Supplemental Materials

A Additional figures and tables

This Appendix provides some additional Figures that complement the analyses carried throughout the main text. In the first place, Figure A1 shows the distribution of the instrument for the ENPP (our main specification) and for the absolute number of parties (a robustness check whose results are included in Figure 5 and shown in Figure D1).

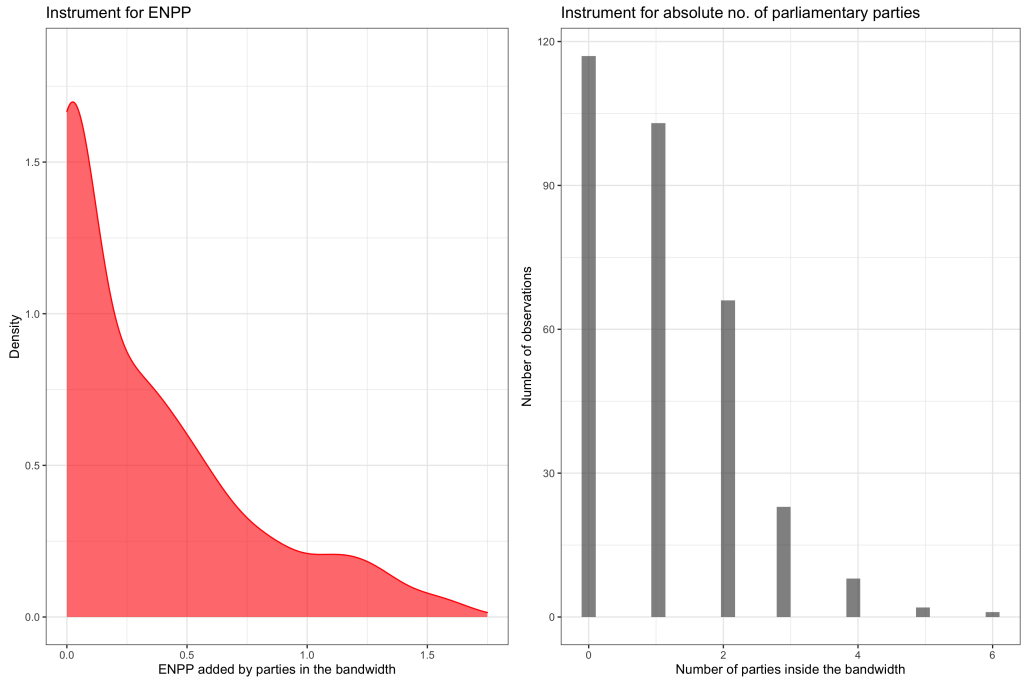


Figure A1: Distribution of the instrument.

We check for sorting around the threshold. If parties could sort into one side of the threshold, this could present a threat to the ignorability assumption. To check for this possibility, we run the McCrary test on the party-level dataset. The forcing variable is the distance between each party's vote share and the electoral threshold in its country. We measure this distance as a proportion of the threshold. For example, all parties with a vote share of 0 will have a value of -1 in this variable, meaning that they are one threshold away from making it to parliament. Parties with a vote share exactly equal to the threshold will have a value of 0. A party with a vote share that is twice as large as the electoral threshold will have a value of 1 in this variable. As the Figure shows, we

find no evidence of sorting. The p-value for the null of no-sorting is 0.95.

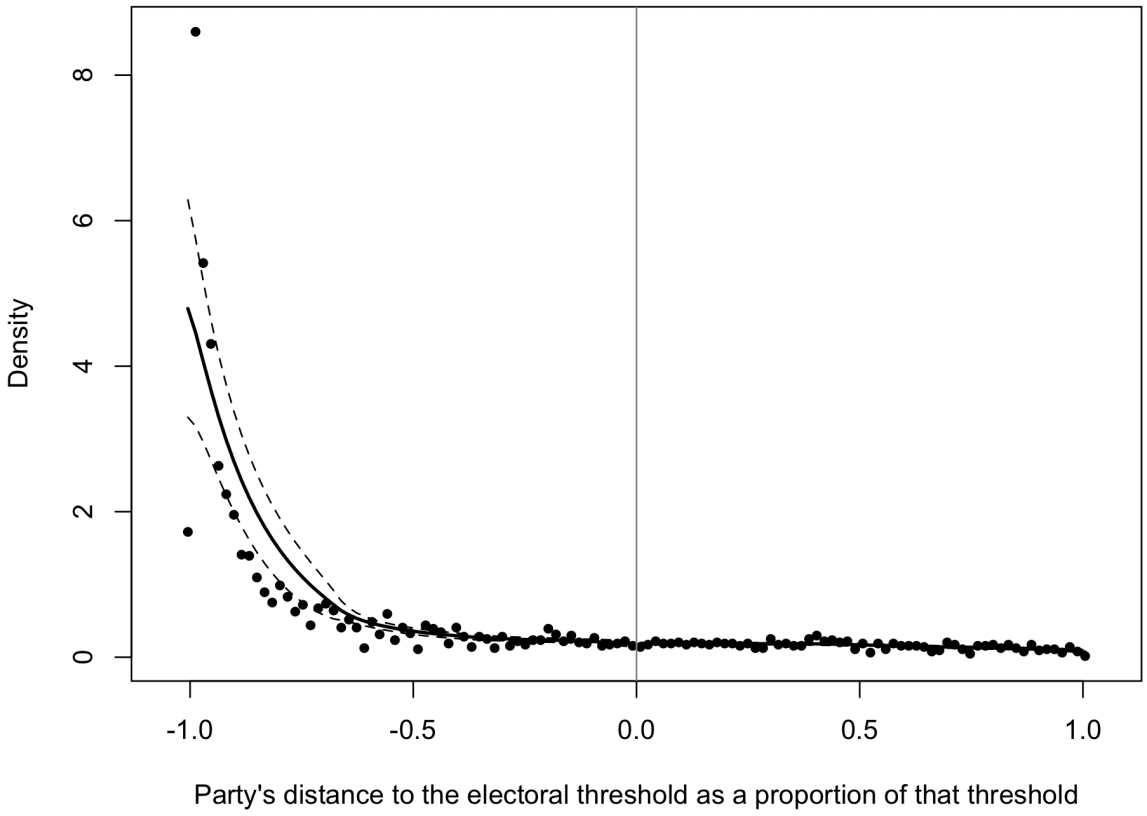


Figure A2: McCrary test for sorting around the threshold.

We then report the results of a number of regression discontinuities run on the party-level dataset, using placebo outcomes: year, a dummy for year above 1989, the left-right position of the party according to the Comparative Manifesto Project, and dummies for each country. Should we find significant effects on either of these variables, they could be taken as evidence of threats to the exclusion assumption. Figure A3 shows these results. All dependent variables have been standardized. Again, the forcing variable is the distance between the party's vote share and the electoral threshold in the country, measured as a proportion of the threshold. We replicate the analyses twice: in the left-hand panel, we use the bandwidth used in our main analyses: 50% around the threshold. In the right-hand panel, we estimate each regression discontinuity using its own optimal

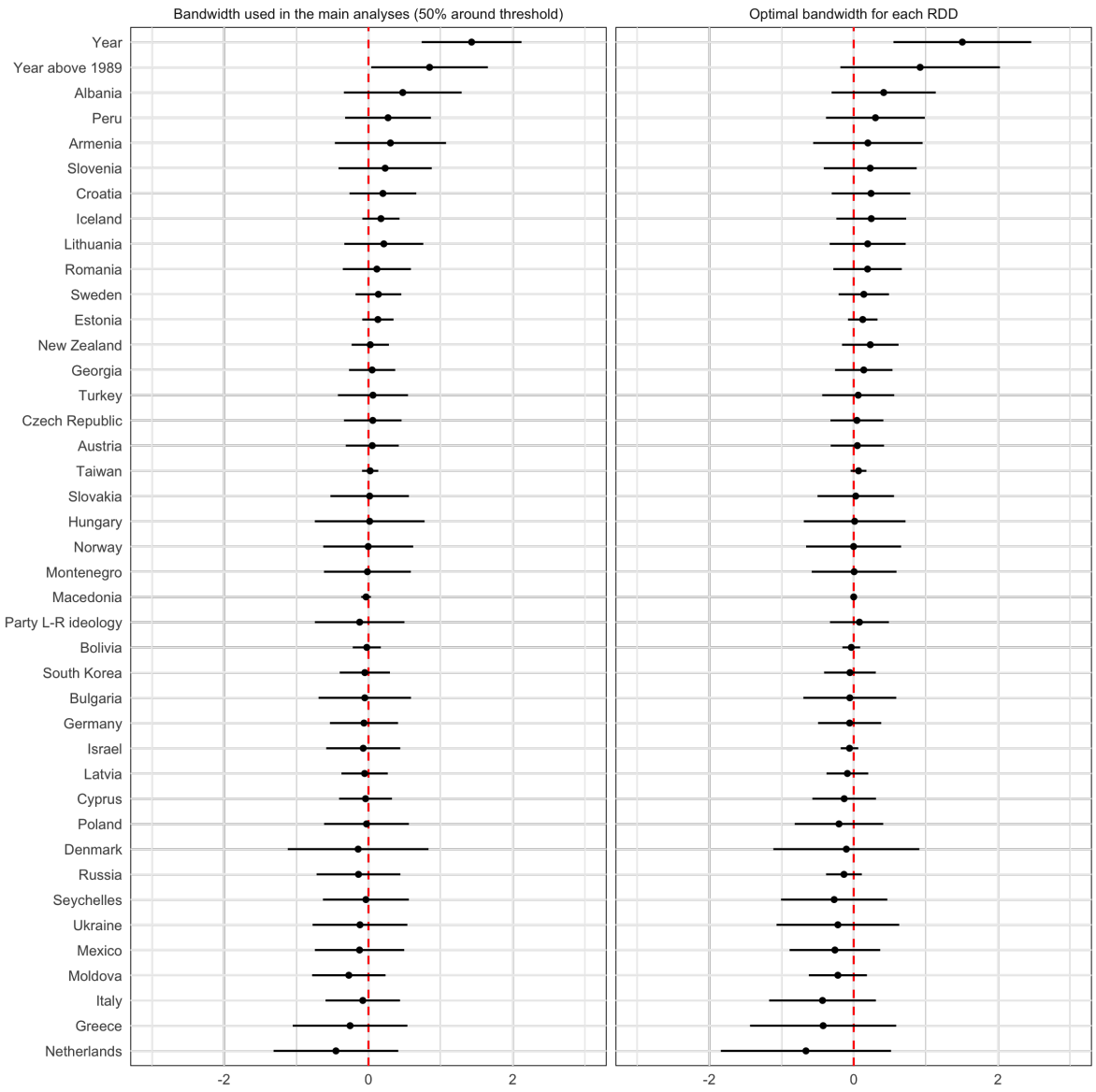
bandwidth, as calculated by the package rdrobust (Calonico et al. 2017). Out of the 41 placebos we analyse, only one (year) reaches the threshold of statistical significance at the 95% level in both models.

A possible concern with the main analyses shown in Figure 4 is that the effects on turnout may be driven by disillusion. When less parties narrowly make it to parliament, more voters are likely to be disappointed with the result of the election. This might affect their probability to turnout.

While it is unlikely that this mechanism may affect the other outcomes we draw upon, it is a likely alternative explanation for turnout. For this reason, we look carefully into it. If this is the case, then there should be more wasted votes when less parties narrowly make it to parliament (meaning, when our instrument has lower values). To check for this possibility, we correlate our instrument with the share of wasted votes—the sum of vote shares for parties that did not make it to parliament.

The results are shown in Figure A4. As the Figure shows, if anything, there is a positive correlation. Higher values of our instrument (meaning, more parties narrowly entering the parliament) is associated with a higher share of votes being wasted. It thus seems unlikely that this alternative explanation may be driving our results on turnout.

Figures A5 through A6 show the first stage when we use alternative specifications of the instrument.



Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized.

Figure A3: Regression discontinuities using placebo outcomes.

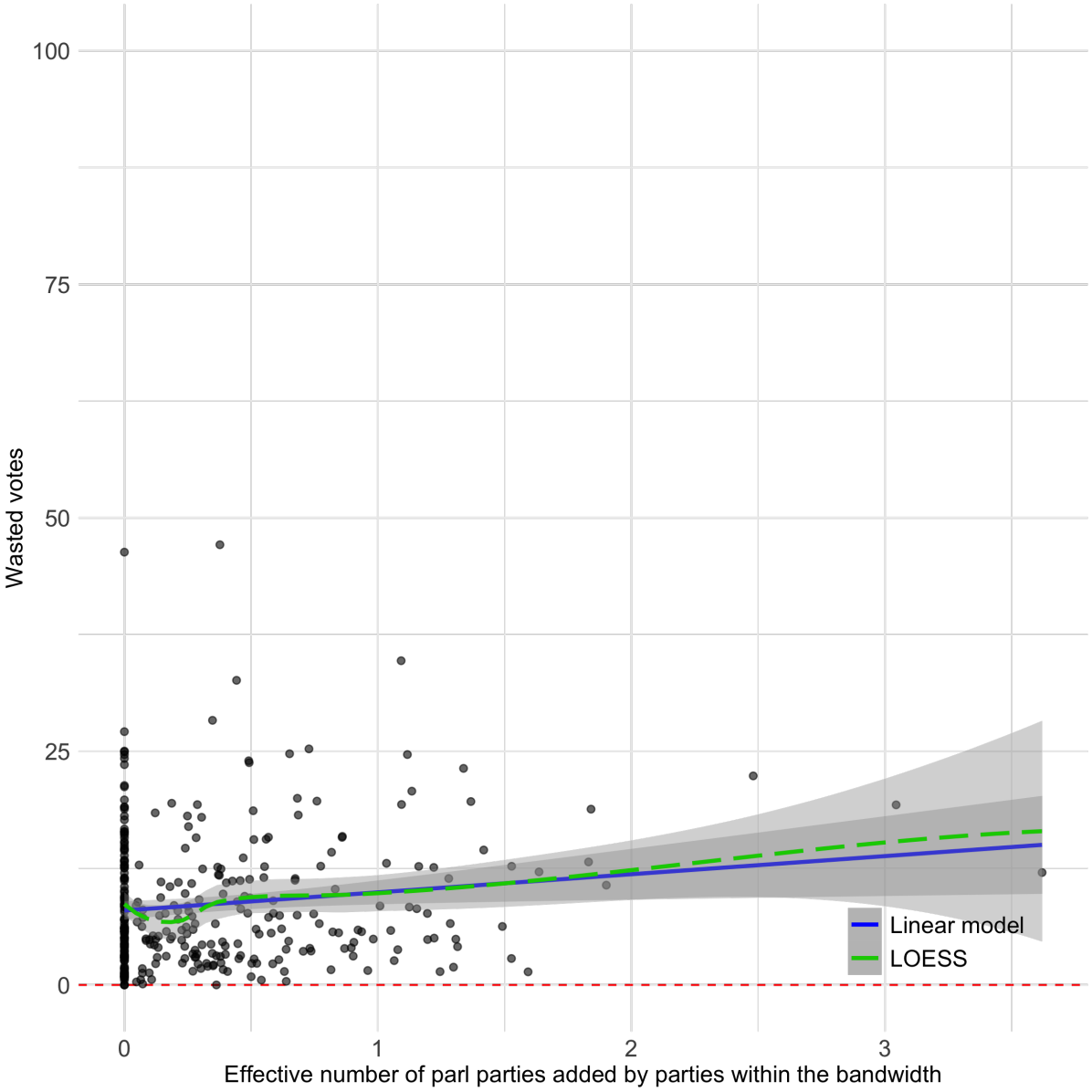
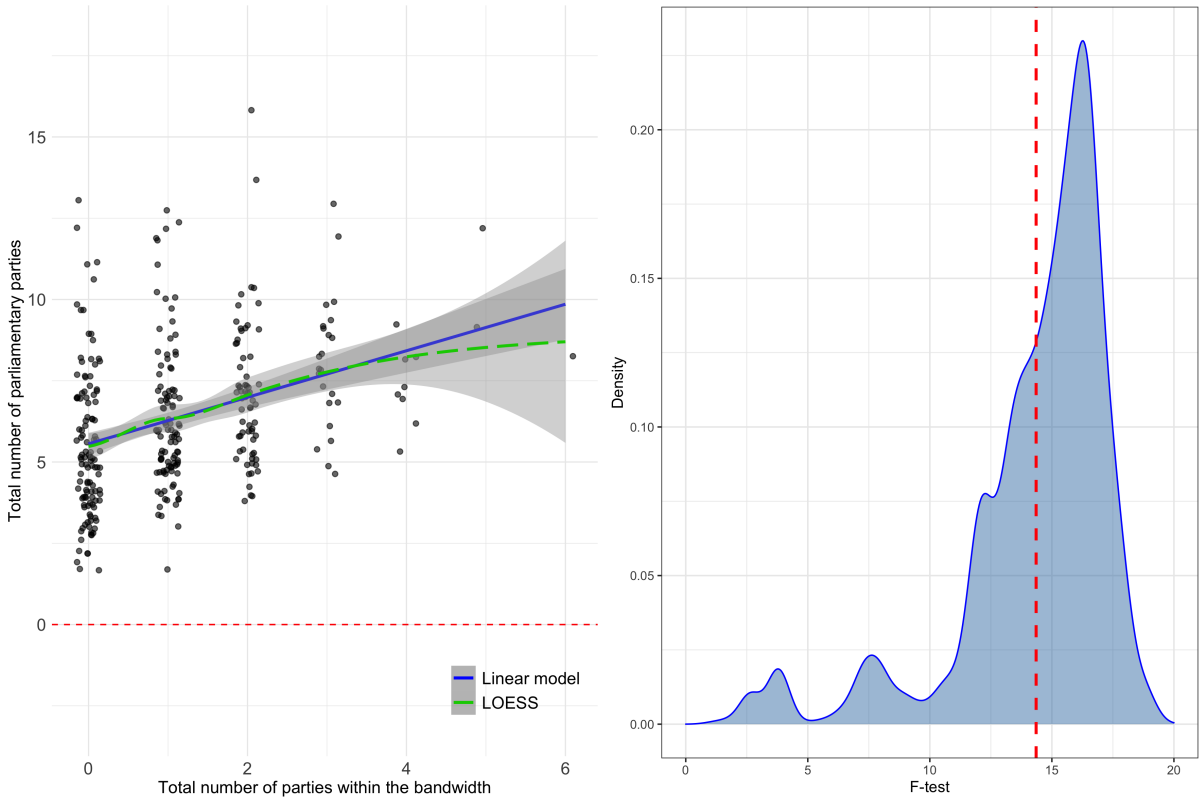
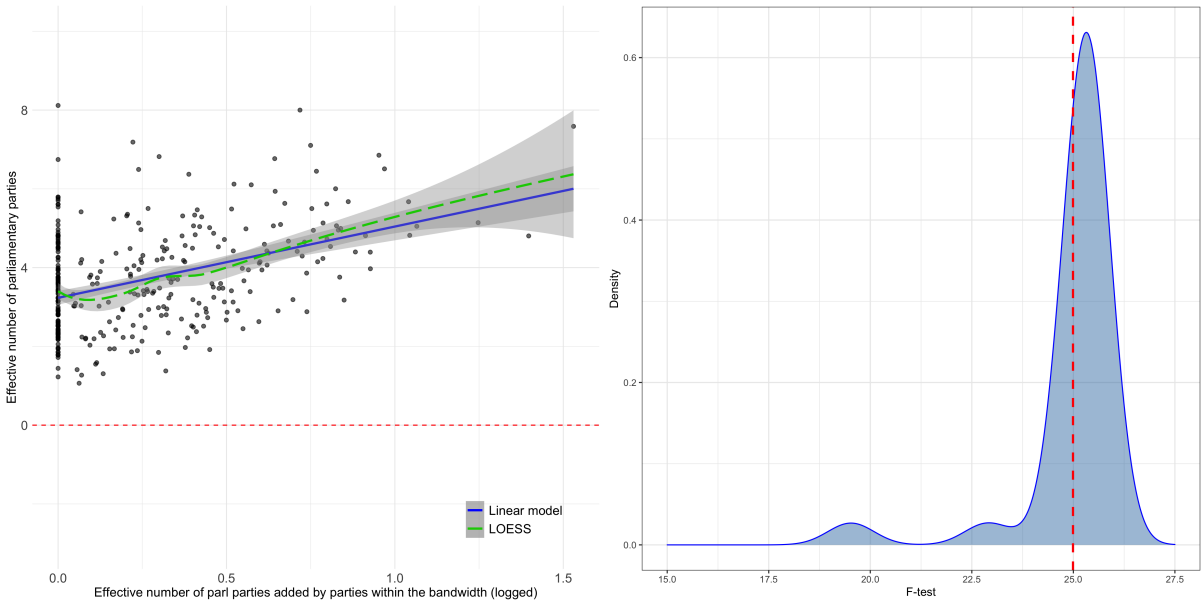


Figure A4: Correlation between the sum of wasted votes and our instrument.



Note: Vertical red line in the right-wing panel represents the average value of the F-test.

Figure A5: First stage for the alternative specification of the treatment variable (using the absolute, instead of effective, number of parties).



Note: Vertical red line in the right-wing panel represents the average value of the F-test.

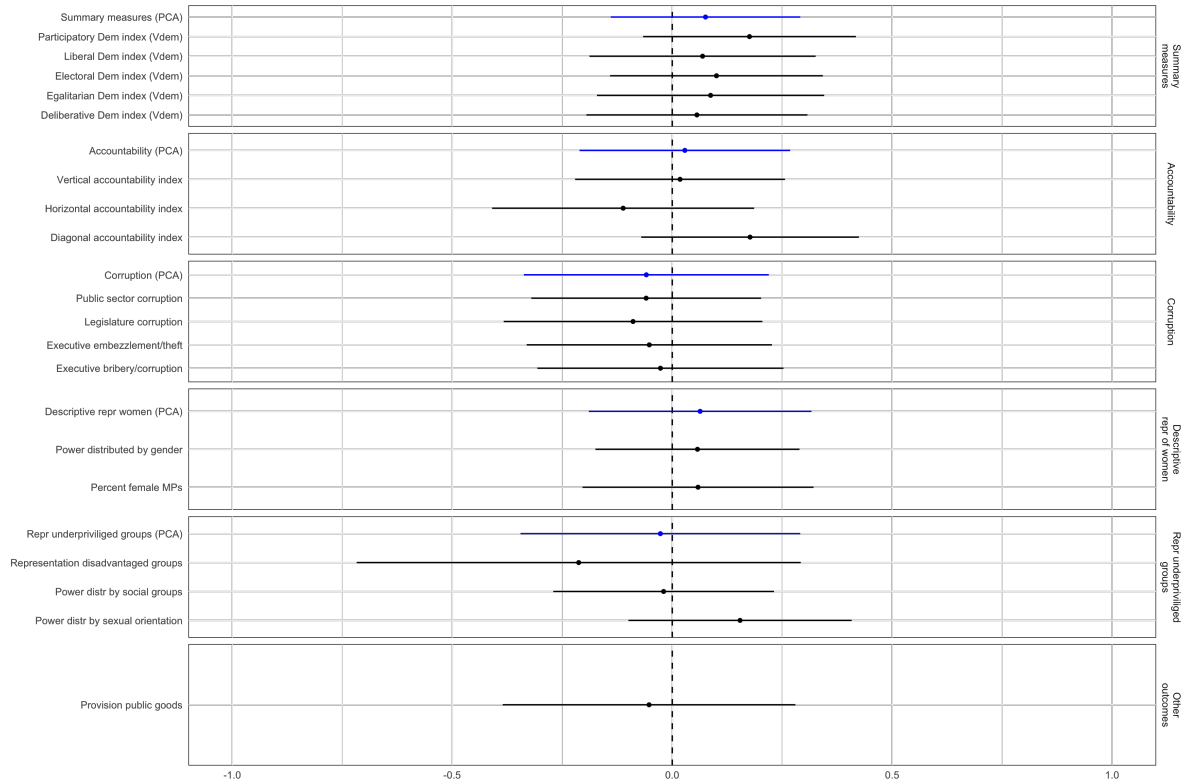
Figure A6: First stage using the logged insrument.

B Long-term effects?

Another potential objection to our analyses is that it may take some time for the effect of the number of parties to be felt. In our main set of analyses, we measure our outcomes at the end of the legislature at the beginning of which the instrument and treatment are measured. This may be too short a time span for party-system fragmentation to affect some outcomes. On the one hand, as discussed by Duverger (1954), electoral systems affect party system via a mechanical and a psychological effect. The former lasts for a single legislative period, while the latter has a more long-term effect. Psychological effects may be the main drivers of some outcomes we are interested in, such as the ones related to participation, which means that the effect on these outcomes may be felt mostly in the longer term. Moreover, some of the outcomes we are interested in—such as the ones concerning corruption—are, by nature, likely to remain relatively stable in the short-term. As such, it might be argued that it is unlikely that one can find an effect in the following legislative period. For these two reasons, it is important to assess whether the effects change in the longer term.

To be sure, we are not in the position to fully look into the long-term societal effects of party-system fragmentation. Still, in the following we try to provide suggestive evidence that increasing the time gap between treatment and outcomes does not change our results. Figures B1 and B2 replicates the 2SLS models shown in Figure 4, with the difference that the outcome variables are measured later: five years after the treatment (Figure B1) and ten years after the treatment (Figure B2).

Again, we find no consistent evidence of an effect of party-system fragmentation on the quality of democracy. We do find a significant effect on the index of diagonal accountability five years after treatment when we use our first estimation strategy, but not ten years after the treatment. In turn, the other accountability outcomes yield much smaller coefficients, which are also far from statistical significance, and whose direction is contradictory with one another. Moreover, given the high number of outcomes that we use, one should be cautious to draw conclusions from a single coefficient being statistically

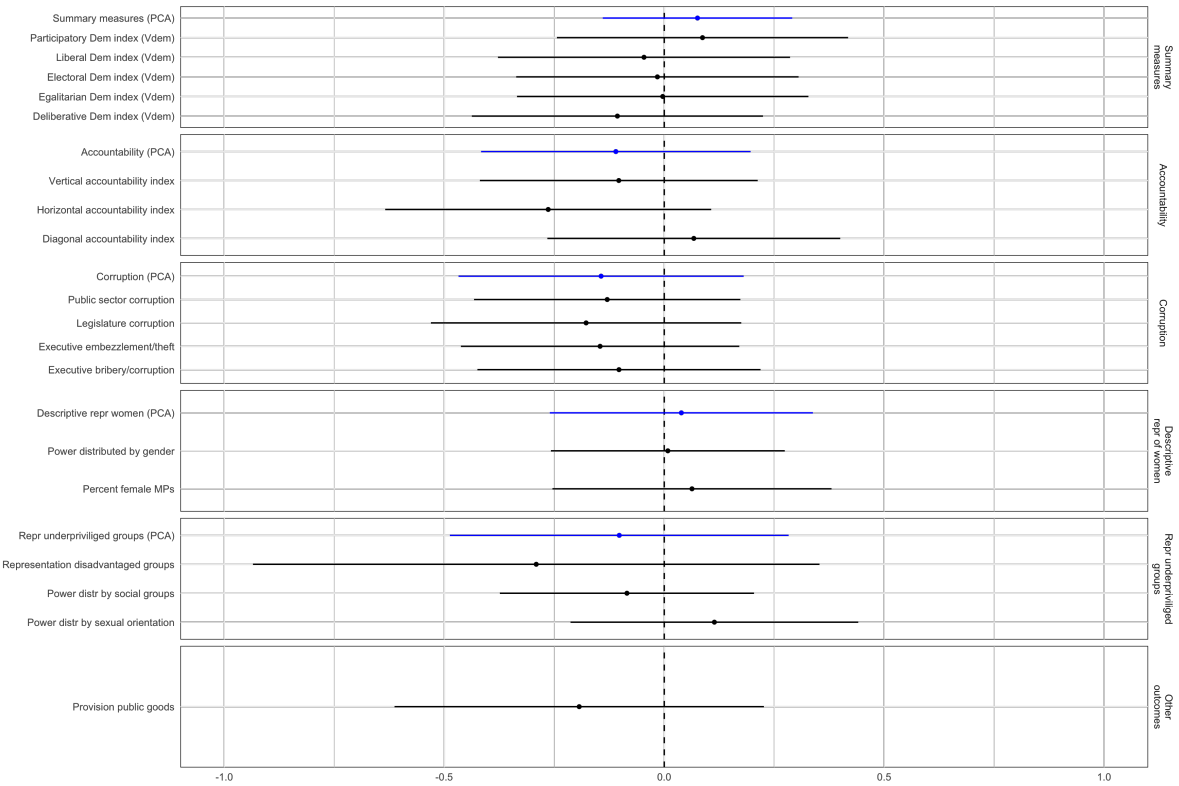


Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized.

Figure B1: Effect of party-system fragmentation on the quality of democracy, measured five years after the treatment.

significant—especially when not robust across different models—since it can be just a product of multiple hypothesis testing (Dudoit, Shaffer and Boldrick 2003).

We also provide a robustness check to the analyses of Figure 7, where we look at whether the effect is conditional on the level of polarization. The difference is that we now focus not on the overall level of polarization, but instead of the level of polarization that is brought about by parties narrowly entering parliament. To this end, we start with calculating the level of polarization of the system, using the index proposed by Dalton (2008). Then, we subtract to the overall level of polarization the level of polarization that the system would have if the parties in the bandwidth had not made it to parliament. This gives us a measure of how much the parties in the bandwidth just above the threshold add to the overall level of polarization of the party system. We thus emulate the logic of our instrument and apply it to polarization. We then replicate the analyses on two

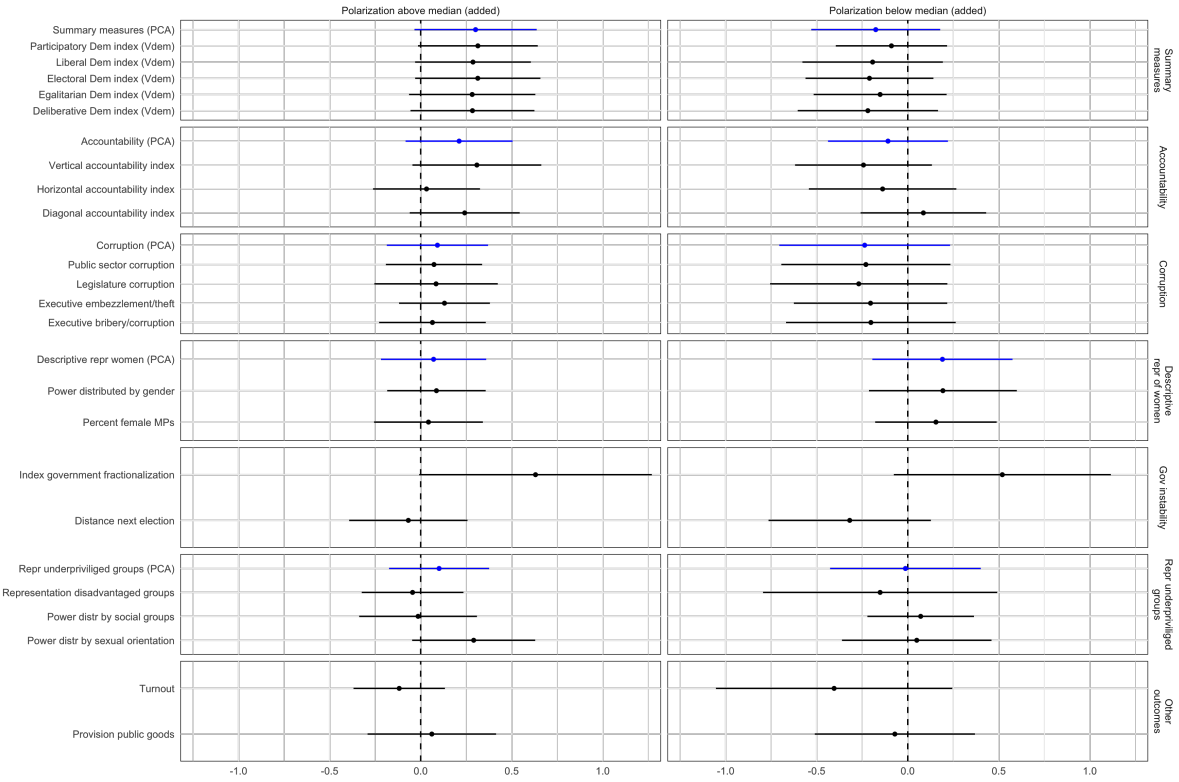


Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized.

Figure B2: Effect of party-system fragmentation on the quality of democracy, measured ten years after the treatment.

subsamples: the subsample where the level of polarization added by parties just above the threshold is below median (in the election where we measure the treatment variable); and the subsample where this value is above median.

The results of this exercise are shown in Figures B3. In general, the results are supportive of what we find in Figure 7. They suggest that, in contexts of high polarization, fragmentation has a more noticeable effect on the quality of democracy.



Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured.

Figure B3: Effect of party-system fragmentation on the quality of democracy, conditional on the level of polarization added by parties just above the threshold.

C Dealing with local effects

A potential concern with our analyses is that our results can only speak to changes in the ENPP that are brought about by variation in the treatment status of parties with vote shares close to the threshold, which are relatively small. This source of variation could render our analysis problematic if either of the following factors were into play. First, it

could be that the theory underlying party fragmentation effects is based on changes in number of parties other than those stemming from the inclusion of small parties in the legislature. Second, it could be that this source of variation alters the ENPP in a very specific fashion, namely increasing it only if its baseline value is below or above a given level. This could be the case if, for example, it is more likely to find parties near the threshold in multi-party systems. We try to address these concerns in turn.

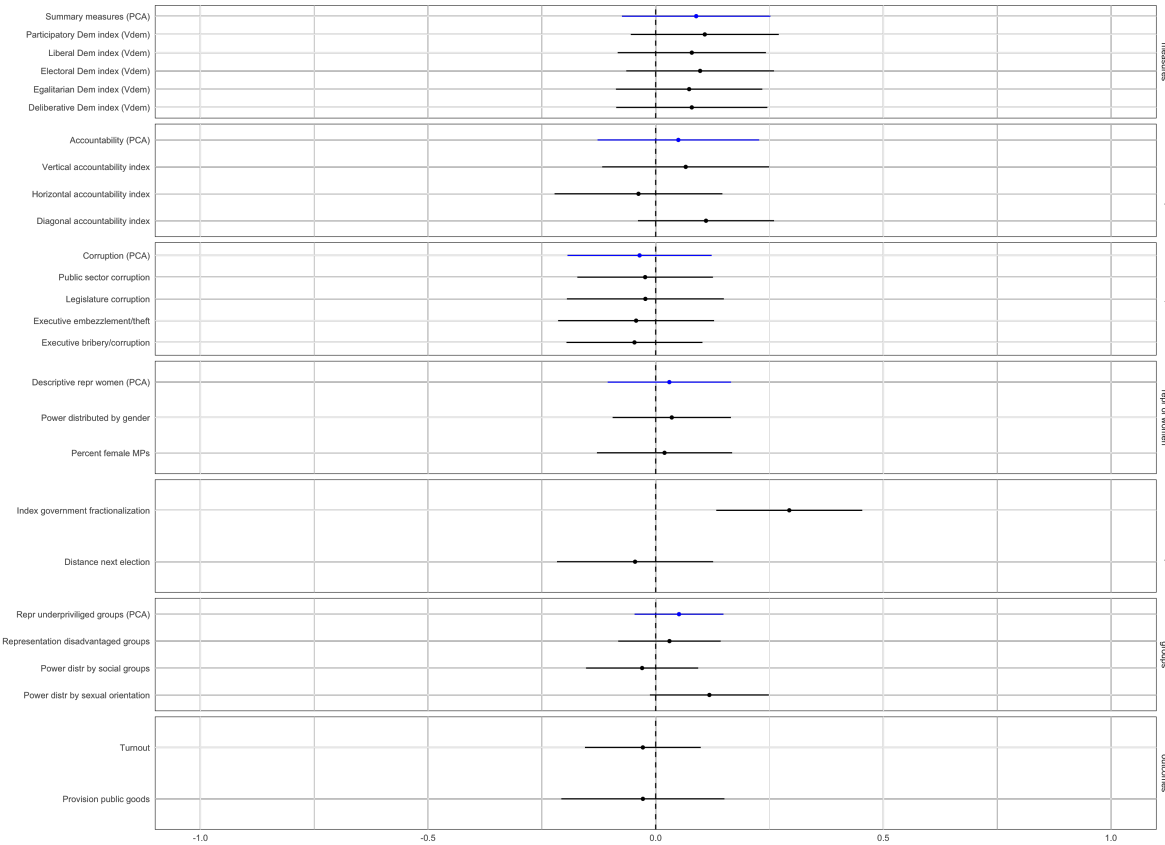
First, the literature does not seem to suggest differential affects of ENPP according to the type of parties in the parliament. If anything, for most of these arguments the type of party entry captured through our identification strategy would qualify as relevant. Take the example of women or minority representation, both based on the idea that more parties imply more points of access for these groups (Norris and Inglehart 2001). Clearly, parties crossing the threshold classify as providers of such points of access. The same logic should hold for the effect of fragmentation on turnout, either in one direction or the other.[18] By the same token, insofar as the higher number of parties translates into more fragmented governments, as we find, the argument about it blurring accountability and increasing corruption should also hold.

Empirically, we try to deal with this concern by repeating our analyses this time using only cases with higher thresholds, which would imply that parties entering the parliament have more seats in parliament. In particular, we look at elections with thresholds above the median, which means that parties within the bandwidth now have vote shares between 4% and 15%. If more dramatic parliamentary entries—by parties that manage to have a large share of the vote—have a more visible effect, we should be more likely to find it when we draw upon this subsample.[19] The results are shown in Figure C1. They are

18. The argument for a negative effect holds that higher fragmentation increases costs of access to information about the (now higher) number of parties in a system; is more likely to put voters under cross-pressure; and enhances the chances of a coalition government—thus reducing the benefits of voting. The arguments for a positive effect are mostly concerned with the higher probability of any given voter finding at least one party in the system that they identify with. Parties with vote share close to the threshold of representation can in principle affect all above-mentioned considerations. Again, it should be noted that we do find that a higher number of parties leads to a higher number of parties in government, a crucial assumption of the arguments about fragmentation reducing the benefits of voting.

19. That said, parties marginally clearing the electoral threshold are not the smallest parties found in parliaments. Elections with only regional thresholds or with exceptions to the national ones incorporate

very similar to the ones shown in Figure 4. All coefficients remain small and far from statistical significance.

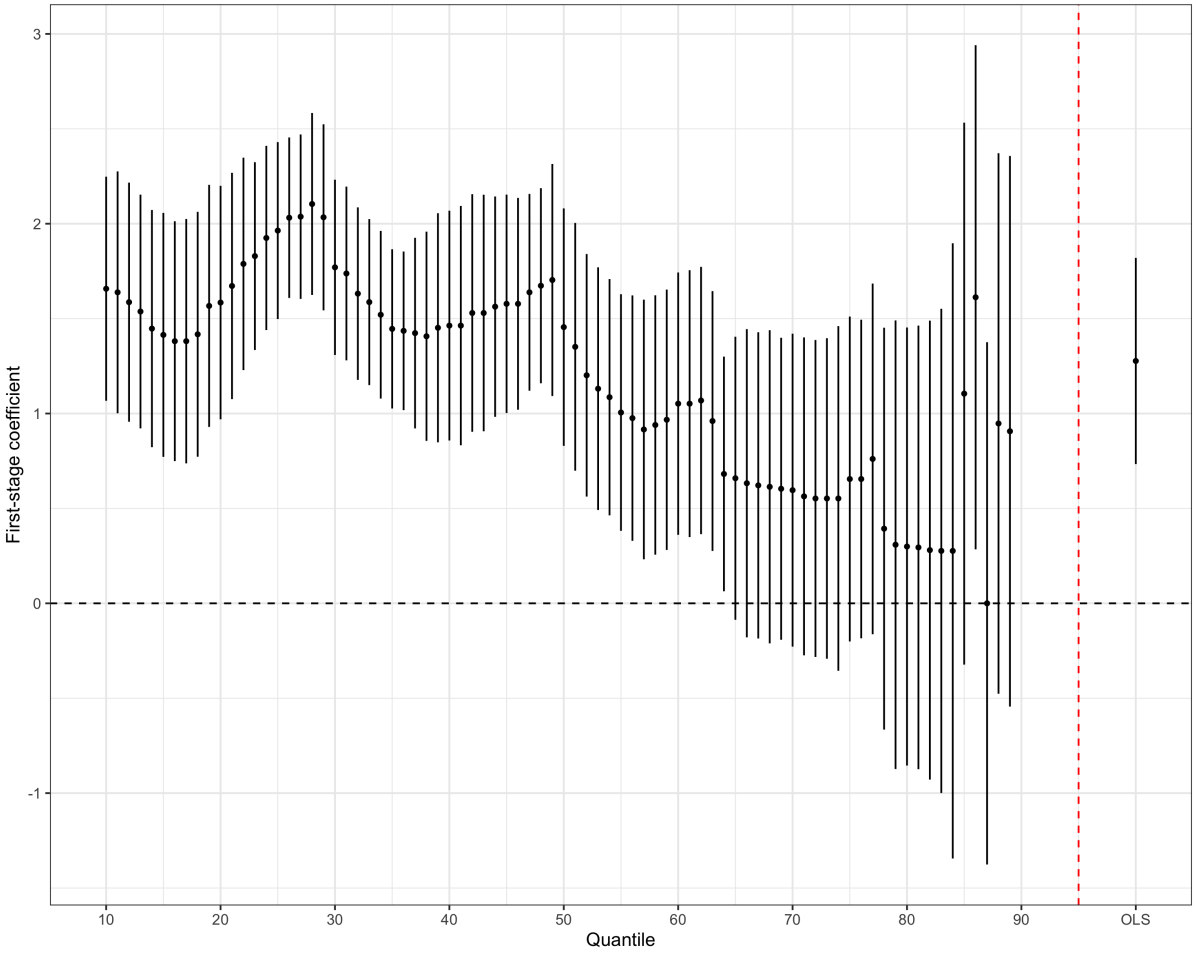


Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured.

Figure C1: Replication of the main analyses in Figure 4 using only elections with thresholds above the median.

Second, to address the possibility that our design increases the ENPP differentially according to its baseline value we report the first stage as a quantile regression. This test allows us to check whether our instrument is equally capable of predicting an increase of ENPP regardless of its baseline value. The results are shown in Figure C2. We reports the OLS regression that uses the full sample, along with quantile regressions for the 0.1 to 0.9 quantiles, in steps of 0.01. While it is true that in some regions our instrument is worse able to predict the level of ENPP, overall we find no clear pattern as we move to higher quantiles.

parties with national vote shares below the parliamentary cutoff.



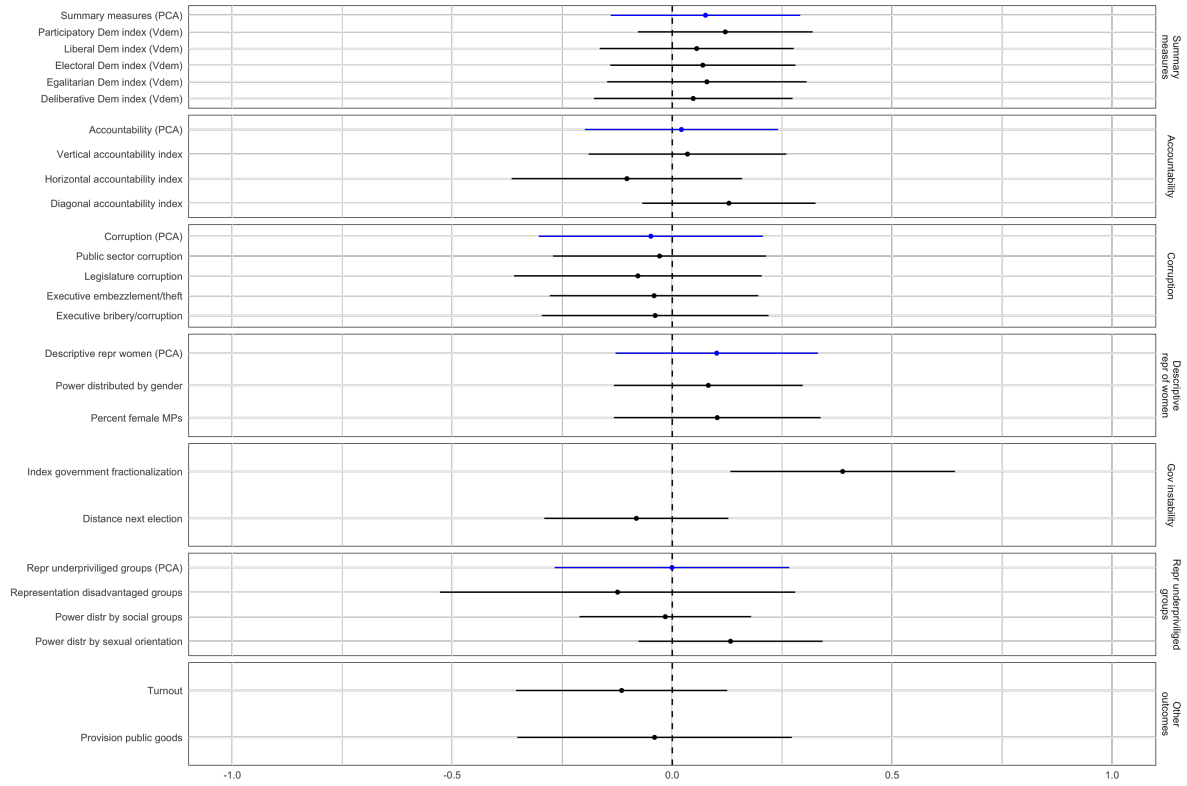
Notes: Vertical lines represent 95% confidence intervals.

Figure C2: First-stage estimates using quantile regression.

D Changing model specifications

As a robustness check, we replicate the main analyses as shown in Figure 4 using the absolute—instead of effective—number of parties as the treatment variable. It should be noted that, in these analyses, our instrument is simply the number of parties within the bandwidth. That is because the number of parties within that bandwidth represents the number of absolute parties that the parties in the bandwidth add to the overall level of fragmentation (as measured by the absolute number of parliamentary parties) in a given election. The results, shown in Figure D1, remain very similar to those of Figure 4.

A possible objection to the main analyses shown in Figure 4 is that, in the bandwidth we use which determines which parties are used to calculate the instrument is

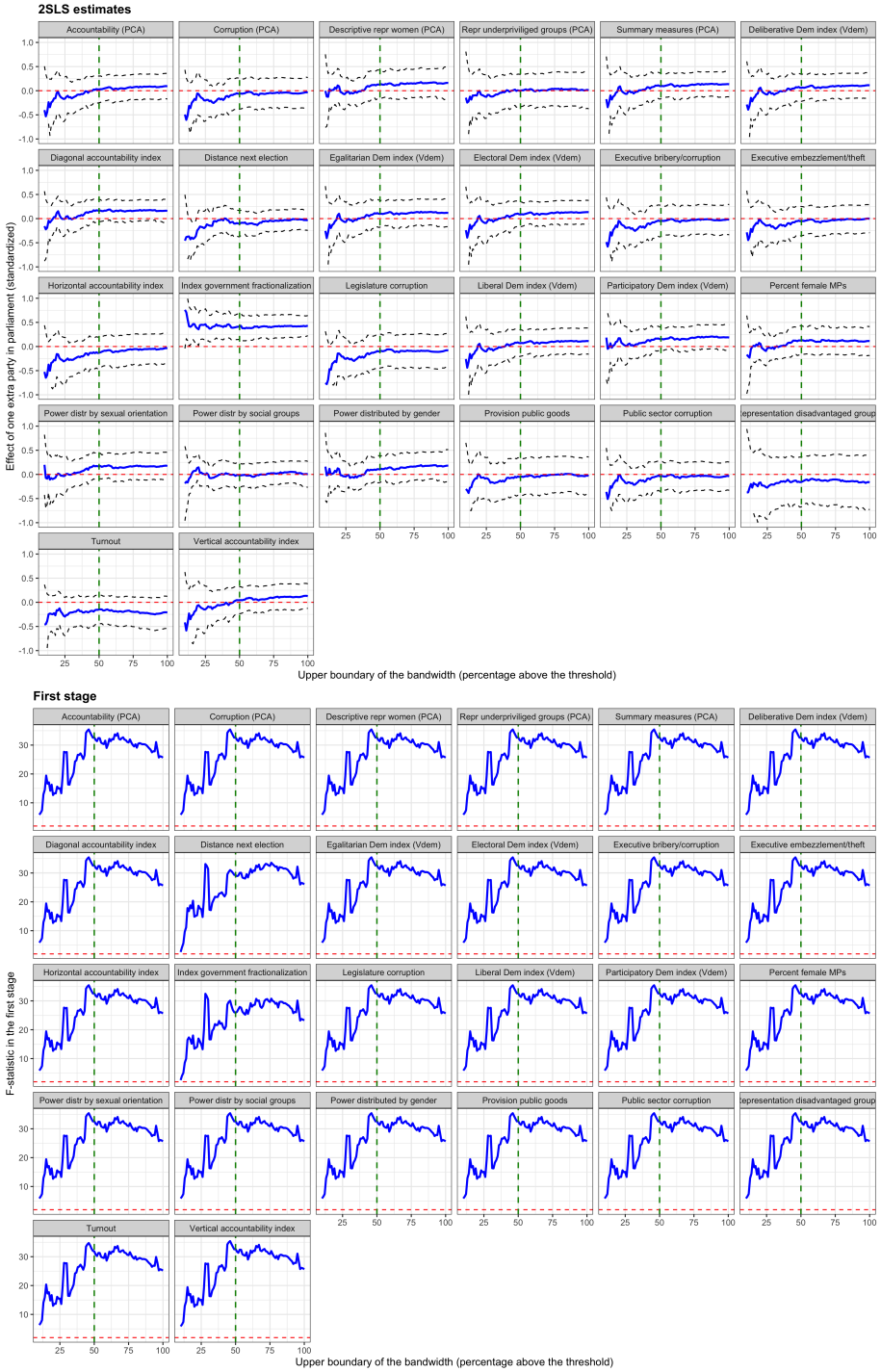


Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured.

Figure D1: Replication of the main analyses in Figure 4 using the absolute number of parties as the treatment variable.

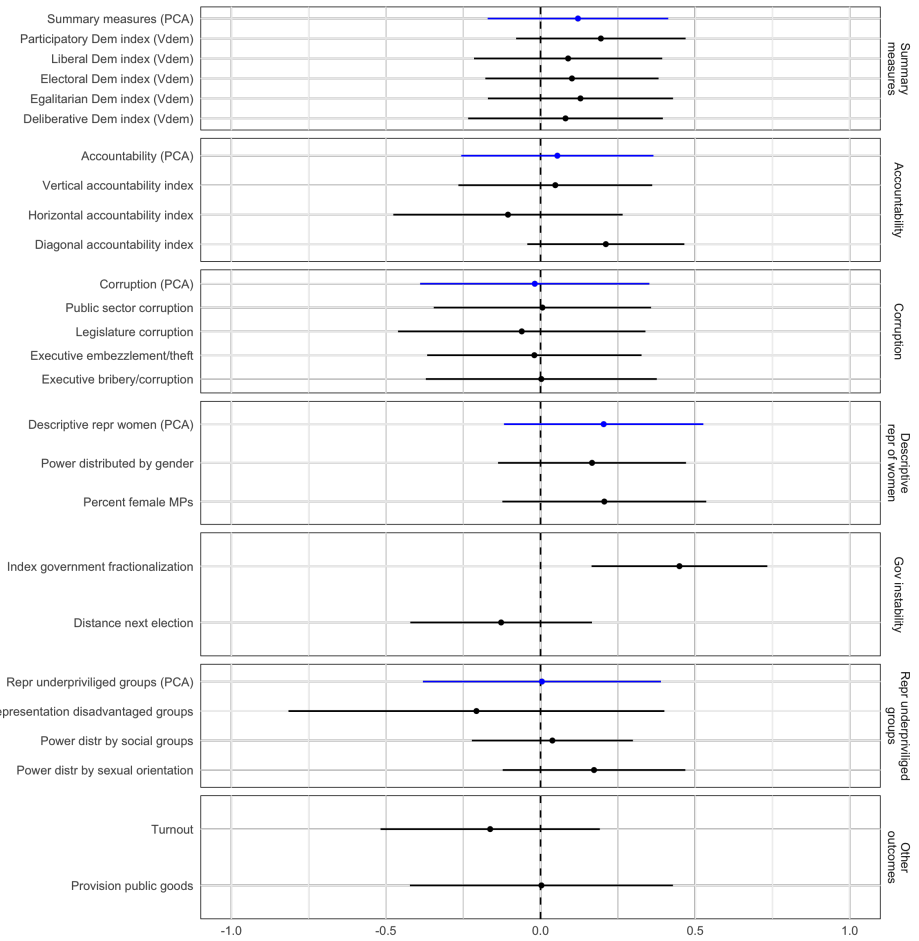
rather arbitrary. To tackle this problem, Figure D2 replicates these analyses employing different bandwidths. We use all bandwidths with an upper bound between 10% above the threshold and 100% above the threshold—in which case the upper bound is twice the electoral threshold. The lower bound of all these bandwidths is always the electoral threshold. The results remain substantively unchanged as the bandwidth is moved. The Figure also shows the first stage for all these analyses. It shows very clearly how we have a strong first stage regardless of the bandwidth we use. Moreover, it shows that our main specification (a bandwidth bordered by the electoral threshold on one end and the value 50% above that threshold on the other end) provides us with one of the strongest first stages.

One thing to note is that our instrument has a rather skewed distribution. For this



Notes: Dashed lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured. Vertical green lines represent the upper bound 50% above the threshold, which we use in the main specification. Red horizontal line in the lower panel represents the critical value of 1.96.

Figure D2: 2SLS estimates and first-stage with different bandwidths.



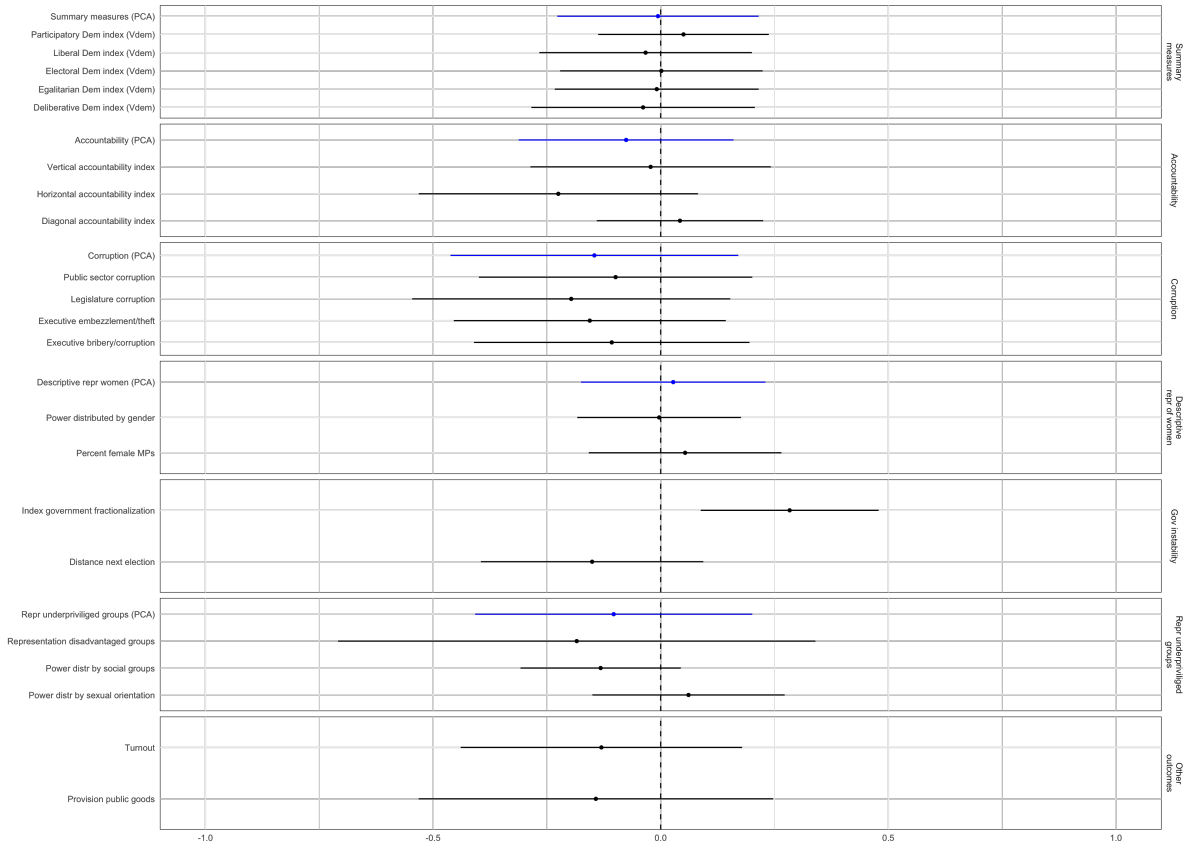
Notes: Dashed lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured.

Figure D3: 2SLS estimates, using the log-transformation of the instrument.

reason, we replicate the analyses shown in Figure 4 using the logged transformation of the instrument. As shown in Figure D3, the results remain very similar.

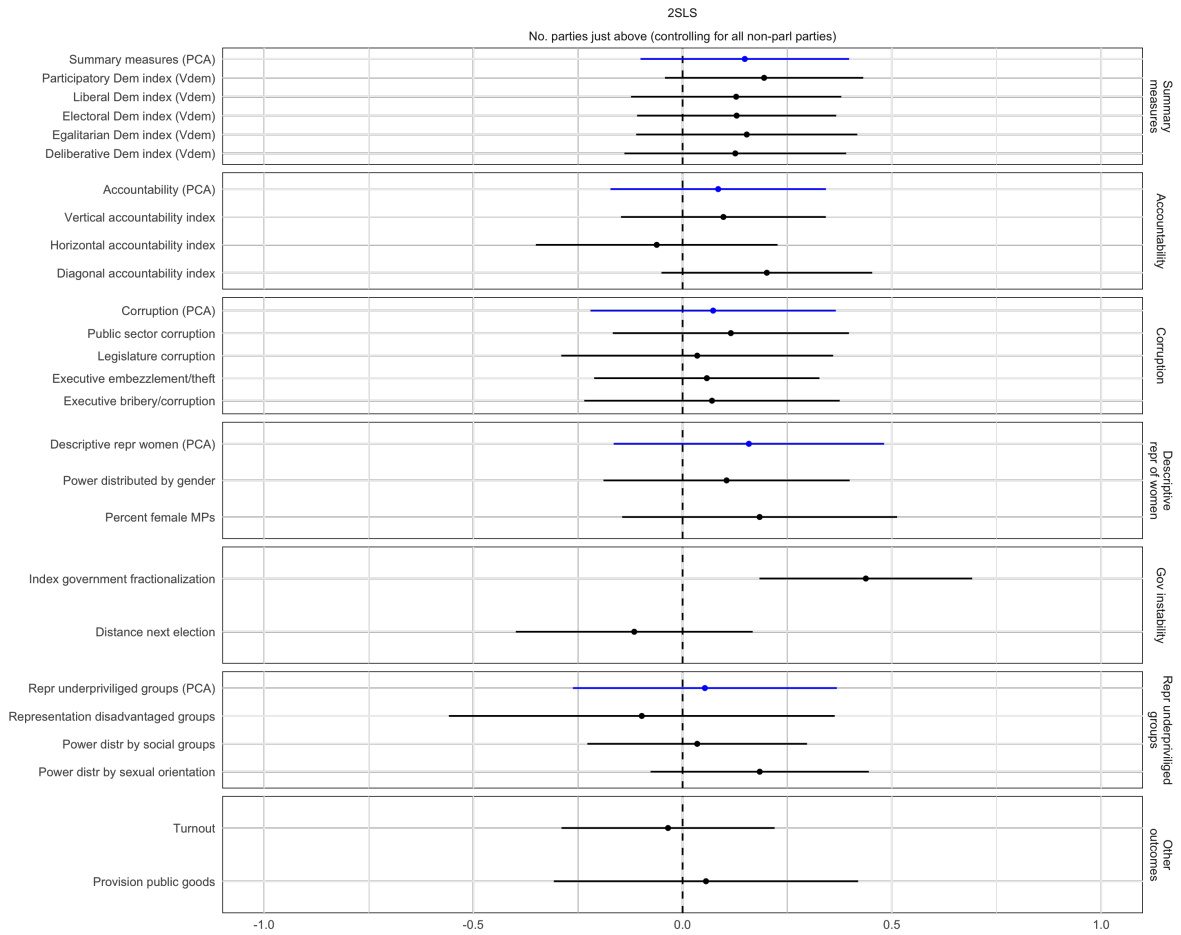
One possible concern with the analyses shown in Figure 4, which draw upon all elections that had an electoral threshold in place, is that they include countries from many regions and from some countries that are not fully democratic. We address this concern by replicating the analyses restricting the sample to European countries—-the region from which we have more observations. As shown in Figure D4, results remain very similar when we draw upon this more homogeneous sample.

We then report another alternative specification. In the main analyses, we control for



Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured.

Figure D4: Replication of the main analyses on European countries only.

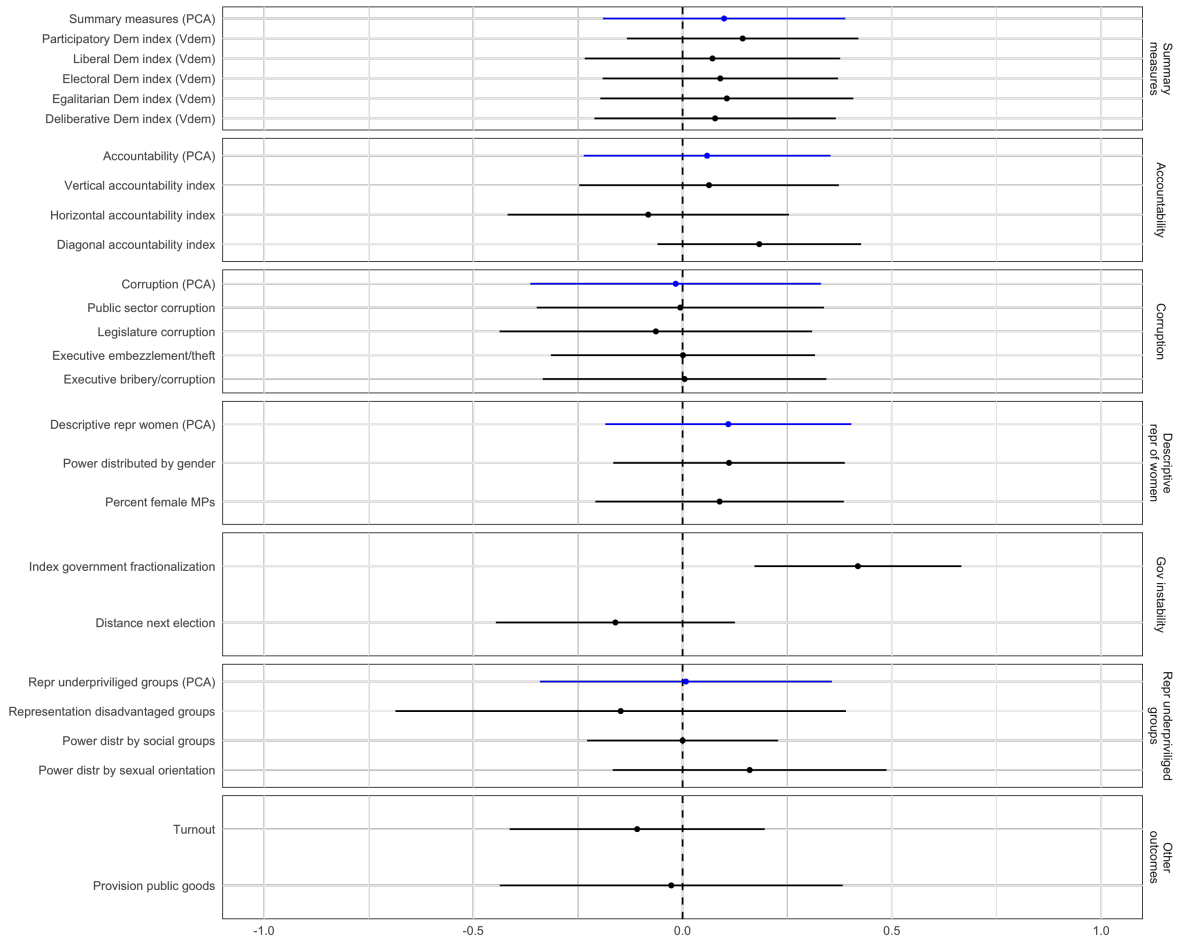


Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured.

Figure D5: Replication of the main analyses controlling for the number of non-parliamentary parties, regardless of how far they were from the threshold.

the number of parties that narrowly entered parliament, by including a control variable that captures the number of parties within the same bandwidth we use to calculate our instrument, but *below* the threshold. Here, we replace that control with a control for the number of parties that did not enter parliament, regardless of how far they were from the threshold. To be sure, this is a measure that is more prone to measurement error, because it is harder to be sure that one has data on all parties that ran for a given election as those parties become smaller. Still, we think this is an important robustness check.

The results are shown in Figure D5. As the Figure shows, the results with this alternative specification are very similar to the main ones.



Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized. Dependent variables are measured at the end of the legislature at the beginning of which the treatment is measured.

Figure D6: Replication of the main analyses controlling for the number of non-parliamentary parties, regardless of how far they were from the threshold.

Finally, we report a robustness check where we draw exclusively from the Dinas, Riera and Roussias (2015) dataset to calculate the ENPP. As mentioned in the main text, for our main analyses we use data on parliamentary parties as retrieved from CMP, because doing so alllows us to collect data on their ideological position. However, as also noted in the text, this dataset occasionally does not include data on some parties, which might lead us to underestimate the ENPP. To make sure this choice is not driving the results, we replicate the right-hand-side facet of Figure 4 using only the Dinas, Riera and Roussias (2015). The results, shown in Figure D6, remain very similar to the main ones.

Different First-Stage Estimation: Continuity in Potential Outcomes

Our strategy to estimate the first stage builds on an intuitive understanding of the RD. However, it draws on a conceptualization of the RD as a local randomized experiment (Cattaneo, Frandsen and Titiunik 2015), which implies as-good-as-random assignment of treatment status in the neighborhood of the threshold. Standard RD can work under a weaker assumption at the cutoff, namely the continuity of potential outcomes while crossing the cutoff point. Put simply, instead of requiring balance at average levels of unobservables below and above the threshold, all we need to assume is that that the only change that takes place right at the cutoff is due to treatment. With finite samples, this assumption typically comes with a cost of modelling assumptions when it comes to estimation (Branson and Mealli 2018), as it requires a linear or polynomial function to extrapolate point estimates at the cutoff point.

As a way to approximate the "continuity" framework, we provide an alternative form of estimating the first stage, where we predict the number of parliamentary parties as a result of whether the parties closest to the threshold, on either side of the cutoff, cross it or not. We start by centering parties' vote share around the cutoff. We then find the four parties closest to the threshold and for each we use its centered vote share as the forcing variable together with our instrument—a dummy that switches on if this party is at or above the threshold. To flexibly account for the distance between each party's vote share and the threshold, we include up to the fourth polynomial of the distance between each party and the cutoff.

Formally, for k parties, where $k = \{1, \ldots, 4\}$ and m polynomials, where $m = \{1, \ldots, 4\}$, the first-stage is given by the following equation:

$$\text{NP}_j = \alpha + \beta_k \text{Above}_{kj} + \sum_{k=1}^{4} \sum_{m=1}^{4} \gamma_{mk} (\text{Vote}_{kj} - c_j)^m + u_j, \tag{1}$$

where NP represents the number of parties in parliament (be it measured as the ENPP, as in our main specification, be it measured as the absolute number of parties, as in the

alternative specification shown in Figure D1); k represents the Kth closest party to the threshold in election j; “Above" is a dummy coded one if the vote share for party k in election j was larger than the electoral threshold in that election; c_j denotes the cutoff (the electoral threshold) in place in election j; “Vote" represents the vote share for party k in election j; m represents the mth-order polynomial of the vote share of party k centered around the threshold c_j—which we include to make sure we capture non-linearities in the relation between centered vote share of each party and the probability of treatment; and u denotes the error term.

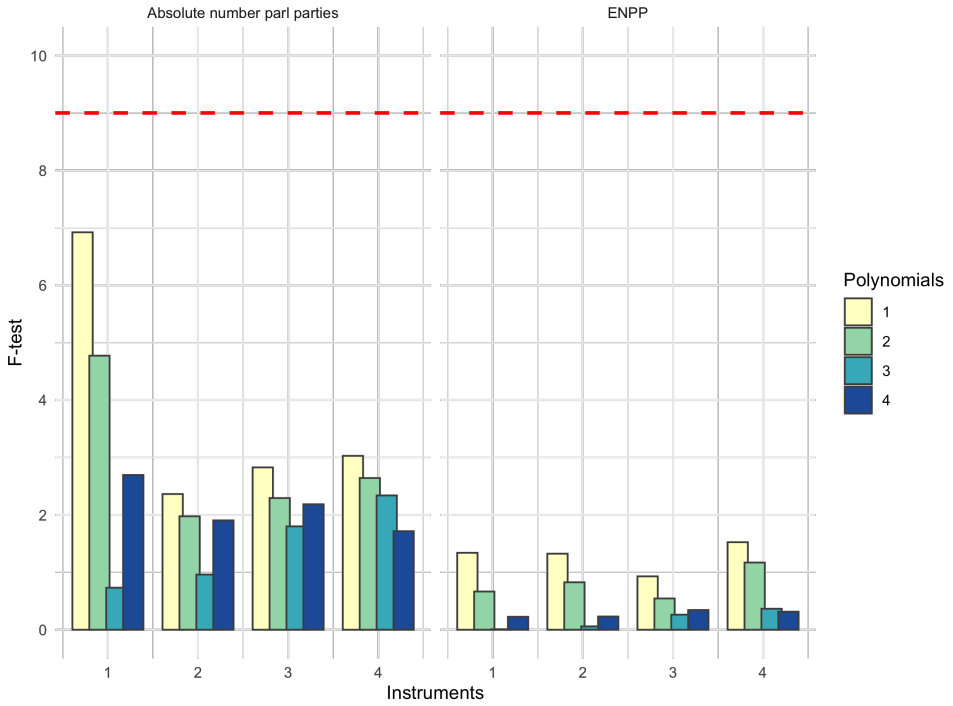
In turn, the second stage is given by the following equation:

$$Y_j = \lambda + \delta \widehat{NP_j} + \sum_{k=1}^{4} \sum_{m=1}^{4} \phi_{mk} (\mathrm{V}_{kj} - c_j)^m + v_j, \tag{2}$$

where Y represents any one of our outcomes; and v represents the error term.

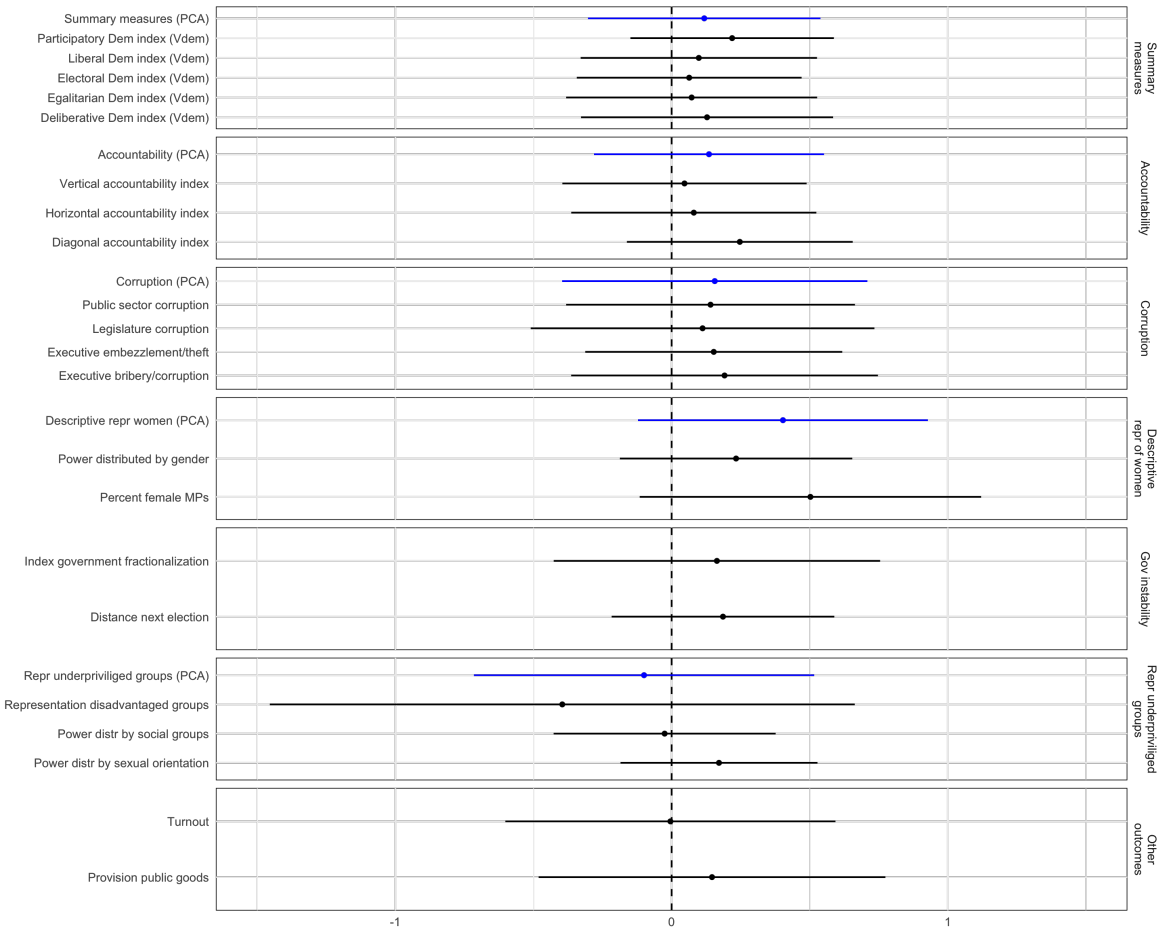
This strategy yields a total of sixteen different models (four instruments times four polynomials). Figure D7 shows the first stage for each of these specifications. Unfortunately, as the figure shows, the first stage in this strategy is often weak. The only model where we come clsoe to the threshold of $F = 9$ is for the model using a single instrument, with a single polynomial of distance to the threshold, and using the absolute number of parliamentary parties as the treatment variable. This finding is consistent with recent RDD literature, which recommends the use of a single polynomial (Gelman and Imbens 2019). We thus replicate the main analyses of Figure 4 using this specification.

The results are shown in Figure D8. As with the main analyses, we fail to fin any effect of the number of parties on the quality of democracy.



Notes:
Horizontal red line represents the threshold of $F = 9$.

Figure D7: First stage for sixteen possible specifications of our alternative estimation strategy (the treatment status of the parties closest to the threshold).



Notes: Lines represent 95% confidence intervals. Standard errors are clustered by country. All dependent variables are standardized.

Figure D8: Replication of the analyses shown in Figure 4 using an alternative approach to the estimation of the first stage.