# Women and Power: Unpopular, Unwilling, or Held Back? Comment

Manuel Bagues[*], Pamela Campa [†]

May 22, 2017

## Abstract

Casas-Arce and Saiz (2015) study how gender quotas in candidate lists affect voting behavior using evidence from the 2007 Spanish local elections, following the introduction of gender quotas in municipalities with more than 5,000 inhabitants. Using a differences-in-differences strategy, they show that parties that listed fewer female candidates in the previous election obtained relatively more votes in the subsequent election in larger municipalities, a pattern that they attribute to the impact of the quota. In this comment we provide a number of robustness tests, placebos, and new estimates from a regression discontinuity design suggesting that this relationship is spurious. Overall, the evidence indicates that the quota did not have an economically or statistically significant impact on the electoral performance of parties that were less feminized.

[*]Aalto University, Economics Department; email: manuel.bagues@aalto.fi
[†]University of Calgary, Economics Department; email: pcampa@ucalgary.ca

# 1 Introduction

As Casas-Arce and Saiz (2015) (hereafter, CS) point out, a better understanding of how voters react to gender quotas in candidate lists may help to shed light on the underlying causes for the underrepresentation of women in politics. If it is due to the lack of qualified women who are willing to participate in politics, or to the existence among voters of negative stereotypes about the ability of female politicians, candidates who enter politics through quotas would tend to attract fewer votes. However, if the lack of women is due to discrimination by party leaders, the introduction of gender quotas might force parties to replace some male candidates with female candidates who are more popular among voters.

CS analyze empirically the impact of gender quotas on voters' behavior using information from the introduction of quotas in Spanish local elections in 2007. In the context of a proportional representation electoral system with closed lists, a quota requiring at least 40% of candidates of each gender in candidate lists was introduced in municipalities with more than 5,000 inhabitants. In order to limit the systematic placement of the under-represented sex at the bottom of electoral lists, the quota also applied to every five-position bracket of the list. The quota increased the share of female candidates in these municipalities by around 8 p.p. (21%) (Casas-Arce and Saiz, 2011; Campa, 2011).

CS show that party lists that had fewer women in the 2003 election and, therefore, were expected to be affected to a larger extent by the quota, tended to obtain relatively better electoral results in 2007 in municipalities with more than 5,000 inhabitants, a pattern they attribute to the impact of the quota. This is an important result which would point towards the existence of agency problems in political parties as the source of the underrepresentation of women in politics. The magnitude of the estimates suggest that these agency problems are severe. Quotas would have increased the electoral support for lists that had no female candidates in the previous election by 6.6 percentage points, about 54% of a standard deviation. The size of this effect is one order of magnitude higher than the impact of other determinants of voting behavior considered in the literature, such as economic shocks.[1]

[1] For instance, Bagues and Esteve-Volart (2016) show that in Spanish general elections each additional point of GDP growth is associated to 0.3 p.p. more votes for the incumbent. In the context of U.S. presidential elections, Fair (2009) finds a figure of 0.7 p.p.

This Comment presents a re-analysis of CS. There are at least two reasons why their findings may deserve further scrutiny. First, their analysis considers a very short time span. Quotas were approved in March 2007, just two months before the May local elections. Some of the mechanisms through which quotas may help to improve candidates' quality may require a longer time span. In the short term, parties' capability to attract the most talented women to their lists may be limited and, moreover, new candidates are likely to lack political experience, a feature that may be valued by voters. More time may also be needed in order to break down negative stereotypes regarding female politicians, both among party leaders and voters, or to generate a change in the internal power structure of political parties (Beaman et al., 2009). In this respect, the large positive impact observed by CS is even more remarkable. Second, CS' empirical strategy relies on the implicit assumption that voting behavior in small municipalities provides a reliable counterfactual of what would have happened in large municipalities in the absence of the quota. This is a non-trivial assumption that might not be satisfied if the timing of political and social changes is somehow related to municipality size.

We provide a number of robustness tests, placebos and a discontinuity-in-differences analysis that cast doubts on the validity of CS' interpretation of the empirical evidence. We show that the main identifying assumption underlying CS' difference-in-differences strategy is not supported by the data. Party lists which had fewer women in the 2003 election generally obtained better electoral results in 2007 in relatively larger municipalities independently of whether these municipalities were affected by the quota. As a result, when we control linearly for population in CS' main specification, the estimated effect of quotas changes sign, becoming negative, and is not significantly different from zero. CS' results are also very sensitive to the inclusion of other additional controls, such as the share of votes that parties received in the previous election. Furthermore, using CS' main specification, we conduct a number of placebo tests considering the impact of 'fake' quotas on the sample of municipalities that had fewer than 5,000 inhabitants and, therefore, were not affected by the quota. These placebo tests, conducted at 100-intervals between 1,000 and 4,000 inhabitants, yield statistically significant results in over 80% of the cases and the estimates are of a similar magnitude to the effect found by CS at the 5,000 threshold.

CS also study voting behavior within a subset of municipalities where they expect the impact of quotas to be particularly relevant. They focus on municipalities where there were two main parties and one of the parties, in 2003, had less than two women within the top 5 positions, i.e. the *male holdout*. CS show that *male holdouts* tended to obtain more votes in 2007 in municipalities subject to the quota, but our re-analysis indicates that these results are also sensitive to the inclusion of additional controls, the way standard errors are computed, and an arguably more reasonable definition of *male holdouts*.

Finally, we exploit the existence of a population threshold that determines which municipalities are subject to the quota to estimate the impact of quotas on voting behavior using a discontinuity-in-differences approach, which relies on milder identifying assumptions than the differences-in-differences strategy adopted by CS. According to this analysis, quotas did not have a economically or statistically significant impact on voting behavior. Overall, this comment provides a cautionary note about the need to conduct standard robustness tests in studies that rely on a differences-in-differences approach.

# 2 Replication

We use the dataset provided by CS, which includes electoral information for all party lists that participated in Spanish local elections in 2003 and 2007 in municipalities with more than 250 and less than 10,000 inhabitants. Additionally, we complement this dataset using information on the electoral results obtained by these lists in the 1999 election.[2]

The structure of the replication is as follows. First, we reexamine the main analysis of CS, which considers all party lists in the dataset (see CS *Section 4.C: Relative Growth of Female Candidates and Vote Share*). Second, we analyze the impact of quotas on the subsample of *bipartisan male holdouts* (see CS *section 4.B: A Natural Experiment Induced by the Quota: "Bipartisan Male Hold-out" Lists*). Finally, we provide novel evidence on the impact of gender quotas on voting behavior using a discontinuity-in-differences approach.

## 2.1 Impact of Quotas on Electoral Results: All Lists

We focus on the reduced form estimation reported by CS in Table 5, column 1. To estimate how the introduction of quotas in 2007 affects the electoral results of party lists that had relatively fewer

[2]Source: Spanish State Department (*Ministerio del Interior)*

women in the 2003 election, CS estimate the following equation:

$$\begin{aligned}\Delta Votes_{pm} =& \beta_0 + \beta_1[\max\{0, \text{quota - } female^{2003}_{pm}\} \times large_m] + \\ &+ \beta_2 female^{2003}_{pm} + \beta_3 (female^{2003}_{pm})^2 + \lambda_p + \phi_m + \varepsilon_{pm}\end{aligned} \tag{1}$$

where $\Delta Votes_{pm}$ stands for the change in the vote share received by party $p$ in municipality $m$ between 2003 and 2007; *quota* is equal to 40%, $female^{2003}_{pm}$ is the share of female candidates on the list in 2003; $large_m$ is an indicator that takes value one in municipalities with more than 5,000 inhabitants, and $\phi_m$ and $\lambda_p$ are, respectively, municipality and party fixed-effects. Each observation is weighted by the vote share obtained by the list in the previous election and standard errors are clustered at the regional level. In what follows we refer to municipalities with more and less than 5,000 inhabitants as *large* and *small* municipalities respectively.

As expected, our replication using the same specification and the same dataset as CS provides identical results. The further a list is from the 0.40 threshold in 2003, the larger the improvement in its electoral performance in 2007 in large municipalities, relative to the performance of similar lists in smaller municipalities (see Table 1, column 1). The magnitude of the effect is substantial: lists that had no women in 2003 obtain 6.6 p.p. more votes in municipalities subject to the quota, about 54% of a standard deviation.[3] This estimate would capture the causal impact of the quota under the usual parallel trends assumption: the evolution of electoral outcomes in small municipalities provides a reliable counterfactual of what would have happened in large municipalities absent the quota. While this assumption is essentially untestable, below we examine several standard robustness tests that were not reported in CS.

**Parallel trends** A standard way to investigate the validity of a difference-in-differences strategy is to compare the evolution of the outcome variable in previous periods in the treatment and the control group. The existence of parallel trends in the past is neither sufficient nor a necessary condition for consistency per se, but it is typically considered as supportive evidence. We estimate equation (1) using as the left-hand side variable the variation in the share of votes obtained by each

[3]The standard deviation of the variation in the share in votes is equal to 0.12.

party list between 1999 and 2003. As shown in column 5 of Table 1, the evidence is consistent with CS' empirical strategy. The evolution of electoral results between 1999 and 2003 does not vary significantly between large and small municipalities.[4]

**Placebos** The fact that municipality size did not play a significant role in the previous election does not necessarily guarantee that this applies also to the 2007 election. In order to examine the plausibility of the identifying assumption, we estimate placebo regressions in the subsample of municipalities that were not affected by the quota (municipalities with less than 5,000 inhabitants). More precisely, using equation (1), we study the 'impact' of placebo quotas at all possible cutoffs between 1,000 and 4,000 inhabitants, at increments of 100. We report these results in Figure 1. The placebo analyses yield significant positive effects in over 80% of the cases and the magnitude of the estimates is comparable to CS' findings at the 5,000 threshold. These results strongly suggest that CS' main identifying assumption is unlikely to hold in the 2007 elections.

**Additional controls** The above placebo tests suggest that the electoral performance in 2007 of party lists that had relatively fewer women in the 2003 election differed depending on municipalities' size. To explore this possibility, we augment equation (1) by adding a linear interaction between population size and how far away the list was in 2003 from satisfying the quota requirement (in bold below).

$$\begin{aligned} \Delta Votes_{pm} = {} & \beta_0 + \beta_1[\max\{0, \text{quota - } female^{2003}_{pm}\} \times large_m] + \\ & + \beta_2\boldsymbol{[\max\{0, \textbf{quota - } female^{2003}_{pm}\} \times population_m]} + \\ & + \beta_3 female^{2003}_{pm} + \beta_4 (female^{2003}_{pm})^2 + \lambda_p + \phi_m + \varepsilon_{pm} \end{aligned} \tag{2}$$

As shown in Table 1, column 2, the evolution of voting behavior between 2003 and 2007 is captured better by a linear function of population than by the *large* municipality dummy. In fact, once we control for population linearly, being in a municipality with more than 5,000 inhabitants has, if anything, a negative impact on the electoral performance of party lists that had fewer

[4] Ideally we would like to conduct also a similar exercise taking into account the share of female candidates in the 1999 elections. Unfortunately, this information is not, up to the best of our knowledge, readily available.

women in the 2003 election, although this effect is not statistically significant at standard levels. To investigate further why population size correlates with electoral performance, we also include in the specification the share of votes received by the list in the 2003 election. This variable is highly predictive of the variation in electoral support between 2003 and 2007 and, moreover, it 'explains' why parties that had fewer female candidates in the 2003 election obtained relatively better results in 2007 in large municipalities (see Table 1, column 3). In sum, once we control either for population size or for the share of votes obtained by the party list in the previous election, there is no support for the claim that quotas significantly affect voting behavior.

### 2.1.1 Other minor issues

There are several additional potential concerns with the estimation of equation (1) which we discuss below. As we show in Table 1, column 4, these considerations do not affect the main conclusions of the above analysis.

**Clustering of standard errors when the number of clusters is small** CS cluster standard errors at the regional level using the standard clustered sandwich estimator, in a context where there are only 17 regions. As pointed out by Cameron et al. (2008), when the number of clusters is small, this procedure can overstate the precision of the estimator. To investigate the relevance of this issue, we also report bootstrap cluster-robust p-values.

**The actual quota is equal to 6/13 (around 46.2%), instead of 40%.** Due to indivisibilities, the minimum share of candidates of either gender required by quotas is above 40%. More precisely, the quota required that in municipalities with more than 5,000 inhabitants at least 6 out of the 13 candidates are female (and male). Therefore, it might be more reasonable to estimate equation (1) taking into account a quota of 6/13 (around 46.2%) instead of 40%.

**Different functional forms above and below the 5,000 threshold.** Equation (1) fits different functions of population above and below the 5,000 inhabitants threshold. More precisely, while in small municipalities it fits the function $[\beta_2 female_{pm}^{2003} + \beta_3 (female_{pm}^{2003})^2]$, in large municipalities it also includes $\beta_1[\max\{0, \text{quota - } female_{pm}^{2003}\}]$. To limit the potential impact of specific functional form assumptions on the identification, it might be convenient to consider the same functional form above and below the 5,000 threshold, for instance allowing also that in small municipalities electoral performance varies with [$\max\{0, \text{quota} - female_{pm}^{2003}\}$].

**Voters from small municipalities receive a disproportionate weight.** CS weight observations by the vote share obtained by the list at the municipal level in the previous election. In practice, this implies that the weight given to each voter is inversely related to the size of the municipality. As a result, voters in municipalities with less than 1,000 inhabitants, which account for approximately 10% of the total number of voters in the sample, receive close to half of the weight in the regression. This might be particularly problematic if these small municipalities provide a less reliable counterfactual for what would have happened in municipalities with more than 5,000 inhabitants in the absence of the quota. In order to give the same weight to all voters, it may be more appropriate to use as weights the number of votes instead of the share.

## 2.2 Impact of Quotas on Electoral Results: *Bipartisan male holdouts*

The above analysis suggests that, in general, the differential evolution in the electoral results of parties that had fewer female candidates in 2003 in large and small municipalities cannot be attributed to the introduction of the quota. However, it is still possible that quotas affected voting behavior in some specific types of municipalities.

CS study the impact of quotas on "a subset of lists in which the competitive environment was well defined and the quota was uniquely binding" (page 651). In particular, they consider municipalities where political competition is limited to two main parties ('bipartisan towns') and "one of the main parties was initially fielding less than 40 percent of females [**in the top five positions**] (we call these lists 'male holdouts') while their main competitor was already above that threshold in the 2003 elections" (CS, page 652).[5,6]

[5] In particular, CS select municipalities where two parties obtained 80% of the seats in the previous election. A possible drawback of this criterion is that, due to indivisibilities, it establishes different thresholds for municipalities of different size. For instance, in municipalities with more than 1,000 and less than 2,000 inhabitants, it requires that the main two parties have obtained 8 out of the 9 seats of the council (89%), while in municipalities with more than 2,000 and less than 5,000 inhabitants it requires that the main two parties should have at least 9 out of the 11 seats (82%). A possible alternative criteria, which does not significantly affect the analysis, would be to define bipartisan municipalities in terms of the share of votes obtained by the main two parties instead of the number of seats.

[6] We have added the text in bold to the above definition to make it more clear. Otherwise, the definition provided by CS seems to suggest that they consider the share of women in the overall list (the same problem applies also to

In order to study the impact of the quota on the electoral performance of *bipartisan male holdouts*, CS implement a difference-in-differences strategy. More specifically, CS estimate the following equation:

$$\Delta Votes_{pm} = \beta_0 + \beta_1 large_m + \varepsilon_{pm} \tag{3}$$

where $\Delta Votes_{pm}$ is the variation between 2003 and 2007 in the electoral performance of the male holdout party list $p$ in municipality $m$, and *large* is an indicator for municipalities with more than 5,000 inhabitants.

CS estimate this equation in two different samples. In both cases the 'treatment' group includes male holdouts in municipalities with more 5,000 inhabitants but, as the control group, CS consider two alternatives: (i) a control group including all male holdouts in small municipalities and (ii) a subsample of male holdouts in small municipalities matched by propensity score in terms of the electoral results in 2003 (share of votes, herfindahl index, participation rate), the share of women within the top 5 positions in 2003, and party dummies.[7] CS report standard errors clustered by region for the general sample, and OLS standard errors for the matched sample.

As expected, our estimation using the same definition of male holdouts and the same samples as CS provides similar point estimates. Male holdouts in large municipalities obtain 2.6 p.p. more votes compared to the unmatched sample of male holdouts in smaller municipalities and 4.2 p.p. more votes than the matched sample. We report this analysis in Table 2, column 1, panels A and B.[8]

A potential threat to the validity of this analysis is the possibility that the treatment and the control group differ in some relevant dimension that somehow was correlated with electoral performance in 2007. We analyze the robustness of results to the inclusion of controls, taking into account the set of variables used in the matching procedure and two additional variables that were not considered but nonetheless might be relevant: the share of women in the list in 2003 and

the definition provided in the notes to Table 2). Instead, the authors only consider in their calculations the top five positions of the list, without taking into account whether the rival of the male holdout satisfies the requirement of including at least 40% of women in the overall list. According to our calculations, this condition is not satisfied by 49% of the lists.

[7] In particular, CS match each of the 144 male holdouts in the quota municipalities with the 10 closest ones in non-quota municipalities.

[8] CS report these results in Table 4, column 1 and in Table 3, column 4, row 5.

population size.[9]

The inclusion of controls leads to a decrease in the magnitude of the estimates and, in the case of the matched sample, the impact of quotas is not statistically different from zero (Table 2, panels C and D, column 1). We also report wild-bootstrap cluster-robust p-values to account for the possible inconsistency of the estimated clustered standard errors, which CS calculate using the standard sandwich estimator in a context when there are only 17 clusters. These p-values tend to be larger and they are above the standard 5% level both for the matched and for the unmatched sample.[10]

### 2.2.1 Definition of male holdouts

CS argue that "(t)owns with a male holdout list are interesting to study because only one of the major parties was affected by the quota" (page 652). Given that, in practice, CS' definition of male holdouts only takes into account the top five positions of the ballot (see footnote 6), this statement implicitly assumes that voters are not affected by the identity of candidates placed below the fifth position. In order to relax this assumption, we examine a stricter definition which considers only municipalities where, in 2003, one of the two main parties satisfied all the quota requirements while the other party did not satisfy at least one of the quota requirements.

Furthermore, we propose two additional definitions of male holdout that also take into account the information provided by bipartisan towns where neither of the two main parties satisfied all the requirements of the quota in 2003, but one of them was less feminized and, plausibly, more affected by the introduction of the quota. The first definition is based on the share of women in the overall list. More precisely, we consider municipalities where one of the main two parties has less than 40% of women in the list and it has fewer women in the list than the rival list (independently of whether the rival list satisfies the 40% requirement). We denominate this group 'lists with fewer female candidates'. Similarly, it is also possible to consider, in the spirit of the definition of male holdout implemented by CS, a similar definition that takes into account only the top five positions

[9] CS take into account in the matching procedure the share of women in the top 5 positions (which they denominate 'Initial share of women on list'), but not the overall share of women in the list. Our analysis of the database reveals that male holdouts in large municipalities had in 2003 0.07 more female candidates than the matched control group, a difference that is highly significant.

[10] None of the estimates is significant either if we use OLS standard errors, heteroskedasticity-robust standard errors, or if we cluster standard errors at the provincial level (N=50).

of the list. Thus, in the second definition, we consider municipalities where one of the two main parties has one or zero women within the top 5 positions of the ballot and its rival has a larger number of female candidates within the top 5 positions. We call this group 'lists with fewer top 5 female candidates'.[11]

In order to understand how useful these four different definitions of male holdouts are for identifying exogenous variations in the presence of female candidates in the list, we estimate equation (3) using as the dependent variable either the share of women in the list or the share of women within the top 5 positions. As shown in Table A1, the four definitions help to identify a significant increase in the degree of feminization of male holdout lists in large municipalities, but the more general definitions that we propose, i.e. 'lists with fewer female candidates' and 'lists with fewer top 5 female candidates', rely on a larger sample and they help to identify a stronger and a more precise variation in the degree of feminization of male holdouts.

We re-execute CS' analysis using these three alternative definitions of male holdouts. Estimates tend to be slightly smaller, but in general significantly positive, when no controls are considered (see Table 2, panels A and B, columns 2-4). However, when we take into account controls, none of the estimates are statistically different from zero (panels C and D, columns 2-4).

## 3 Discontinuity-in-Differences Design

CS' difference-in-differences estimates are very sensitive to the inclusion of controls, reflecting the existence of relevant time-variant differences in voting behavior between small and large municipalities. A natural way to deal with this problem is to exploit the design of the quota, which was implemented based on a precise population threshold, to estimate a regression discontinuity design (RDD). In order to allow for the possibility that there exist some time-invariant differences in voting behavior between small and large municipalities, we follow a discontinuity-in-differences approach and we consider the outcome variable in differences ($\Delta Votes_{pm}$). In particular, we compare how the share of votes received by male holdouts evolves in municipalities slightly above and below the

[11]The difference between this definition and the one proposed by CS is that this definition also includes municipalities where one of the main two parties had no women within the top 5 positions and the other party had only one woman.

5,000 inhabitants threshold using the following equation:

$$\Delta Votes_{pm} = \beta_0 + \beta_1 large_m + \beta_2 f(population_m) + \varepsilon_{pm} \tag{4}$$

where the running variable is $population_m$, the dummy variable $large_m$ denotes treatment status, and we estimate a local linear regression with triangular kernel using the optimal bandwidth proposed by Calonico et al. (2016). We conduct the analysis separately for each of the four different definitions of male holdouts discussed in the previous section.

This approach provides a consistent estimate of the impact of gender quotas in municipalities that have around 5,000 inhabitants under the assumption that there are no other relevant factors that experience a discrete change at this threshold. There are two potential threats to the validity of this empirical strategy. First, some municipalities might try to manipulate their population counts in order to avoid (or to qualify) for this policy. This is unlikely to be a problem in this case given that the law was passed in March 2007 and it was implemented based on the official population in January 2006. Another potential threat to the validity of this strategy is the existence of other policies that rely on the 5,000 population threshold. These policies might potentially have a direct impact on voting behavior or they might induce a manipulation of population figures (Eggers et al., 2015). As Bagues and Campa (2017) explain in detail, while there are some policies that are implemented based on the 5,000 threshold, no relevant changes occurred during this period that had a differential effect above and below the threshold.

Before turning to the impact of the quota on electoral results, we analyze the impact of quotas on the presence of women in the list ($\Delta Female_{pm}$). The optimal bandwidth is somewhere between 1,000 and 2,000 inhabitants, depending of the definition of male holdout considered. In all four cases, the introduction of the quota leads to a significant increase in the share of female candidates in large municipalities of approximately 10 p.p. (Table 3, upper panel, columns 1-4). However, when we consider as the outcome variable the variation in the share of votes received by male holdouts ($\Delta Votes_{pm}$), we do not observe a significant impact (Table 3, lower panel, columns 1-4). Depending on the specification, quotas may increase the support for male holdouts by 2.9 p.p. or they decrease it by 1.3 p.p., but none of these estimates are statistically significant. The RDD graphs in Figure 2 provide further evidence of these findings. These estimates suggest that, overall, there is not enough evidence to conclude that the increased feminization of lists brought about by the quota had any significant impact on voting behavior, at least in municipalities with approximately 5,000 inhabitants.

## 4 Conclusion

In this Comment, we re-analyze CS' study of the impact of gender quotas on voters' behavior using data from the 2007 Spanish local elections, when candidate quotas were introduced in municipalities with more than 5,000 inhabitants. CS show that party lists that had fewer women in their ranks in the previous election obtained more votes in the subsequent election in municipalities subject to the quota. This is an important finding that would suggest that the lack of women in politics is due to discrimination by political parties as opposed to a lack of qualified women who are willing to participate or negative stereotypes concerning female politicians among voters.

We present a number of robustness and placebo tests that cast doubts on the validity of CS' empirical strategy: party lists that were previously less feminized obtained better electoral results in larger municipalities regardless of the quota. Furthermore, we exploit the design of the quota to implement a discontinuity-in-differences analysis which fails to reject the null hypothesis that quotas did not affect voters' behavior. Overall, our re-analysis of CS suggests that it is not possible to conclude based on the available evidence that quota candidates tend to attract more votes.

There are at least two possible explanations for our findings. Quotas raised the presence of women among candidates and council members but they did not help women reach powerful positions such as party leader or mayor, which might be what voters ultimately care about (Casas-Arce and Saiz, 2011; Campa, 2011). Moreover, the elections took place only two months after the approval of the quota. Some of the mechanisms through which quotas affect voting behavior may require more time. An analysis of the longer term effects of quotas on voting behavior may be needed in order to get a better understanding of why women have failed historically to achieve equal representation with men in politics and whether quotas can help to correct this problem (Bagues and Campa, 2017).

## References

Bagues, M. and P. Campa (2017). Can gender quotas in candidate lists empower women? Evidence from a regression discontinuity design. *Mimeo*.

Bagues, M. and B. Esteve-Volart (2016). Politicians' luck of the draw: Evidence from the spanish christmas lottery. *Journal of Political Economy 124*(5), 1269–1294.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: female leadership and gender bias. *Quarterly Journal of Economics 124*(4), 1497–1540.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016). Rdrobust: Software for regression discontinuity designs. *University of Michigan*.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Campa, P. (2011). Gender quotas, female politicians and public expenditures: Gender quotas, female politicians and public expenditures: Quasi-experimental evidence. Working Paper no. 157, Econpubblica, Universita Bocconi.

Casas-Arce, P. and A. Saiz (2011). Women and power: Unwilling, ineffective, or held back? *IZA Discussion Paper No. 5645*.

Casas-Arce, P. and A. Saiz (2015). Women and power: Unwilling, ineffective, or held back? *Journal of Political Economy 123*(3), 641–669.

Eggers, A. C., R. Freier, V. Grembi, and T. Nannicini (2015). Regression discontinuity designs based on population thresholds: Pitfalls and solutions. *IZA Discussion Paper Series 9553*.

Fair, R. (2009). Presidential and congressional vote-share equations. *American Journal of Political Science 53*(1), 55–72.

**Table 1: Impact of Quotas on Electoral Results: All Lists**

| Dependent variable: | Δ Vote share (2003-2007) | | | | Δ Vote share (1999-2003) |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| (Quota - $female^{2003}$) x large | 0.166 | -0.076 | -0.015 | 0.046 | -0.078 |
| | (0.056)*** | (0.096) | (0.054) | (0.089) | (0.069) |
| (Quota - $female^{2003}$) x population | | 0.043 | 0.018 | 0.001 | |
| | | (0.020)*** | (0.015) | (0.023) | |
| Vote share in 2003 | | | -0.222 | -0.230 | |
| | | | (0.013)*** | (0.018)*** | |
| Bootstrap p-values | 0.00 | 0.328 | 0.811 | | 0.176 |
| Weights | Vote (%) | Vote (%) | Vote (%) | Votes | Vote (%) |
| Quota | 0.4 | 0.4 | 0.4 | 0.46 | 0.4 |
| Adj. R-squared | 0.421 | 0.422 | 0.479 | 0.435 | 0.469 |
| N | 11562 | 11562 | 11562 | 11341 | 10263 |

*Note*: *Large* is a dummy variable that takes value one for municipalities with more than 5,000 inhabitants, and $female^{2003}$ is the share of women in the list in the 2003 election. All regressions include controls for a quadratic polynomial of female share in 2003, municipality fixed effects and party fixed effects. Column (4) also includes *(Quota - $female^{2003}$)* as a control. Standard errors clustered by region (N=17) in parenthesis. Bootstrapped p-values are cluster-robust bootstrap p-values for the estimator of *(Quota - $female^{2003}$) x large*. Significance levels: 1% ***, 5% ** and 10% *

**Table 2: Impact of Quotas on Electoral Results: Bipartisan Male Holdouts**

| Male holdouts: | **CS definition** | **Strict definition** | **Fewer Female Candidates** | **Fewer Top 5 Female Candidates** |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | | *Panel A. Unmatched sample; No controls.* | | |
| Large municipality | 0.026 | 0.030 | 0.020 | 0.020 |
| | (0.011)** | (0.012)** | (0.008)** | (0.012) |
| N | 1498 | 1127 | 2343 | 1927 |
| | | *Panel B. Matched sample; No controls.* | | |
| Large municipality | 0.042 | 0.025 | 0.022 | 0.034 |
| | (0.008)*** | (0.010)*** | (0.007)*** | (0.007)*** |
| N | 735 | 530 | 1085 | 851 |
| | | *Panel C. Unmatched sample; With controls.* | | |
| Large municipality | 0.022 | 0.003 | -0.001 | 0.009 |
| | (0.010)** | (0.015) | (0.007) | (0.008) |
| Bootstrap p-values | 0.077 | 0.878 | 0.941 | 0.273 |
| N | 1498 | 1127 | 2343 | 1927 |
| | | *Panel D. Matched sample; With controls.* | | |
| Large municipality | 0.021 | -0.009 | -0.001 | 0.010 |
| | (0.015) | (0.018) | (0.012) | (0.014) |
| Bootstrap p-values | 0.138 | 0.646 | 0.816 | 0.446 |
| N | 735 | 530 | 1085 | 851 |

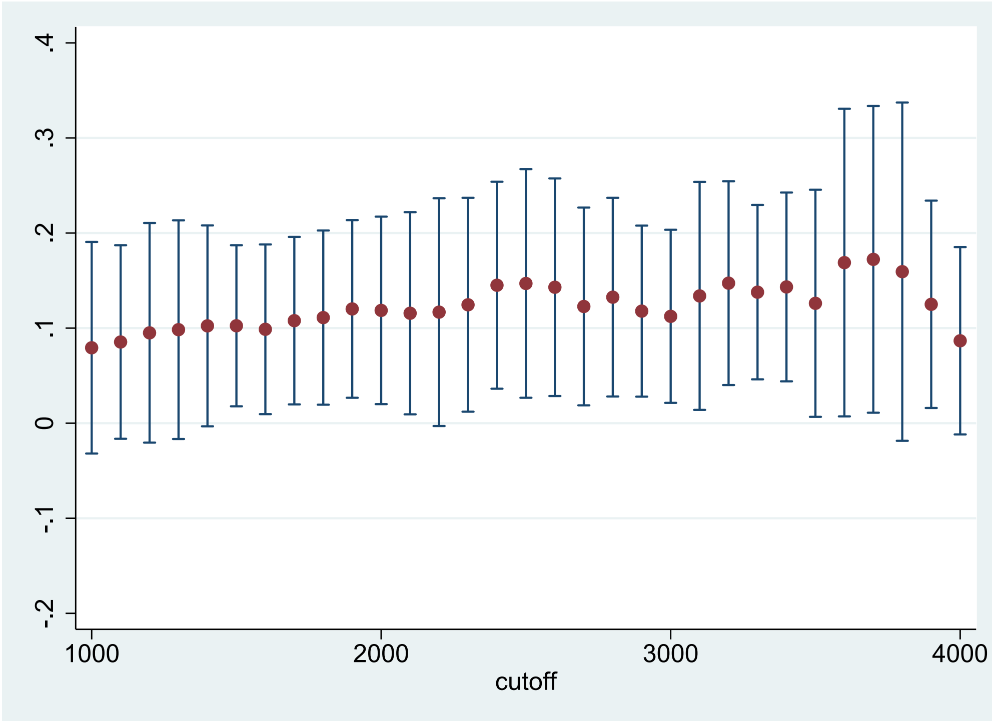
*Note*: Each cell reports the results from a different regression. The dependent variable in all regressions is $\Delta$ *Share of Votes* 2003-2007. *Large municipality* is a dummy variable that takes value one for municipalities with more than 5,000 inhabitants. In panels A and C the sample includes all *male holdouts*, according to the corresponding definition. In panels B and D the sample includes all *male holdouts* in large municipalities and a matched sample of *male holdouts* in small municipalities. In panels C and D, the set of controls includes information about the electoral results in 2003, the gender composition of lists in 2003, a quadratic polynomial of population and party dummies. In Panels A and C we report standard errors clustered by region (N=17) in parentheses, as in CS specification. In Panels B and D we report OLS standard errors in parentheses, as in CS specification. Bootstrapped p-values are cluster (by region)-robust bootstrap p-values for the estimator of *Large municipality* Significance levels: 1% ***, 5% ** and 10% *

**Table 3: Impact of Quotas: Discontinuity-in-differences**

| Male holdouts: | **CS definition** | **Strict definition** | **Fewer Female Candidates** | **Fewer Top 5 Female Candidates** |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | | *Panel A. Share of women in the list* | | |
| Large municipalities | 0.099*** | 0.093** | 0.120*** | 0.102*** |
| | (0.031) | (0.040) | (0.029) | (0.030) |
| BW Loc. Poly. (h) | 1741 | 1254 | 1291 | 1872 |
| Obs left of c | 132 | 57 | 126 | 170 |
| Obs right of c | 88 | 48 | 114 | 107 |
| | | *Panel B. Vote share* | | |
| Large municipality | 0.029 | -0.013 | 0.009 | 0.025 |
| | (0.026) | (0.039) | (0.025) | (0.026) |
| BW Loc. Poly. (h) | 1779 | 1543 | 1454 | 1668 |
| Obs left of c | 134 | 74 | 148 | 135 |
| Obs right of c | 88 | 59 | 123 | 101 |

*Note*: Each cell reports results from a bias-corrected robust discontinuity in differences estimation (see equation (4)), where the bandwidth was chosen according to the MSE-optimal bandwidth selector (see Calonico et al. (2016)). The polynomial in population is chosen to be linear and is allowed to have a different slope on the two sides of the 5,000 threshold. Observations are weighted by distance from the threshold, using a triangular Kernel. The dependent variable in panel A is $\Delta$ *Share of Women in the List* 2003-2007, while in panel B is $\Delta$ *Share of Votes* 2003-2007. Significance levels: 1% ***, 5% ** and 10% *

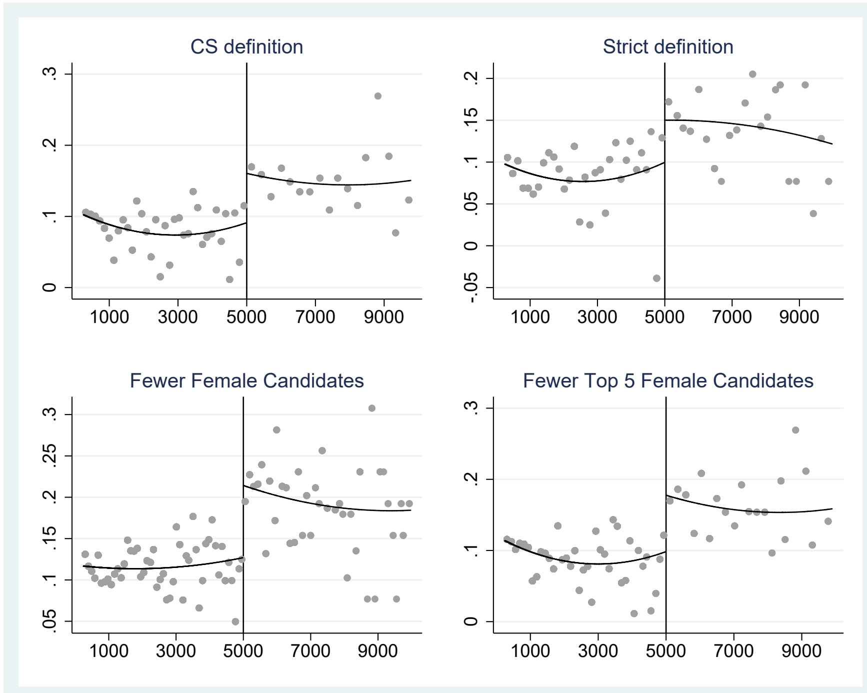
**Figure 1: Placebo regressions**



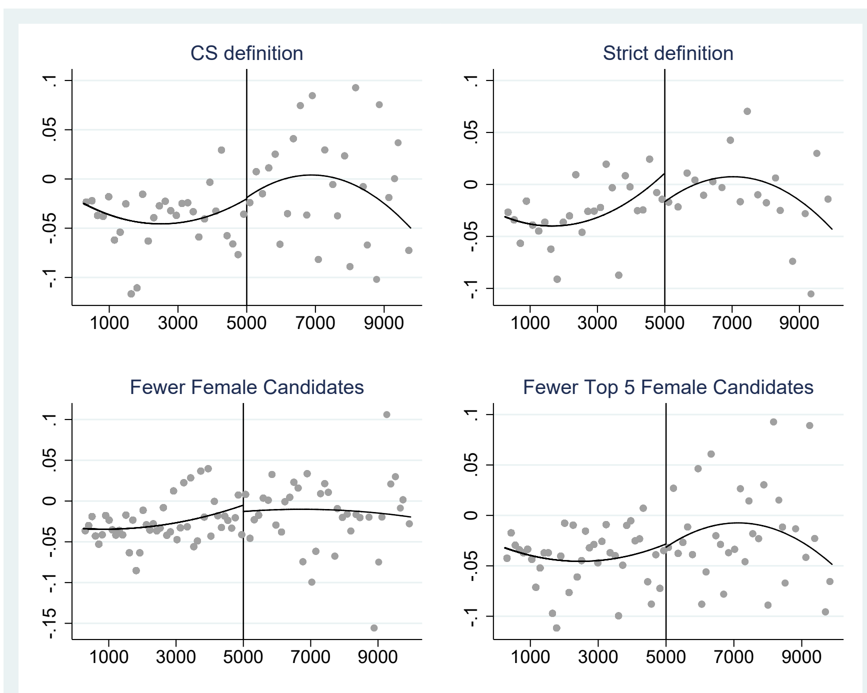

Notes: The placebo regressions estimate variations of equation (1), where the variable *large* takes value 1 for population above different thresholds, growing from 1000 to 4000 at increments of 100. The sample is made of municipalities with population below 5000.

## Figure 2: Discontinuity-in-differences Graphs

### (a) Share of Women in the List




### (b) Votes




Note: These graphs provide information on the the change in the share of female candidates and in votes for male holdout lists in bipartisan municipalities, by population of the municipality. Dots are means, lines are fitted values from second-order polynomial regressions. Gender quotas were implemented in the 2007 elections in municipalities which had more than 5,000 inhabitants in January 2006. the X-axis represents the population of the municipality on January 2006.

**Table A1: Impact of Quotas on the Share of Female Candidates: Bipartisan Male Holdouts**

| Male holdouts: | **CS definition** | **Strict definition** | **Fewer Female Candidates** | **Fewer Top 5 Female Candidates** |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | *Panel A. Share of women in the list; Unmatched sample; With controls.* | | | |
| Large municipality | 0.079*** | 0.063** | 0.082*** | 0.083*** |
| | (0.013) | (0.015) | (0.010) | (0.012) |
| | *Panel B. Share of women top 5 positions; Unmatched sample; With controls.* | | | |
| Large municipality | 0.083*** | 0.046** | 0.082*** | 0.094*** |
| | (0.016) | (0.022) | (0.014) | (0.016) |
| | *Panel C. Share of women in the list; Matched sample; With controls.* | | | |
| Large municipality | 0.085*** | 0.052*** | 0.087*** | 0.089*** |
| | (0.016) | (0.021) | (0.012) | (0.015) |
| | *Panel D. Share of women top 5 positions; Matched sample; With controls.* | | | |
| Large municipality | 0.096*** | 0.025 | 0.072*** | 0.113*** |
| | (0.021) | (0.028) | (0.017) | (0.019) |

*Note*: Each cell reports the results from a different regression. The dependent variable is $\Delta$ *Share of Female Candidates* 2003-2007 in Panels A and C, and $\Delta$ *Share of Top 5 Female Candidates* 2003-2007 in Panels B and D. *Large municipality* is a dummy variable that takes value one for municipalities with more than 5,000 inhabitants. In panels A and B the sample includes all *male holdouts*, according to the corresponding definition. In panels C and D the sample includes all *male holdouts* in municipalities and a matched sample of *male holdouts* in municipalities with less than 5,000 inhabitants. The set of controls includes information about the electoral results in 2003, the gender composition of lists in 2003, a quadratic polynomial of population and party dummies. Heteroskedasticity-robust standard errors in parenthesis. Significance levels: 1% ***, 5% ** and 10% *