potential confounders influencing the electoral outcome. This assumption is not likely to hold for many reasons. For instance, the fact that in Vibo Valentia the previous administration was right-wing could have influenced the degree of media exposure for the criminal records of candidates in subsequent elections, for the fear of being punished at the polls[13]. Moreover, media exposure might have been lower in Vibo Valentia because the number of candidates with criminal records was lower with respect to Lecce. We tend to exclude the possibility of lower political involvement or interest in Vibo: turnover in these two elections was almost equal, with 69.8% of voters casting their preference in Lecce and 67.4% in Vibo Valentia. Also, these two cities are comparable among some of the proxies of social capital used in the literature (Guiso et al. (2004)), such as the number of non-profit associations per capita (0.00541 in Vibo and 0.00611 in Lecce). Thus, it seems that in terms of citizens' civic engagement we are looking at comparable cities, and therefore it is less likely that newspapers did not report

[13] The list of candidates of the right-wing mayor Maria Limardo had two candidates with existing criminal records.

criminal records' content for a lower readers' interest. However, proving the exogeneity of media exposure would be extremely hard, therefore we cannot claim our treatment effect to be causal. Thus, we offer the following evidence as purely anecdotal, since it is unlikely to prove a causal effect of having a "dirty" candidate on electoral outcomes.

We first start from the pooled results: in Tab.7 we show the effect of having a "dirty" criminal record on votes received by the individual candidate, without taking into account media exposure.

Tab. 7: **Effect of having a "dirty" criminal record on votes**

	(1)	(2)	(3)	(4)
VARIABLES	Votes	Votes	Votes	Votes
Criminal Record	-37.871***	-38.025***	-24.645*	-17.394
	(12.416)	(13.236)	(14.143)	(13.951)
City FE	NO	YES	YES	YES
Coalition FE	NO	NO	YES	YES
List FE	NO	NO	NO	YES
Observations	1,279	1,279	1,279	1,279
R-squared	0.001	0.004	0.080	0.204

Robust standard errors in parentheses are clustered by candidates' lists
*** p<0.01, ** p<0.05, * p<0.1

Notes: OLS Regressions. The dependent variable is the number of votes received by an individual candidate, while the explanatory variable of focus is a dummy taking a value of one in case the candidate has received a final penal sentence. Candidates included are the ones who ran for the elections on the 26th May 2019, in the cities of Lecce and Vibo Valentia.

As we can notice from Tab.7, the effect of having a criminal record on votes is negative, as we would have expected. However, this effect becomes non-significant once we add List fixed effects[14]. In other words, within the same list there seem to be no consequences for having committed a crime and having been sentenced in the past[15]. As we pointed out before, this is meant to be pure suggestive evidence, since we did not include the interaction with media

[14] P-value is 0.22.

[15] I investigated whether voters, once they realize that a list ranks some individuals with a criminal record, opt for not voting for that list at all. To do so, I created a dummy "list criminal record", which takes value 1 in the case of a candidate in the list with a criminal record. The dummy has a negative effect on votes across different specifications, even without the interaction with media exposure. This is still not enough to state that voters don't vote at all for lists with criminals, but it is a signal that might lead in this direction.

exposure. Therefore, we now move to the Difference in Difference results presented in Tab.8.

Tab. 8: **Effect of having a "dirty" criminal record on votes**

	(1)	(2)	(3)
VARIABLES	Votes	Votes	Votes
Criminal Record	5.762	14.080	21.556
	(17.169)	(23.316)	(21.310)
Media Exposure	11.797	-8.634	-16.025***
	(14.610)	(13.055)	(0.000)
Criminal Record*Media Exposure	-65.697***	-58.080**	-58.134**
	(20.570)	(26.515)	(25.167)
Coalition FE	NO	YES	YES
List FE	NO	NO	YES
Observations	1,279	1,279	1,279
R-squared	0.005	0.081	0.204

Robust standard errors in parentheses are clustered by candidates' lists
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Difference in Difference results. The dependent variable is the number of votes received by an individual candidate, while the explanatory variables of focus are the dummies Criminal Record, taking the value of one in case the candidate has a criminal record, and Media Exposure, taking the value of one if local media talked about candidates' criminal records before elections. The interaction between these two dummies should capture our treatment effect, or the effect of "dirty" candidates being exposed by local media on the votes received. The candidates included are the ones who ran for the elections on the 26th of May 2019, in the cities of Lecce and Vibo Valentia.

We can see from Tab.8 that the effect of having a "dirty" criminal record on votes received is negative and significant only once interacted with Media Exposure. The average treatment effect is a decrease of 58 votes received per candidate once we include List and Coalition fixed effects. Thus, this evidence confirms the importance of local media talking about the quality of candidates in order to influence voters in their decisions at the polling booth. Still, we need to point out that the variable Media Exposure identifies all candidates from Lecce, thus it might be possible that it captures something unobserved other than the local media effect, for instance a different voters' sensibility.

Clearly, voters might also have a different sensibility on the kind of crimes, depending on their ideological position or on the candidates' one. For instance, the war on drugs has always been a priority for right-wing parties, thus a candidate coming from this political side being sentenced for drug dealing could face a stronger electoral backlash with respect to a left-wing candidate. It would be thus interesting to understand whether voters' reaction was different with respect to diverse types of sentences. Unfortunately, the number of “dirty” candidates for each offense type is not numerous enough to allow a heterogeneity analysis with respect to the kind of crimes. We will therefore exclusively perform an analysis with respect to the ideological side (right wing vs left wing of candidates), to check whether revealing criminal records generates a differential effect with respect to belonging to a specific political faction. Hence, we now conclude this Section with a heterogeneity analysis, by looking at the differential effect of having a “dirty” criminal record for candidates from either the right or the left wing. We create two dummies, center-right and center-left, identifying candidates either in the left or right-wing coalitions in both cities. We interact these dummies with our treatment effect and check whether the news about criminal records influenced differently voters from opposite ideological sides. We have 36% candidates from the left wing and 53% from the right wing in Vibo Valentia, while these percentages are respectively 34% and 50% in Lecce; the remaining candidates have no clear ideological position, being from either civic lists or from the Five Star Movement. In Tab.9 we show this heterogeneity analysis.

Tab. 9: **Effect of having a "dirty" criminal record on votes**

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Votes	Votes	Votes	Votes	Votes	Votes
Criminal Record	42.437**	49.151**	52.584**	-6.878	5.451	15.786
	(15.952)	(22.442)	(21.121)	(26.192)	(37.011)	(34.265)
Media Exposure	48.492**	-3.421	-16.025***	-18.081	-8.799	-16.025***
	(19.196)	(16.280)	(0.000)	(16.651)	(13.119)	(0.000)
Right Wing	34.536*	35.365**	-4.511***			
	(17.370)	(14.608)	(0.000)			
Crim.Record*Media Exp.	-119.159***	-116.127***	-77.390***	-31.697	-42.448	-50.958
	(22.899)	(38.444)	(27.572)	(28.174)	(38.866)	(37.308)
Crim.Record*Right Wing	-80.202***	-86.916***	-76.044***			
	(23.274)	(28.146)	(24.886)			
Media Exp.*Right Wing	-71.153**	-56.066**	-3.298***			
	(26.904)	(21.889)	(1.011)			
Crim.Record*Media Exp.*Right Wing	114.945***	118.950***	61.300*			
	(31.155)	(44.075)	(34.568)			
Left Wing				-21.969	5.176	-2.477***
				(17.454)	(11.097)	(0.000)
Crim.Record*Left Wing				33.802	21.473	14.797
				(28.437)	(38.639)	(34.265)
Media Exp.*Left Wing				85.347***	-8.519	6.360***
				(27.424)	(17.182)	(0.000)
Crim.Record*Media Exp.*Left Wing				-91.070**	-90.602**	-28.755
				(35.617)	(44.215)	(37.308)
Coalition FE	NO	YES	YES	NO	YES	YES
List FE	NO	NO	YES	NO	NO	YES
Observations	1,279	1,279	1,279	1,279	1,279	1,279
R-squared	0.031	0.082	0.205	0.055	0.081	0.204

Robust standard errors in parentheses are clustered by candidates' lists

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Difference in Difference heterogeneous results. The dependent variable is the number of votes received by an individual candidate, while the explanatory variables of focus are the dummies Criminal Record, taking the value of one in case the candidate has a criminal record, and Media Exposure, taking the value of one if local media talked about candidates' criminal records before elections. The interaction between these two dummies should capture our treatment effect, or the effect of "dirty" candidates being exposed by local media on the votes received. The heterogeneity is with respect to the ideological side of the candidate's coalition, either right or left wing. The candidates included are the ones who ran for the elections on the 26th of May 2019, in the cities of Lecce and Vibo Valentia.

The interactions with the "Right Wing" dummy yield interesting results: right-wing candidates with criminal records received fewer votes on average (double interaction), but media exposition for the offense makes the effect positive. Thus, it seems that the sensibility of voters was reduced for right-wing politicians since they were not punished as strongly as either centrist or leftist candidates. The kinds of sentences received by right-wing candidates in Lecce were serious ones, from drug dealing to threatening[16] On the other hand, the interactions with the “Left Wing” dummy do not reveal anything noteworthy: there is a negative effect for the triple interaction that becomes not significant once we add list-fixed effects.

To conclude this Section, we presented some anecdotal evidence about the importance of local media presence in the diffusion of information about the candidates. Indeed, if in any list there was a “dirty” candidate with a criminal record, he or she saw a reduction in votes only if local media talked about candidates’ records. There is no substantial evidence of an ideological bias in the punishing of “dirty” candidates, even if the effect seems to be reduced for right-wing politicians. Once again, this evidence is purely anecdotal and the impact that we find is not causal but brings more support to the theory proposed by **?**, stressing the importance of media in the electoral process of selection of public officials.

7 Discussion

In the first set of empirical results that we show, we look at potential Spazzacorrotti’s effects on the characteristics of elected politicians, and we find no significant change in gender, age, educational level, or ideology. Thus, it seems that the publication of CVs did not generate consequences in terms of selection of candidates at the polls. As we can observe only the elected officials and not all the candidates, we cannot know whether party leaders picked different individuals to compose their lists, but we can exclusively assess voters’ final choices. In other words, we see that voters’ choices did not differ after the introduction of the Law, at least in terms of the variable we considered. From our point of view, there are three possible explanations for this null result:

1. Voters did not know the information or did not check them

2. Voters saw this information, but their choices did not change

[16] Right-wing candidates in Lecce were sentenced for drug-dealing, bankruptcy, slander, animal cruelty, defamation, threatening, and fraud. Other coalitions’ candidates were instead sentenced for fraud and violation of the obligations of family support.

3. Voters saw this information and changed their choices in other ways not observable in the current dataset

We start by assessing the first possibility. Since the information on candidates' CVs was published online, we should observe differential effects in the places with either better or worse internet access. Indeed, the quality of the internet should be linked to the frequency of usage by residents (Romarri (2020)) ad therefore to the probability that they checked the new information on the municipality's website. Therefore, we perform a series of Difference in Discontinuity regressions, interacting the main treatment coefficient $TreatPost$ with some variables measuring the strength and speed of internet connection in the municipality. We use data from AGCOM (Authority for Communications Guarantees – the regulator and competition authority for the communication industries in Italy), with 2018 as the year of reference. To be more precise, we interact $TreatPost$ with a dummy indicating a fast internet speed, (i.e. above the median speed), following the strategy of Romarri (2020). Tab.10 presents this additional evidence.

Tab. 10: **Heterogeneous effects with respect to internet speed**

	(1)	(2)	(3)	(4)	(5)
VARIABLES	mean_yearsed	mean_female	mean_age	mean_RW	mean_LW
TreatPost	-0.154	0.042	-1.450	0.070	0.024
	(0.380)	(0.040)	(1.059)	(0.058)	(0.041)
TreatPost*Highspeed	0.249	-0.032	0.781	-0.002	-0.069**
	(0.263)	(0.027)	(0.794)	(0.038)	(0.035)
Observations	864	867	867	867	867
R-squared	0.128	0.531	0.173	0.238	0.179
Number of municipalities	357	357	357	357	357

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Difference in Discontinuity panel regressions. We identify Spazzacorrotti's effect on the elected councillors for each treated municipality belonging to the bandwidth. The variable Treatpost identifies the treatment effect, captured by the interaction between the variables Treatment and PostLaw. We look here at heterogeneous results with respect to the variable high speed, identifying municipalities with internet speed higher than the median value. The dataset is a panel including data on sampled municipalities from the electoral year 2008 until 2022. The bandwidth has been chosen by the minimization of the MSE, as suggested by Calonico et al. (2020). Sampled municipalities are 2216 out of 8100 Italian municipalities, of which only 412 belong to our optimal bandwidth.

From the results of Tab.10, we see that there seem to be no heterogeneous effects of the

Law with respect to internet availability. There is only a small negative effect on leftist candidates elected, but overall we do not have evidence proving that municipalities with faster internet connections were more affected by the Spazzacorrotti introduction with respect to other cities. Therefore, we tend to exclude the possibility that citizens did not check the information, or that the majority of voters did not.

We are inclined to explain our results as a combination of points 2 and 3. In other words, voters became aware of the information contained in the CVs but did not choose younger or more educated voters. Voters can change their voting decisions and punish “bad” candidates, as both Section 6.1 and the literature confirm (Ferraz and Finan (2008), Avis et al. (2018), Kendall et al. (2015)). However, the information contained in the Anagrafe degli Amministratori Locali, from which we built our dataset for the empirical analysis in Section 6, is apparently not comprehensive enough in order to understand how voters changed their choices. Information such as age, gender, education, and ideology, is indeed just a piece of the big picture that constitutes a good or a bad candidate. Criminal records can be way more influential in changing voters’ choices, as the case of Lecce seems to suggest. In this way, we are also offering evidence that voters are more sensitive about possible malfeasances of candidates than about their educational level, which has been frequently used as a proxy of the quality of politicians throughout the literature (Galasso and Nannicini (2011), Gagliarducci and Nannicini (2013)).

8 Conclusion

In this paper, we study the effects of a new law, denominated “Spazzacorrotti”, introduced in Italy on the 31st of January 2019. This law was ideated to strengthen the prevention and repression of crimes against the public administration. Among the various law’s provisions, some novelties were introduced for local elections, in order to improve political selection also at the lower levels of administrative power. Candidates at local elections of cities with more than 15000 inhabitants were asked to publish their Curriculum Vitae and Criminal Records on the municipality’s website at least 45 days before elections, with large fines for non-complying parties. Therefore, voters in these cities had more information available in order to cast their preference for mayors and local councillors. By the means of a Difference in Discontinuity analysis (Grembi et al. (2016)), we look at the effects of this increased transparency on the characteristics of elected public officials, in terms of educational level, age, gender, and ideological leaning. We find no significant results on any of these characteristics, signaling that voters were not changing their choices after the law's introduction, at least for the outcomes that we take into consideration. We move further our analysis by looking at the effects of publishing criminal records and we focus on the cases of two different Southern municipalities, Lecce and Vibo Valentia, which both held elections after the Law's introduction, on the 26th of May 2019. In these two cities, all candidates had to publish their criminal records, and some of them had previously received a final sentence for breaking the law in different ways. The difference between these two cities is that, while both had available criminal records on their website, in Lecce information on sentences captured the attention of local media while in Vibo Valentia there was no further publicity for "dirty" candidates apart from the institutional one. While we cannot establish a causal effect, we find a negative correlation between having received a final sentence and votes received only in the case of local media attention being present. Furthermore, it seems that the correlation was weaker for right-wing candidates, for whom the negative coefficient was reduced with respect to centrist and leftist candidates. Overall, we interpret these results as relevant to understand the impact of transparency on voters' choices. Voters seem to be less interested in knowing characteristics such as the educational level or the age of the candidates, while they pay more attention to the criminal records of public officials. In line with the literature (Ferraz and Finan (2008), **?**), revealing past malfeasance by politicians appears to be negatively correlated with the amount of political support that the candidate receives, also in a developed country such as Italy. For future policy implications, we advise public authorities to give importance to the role of local media in transmitting information to citizens, since voters might be unaware of potentially relevant news or public data. At the same time, local media must be responsible for delivering this information in the most transparent and unbiased way in order to enhance the process of political selection.

References

Eric Avis, Claudio Ferraz, and Frederico Finan. Do government audits reduce corruption? estimating the impacts of exposing corrupt politicians. *Journal of Political Economy*, 126 (5):1912–1964, 2018.

Abhijit Banerjee, Nils T Enevoldsen, Rohini Pande, and Michael Walton. Public information is an incentive for politicians: Experimental evidence from delhi elections. Technical report, National Bureau of Economic Research, 2020.

Gustavo J Bobonis, Luis R Cámara Fuertes, and Rainer Schwabe. Monitoring corruptible politicians. *American Economic Review*, 106(8):2371–2405, 2016.

Massimo Bordignon, Tommaso Nannicini, and Guido Tabellini. Moderating political extremism: single round versus runoff elections under plurality rule. *American Economic Review*, 106(8):2349–70, 2016.

Fernanda Brollo and Ugo Troiano. What happens when a woman wins an election? evidence from close races in brazil. *Journal of Development Economics*, 122:28–45, 2016.

Sebastian Calonico, Matias D Cattaneo, and Max H Farrell. Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210, 2020.

Pablo Casas-Arce and Albert Saiz. Women and power: unpopular, unwilling, or held back? *Journal of political Economy*, 123(3):641–669, 2015.

Simeon Djankov, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. Disclosure by politicians. *American Economic Journal: Applied Economics*, 2(2):179–209, 2010.

Claudio Ferraz and Frederico Finan. Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes. *The Quarterly journal of economics*, 123 (2):703–745, 2008.

Stefano Gagliarducci and Tommaso Nannicini. Do better paid politicians perform better? disentangling incentives from selection. *Journal of the European Economic Association*, 11(2):369–398, 2013.

Vincenzo Galasso and Tommaso Nannicini. Competing on good politicians. *American political science review*, 105(1):79–99, 2011.

Andrew Gelman and Guido Imbens. Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456, 2019.

Veronica Grembi, Tommaso Nannicini, and Ugo Troiano. Do fiscal rules matter? *American Economic Journal: Applied Economics*, pages 1–30, 2016.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. The role of social capital in financial development. *American economic review*, 94(3):526–556, 2004.

Chad Kendall, Tommaso Nannicini, and Francesco Trebbi. How do voters respond to information? evidence from a randomized campaign. *American Economic Review*, 105(1): 322–53, 2015.

Horacio Larreguy, John Marshall, and James M Snyder Jr. Publicising malfeasance: when the local media structure facilitates electoral accountability in mexico. *The Economic Journal*, 130(631):2291–2327, 2020.

Alessio Romarri. Do far-right mayors increase the probability of hate crimes? evidence from italy. *Evidence From Italy (December 6, 2020)*, 2020.

James M Snyder Jr and David Strömberg. Press coverage and political accountability. *Journal of political Economy*, 118(2):355–408, 2010.

Silvia Vannutelli. From lapdogs to watchdogs: Random auditor assignment and municipal fiscal performance in italy. *Job Market Paper: Yale University*, 2021.