杏林社会科学研究
第 34 巻 2 号
2 0 1 9 年 1 月

Did the Size of Municipal Legislatures Affect National Election Outcomes in Japan?

A Preliminary Investigation Using a Natural Experiment

Kota Matsui

1 Introduction

Local political elites often play a significant role not only in local but also in national politics. They provide a pool of ambitious and experienced human resources from which national political parties recruit candidates for national elections (Fiorina 1994), mobilize and monitor voters (Stokes 2005), and serve as a channel to distribute particularistic benefits to districts (Scheiner 2005). Despite its importance, the relationship between local politicians and national political phenomena has been understudied. Using a natural experimental setting in a Japanese election, this article examines an important dimension of such connections; the extent to which the existence of local politicians influence national election outcomes.

In Japan, the Liberal Democratic Party (LDP) has largely dominated national politics since the party was founded in 1955. Although the LDP temporarily lost power in the early 1990s because of an internal conflict and a resulting split, it soon came back to power and maintained its leading party status until 2009 when the Democratic Party of Japan (DPJ) won the Lower House election in a landslide. A number of theories have been proposed to explain the reason the LDP was able to maintain its dominant status for such a long period, such as strategic failures of the Socialist Party, the fragmentation of the opposition parties as a result of the Single-Nontransferable-Vote (SNTV) electoral system, and Japanese culture.

Some of the more recent studies have focused on the LDP's strength at the local level to explain the long-time dominance in the national political arena. Scheiner (2005), for instance, argues that the LDP was able to maintain the majority status at the national level because it had an even stronger dominance at the local level, so that an army of conservative local politicians served as a mobilizer of voters for national LDP politicians in addition to providing experienced candidates. Local politicians had a strong incentive to support national LDP politicians because of the clientelistic structure of Japanese politics. That is, because most of the municipalities heavily depended on fiscal transfers from the central government where the budgetary resources are highly concentrated, municipal legislators had to compete for distributive benefits allocated by national LDP politicians so as to secure their own reelection (Horiuchi et al.2015).

Following the argument that national LDP politicians relied on local politicians' efforts for electoral mobilization and monitoring in return for distributive benefits, Horiuchi et al.(2015) hypothesize that the number of municipal legislators is positively associated with voter turnout and the governing parties' (LDP and Komeito) vote share at the national level [1)]. In order to deal with the common methodological challenges such as omitted variable biases and the endogeneity of the number of local politicians, Horiuchi et al.(2015) employed a difference-in-differences approach, in which they compared municipalities that experienced mergers between two elections in 2001 and 2007, with those that did not experience mergers. Since the number of municipal legislators was dramatically reduced in the newly-established municipalities while the number remained unchanged in municipalities whose borders remained intact, Horiuchi et al.(2015) argue that the larger drop in voter turnout and the governing parties' vote share in the newly-established municipalities between the two elections can be explained by the important role of local politicians as a vote mobilizer.

Although Horiuchi et al.(2015) provide insightful findings accompanied by a clever research design, the study still has potential theoretical and methodological problems, such as whether or not municipalities that chose mergers are really comparable to those that didn't, even after controlling for several observed confounders. The decision to merge with neighboring municipalities is basically made by the municipalities themselves, which makes the self-selection problem highly likely. Therefore, instead of the approach proposed by Horiuchi et al.(2015), this article proposes an alternative research design to study the effect of the number of municipal legislators on national election outcomes.

Specifically, this article exploits an exogenous variation of the legislature size generated by a national law (the Local Autonomy Act, or Chiho-jichi hou) in Japan, which determines the maximum size of the municipal legislature according to population in the municipality, as shown in Table 1. For instance,

Table 1: Population and municipality legislature size in 2008

Population	Maximum size of legislature allowed by law	Number of cities
0-49,999	26	249
50,000- 99,999	30	272
100,000-199,999	34	158
200,000-299,999	38	41
300,000-499,999	46	51
500,000-899,999	56	20
900,000-1,299,999	64	5
1,300,000-1699,999	72	4
1,700,000-2,099,999	80	1
2,100,000-2,499,999	88	1
2,500,000-	96	2

municipalities with population that is slightly smaller than 100,000 can have 30 legislators maximum, while those with a population of 100,000 can have 34 legislators maximum, which means an additional four seats for municipalities with a slightly higher population. If we can assume that municipalities just below such a population threshold are very similar to those just above the threshold with respect to both observable and unobservable characteristics, then it is possible to consider that the maximum size of the legislature is as-if randomly assigned among those municipalities. This as-if randomness creates an exogenous variation in municipal legislature size, which makes it possible to credibly estimate the causal effect of the legislature size [2)]. More specifically, this article tests the hypothesis that the size of the municipal legislature is, ceteris paribus, positively associated with turnout and the governing parties' (LDP and Komeito) vote share at the national level.

A similar discontinuity design is employed by Pettersson-Lidbom (2012), which uses an exogenous variation of the municipal legislature size determined by national laws in Sweden and Finland to study the effect of the legislature size on the size of the municipal government's budget. This article is similar to Petterson-Lidbom (2012) in spirit, and as explained later, we follow the global polynomial specification proposed by Pettersson-Lidbom (2012). However, this research is different in some important respects. First, the outcomes of interest are not the size of municipal governments, but rather the electoral outcomes at the national level. To the best of the author's knowledge, this is one of the first studies in political science that makes use of population thresholds of municipalities to examine the relationship between local politicians and national politics [3)]. Second, a variety of approaches are used to analyze the data with discontinuities and to compare the results. This not only helps evaluate the robustness of the findings, but also contribute to the methodological literature on how best to use regression discontinuity designs (Hahn et al. 2001; Imbens and Lemieux 2008; Green et al. 2009; Dunning 2012).

The remainder of the article is organized as follows. The second section describes the data and empirical models. The third section reports the results, and the fourth section assesses the assumptions of the RD approach to validate the design. The last section summarizes the findings and discusses their implications.

2 Data and Empirical Framework

This article uses the 2007 Upper House election results [4)]. The electoral system used in the election was a combination of the nationwide open-list proportional representation (PR) system and the prefecture-based SNTV districts with the district magnitude ranging from one to five. The analysis uses the governing parties' vote share and voter turnout [5)] in the PR portion [6)]. Since the PR election is conducted at nationwide at-large district, the election result is aggregated at the municipality level for the analysis. Descriptive statistics of the key variables are reported in Table 2.

The data for the independent variable, municipal legislature size, is provided by a survey conducted by the National Association of Chairpersons of City Councils in 2008. Since this association is composed only of cities (about 800), legislature size data for towns (about 750) and villages (about 190) are

Table 2: Summary statistics

Variable	Obs.	Mean	Std.dev.	Minimum	Maximum
Population	804	142521.3	250526.8	4866	3645507
Governing parties (LDP + Komeito) vote share	804	.4148	.0601	.2084	.6911
LDP vote share	804	.2910	.0623	.1450	.5936
Turnout	804	.6039	.0552	.4698	.8132
Number of eligible voters	804	115374.4	201463	4534	2927214

not available. The hypothesis is based on the idea that the LDP's electoral strength has been augmented by local politician's electioneering efforts, and their ability to mobilize and monitor voters is probably higher in less-populous, rural areas (Saito 2010). Therefore, the data limitation might pose an external validity problem, because the vast majority of the cities have larger population than towns and villages, where local politicians' electioneering abilities might matter the most [7)]. However, our data covers every city, whose population ranges from less than 5000 to over 3 million. Since municipal level breakdown of the municipal legislature size in Japan was not publicly available prior to the survey noted above (cf. Horiuchi et al.2015), this article is one of the first systematic studies to use the actual size of legislatures to examine the effect of size on outcomes.

The empirical approach of this article is based on a regression discontinuity (RD) design. In recent years RD approach has increasingly attracted attention in political science and other social sciences as a powerful design to draw secure causal inferences. However, as some researchers have pointed out, the current state of literature is characterized by the lack of consensus on how best to analyze data with discontinuities generated by arbitrary thresholds. Figure 1 places various approaches that have been proposed on a line representing a continuum from the most 'design-based' to the most 'model-based' approaches.

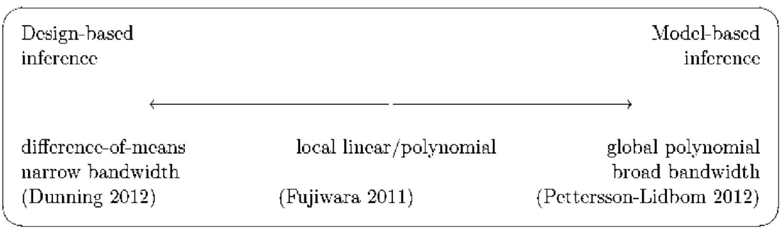

Figure 1: Variety of RD approaches

First, some researchers recommend to use the simple difference in means as the most credible estimator to analyze such data (Dunning 2012). If the design is very strong and a natural experimental interpretation really holds, that is, units just below the threshold and those just above are truly exchangeable, then the simple difference in means should be unbiased and there is no need for additional statistical adjustment. Dunning (2012) goes further to argue that no ex-post statistical adjustment can make inferences more credible if we only have less-than-ideal designs in which the exchangeability assumption is implausible or only weakly holds.

In contrast to this perspective, standard practice currently used is to employ local linear and local polynomial regressions to control for the forcing (and other potentially confounding) variables (Hahn et al. 2001; Imbens and Lemieux 2008). Advocates of this approach suggest that the difference-of-means estimator is biased if the derivatives of potential outcome functions on either side of the key threshold are non-zero in the limit at the threshold. For example, Imbens and Lemieux (2008) argues that "we typically do expect the regression function to have a non-zero derivative, even in cases where the treatment has no effect. In many applications the eligibility criterion is based on a covariate that does have some correlation with the outcome . . . Hence it is likely that the biases for the simple kernel estimator is relatively high." Even if the derivatives are non-zero, causal effect can be identified if the potential outcome function is smooth around the threshold, i.e., there is no jump. When local linear and polynomial regressions are used, the causal effect is defined as the difference between the two quantities: the value of the regression function of potential outcomes under treatment at the threshold, and the value of the regression function of potential outcomes under control at the threshold [8].

A third approach is the use of global polynomial regressions. Since both of the first two approaches only use observations within a certain window around the key threshold, the estimator can be inefficient due to the small number of observations. Thus the advantage of this approach is to gain efficiency by using the full sample and controlling for the forcing variable (i.e. population in this case) with higher-order polynomials (Pettersson-Lidbom 2012; see also Angrist and Lavy 1999). Similarly to the second approach, the causal effect is defined as the difference between the limits of regression functions at the threshold. The negative side of this approach is a possibility of introducing bias from using observations that are very far from the threshold. Observations far from the key threshold tend to have very different values of covariates and potential outcomes from those near the threshold, so the exchangeability assumption is unlikely to hold. Therefore, although this approach still exploits discontinuities at key thresholds, it is considered to be more 'model-based', rather than 'design-based', compared to the first two approaches.

In addition to the modeling strategies discussed so far, RD analyses also require bandwidth choice. Choosing the optimal bandwidth is difficult because there is a trade-off between bias and efficiency. On the one hand, as the bandwidth becomes narrower, municipalities just below and just above the threshold will more likely be similar on average, but the number of observations can often be very small, which makes estimation inefficient. On the other hand, larger bandwidths make estimation more efficient by increasing the number of observations, but it might also introduce more bias because we do not have good counterfactuals for municipalities far above and far below the threshold. In terms of bandwidth, the global polynomial regressions can be considered as a special case in which the bandwidth is infinitely large.

Since each approach has both strengths and weaknesses, and there is still little consensus in literature on the best practice, our strategy in this article is to employ all of the three approaches and compare the results they produce. By doing so we can examine the robustness of the findings, and if the results are sensitive to the choice of estimation frameworks, it would suggest the necessity of more studies on the advantages and disadvantages of the various approaches.

Furthermore, the institutional context analyzed in this article offers two more methodological issues to consider. First, unlike many of the studies that utilize a discontinuity generated by only one threshold, the subject of this article has multiple key thresholds, as shown in Table 1. In general this does not require any modification regarding the basic idea of RD and the estimation framework. As discussed below, however, a choice has to be made either to estimate causal effects separately for each threshold or to combine data from different thresholds.

Second, although the national law sets the maximum size of municipal legislatures, municipalities still can determine the actual size of their legislatures as long as it is below the maximum stipulated by the law. Therefore, the population thresholds creates "fuzzy" discontinuities, in which the treatment assignment only probabilistically affect the actual size of the municipal legislature [9]. Following literature, the intention-to-treat (i.e. whether municipalities are above or below a certain key threshold) is used as an instrumental variable (IV) for an endogenous regressor (i.e. actual size of the legislature) [10].

Based on the discussion above, the rest of this section presents our specifications in more detail. First the simple difference-of-means is estimated:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 B_{2i} + \beta_3 B_{3i} + \cdots + \beta_{10} B_{10i} + \epsilon_i \tag{1}$$

where Y_i is the outcome variable for municipality i, T_i is a treatment indicator that takes the value of one if the municipality i is just above the threshold, and zero if below the threshold. B_{ji} is a dummy variable that takes the value of one if the municipality i belongs to block j, and 0 otherwise [11]. Here the blocks correspond to different thresholds, i.e., block 1 includes municipalities near the threshold population of 50,000, block 2 includes those around population 100,000 and so forth. The idea of including those dummies is that the natural experiment studied here is analogous to a block randomized experiment in which randomization is conducted within each block that is defined prior to the randomization [12]. Equation (1) is the intention-to-treat (reduced form) estimation, and β_1 is the coefficient for the mean difference (ITT effect). As noted above, an IV model is also estimated that uses the actual size of the legislature in municipality i as an endogenous regressor and T_i as an instrument. Each model is estimated with various bandwidths (5000, 10000, 15000, 20000, 25000).

Second the IV models are employed using local linear and local polynomials in the form:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 (pop_i - threshold) + \beta_3 (pop_i - threshold) \hat{X}_i + \epsilon_i \quad (2)$$

where $\hat{X}_i$ is the predicted legislature size from the first stage regression, pop_i represents population in municipality i, and $threshold$ represents the number of population used as the threshold (e.g. 50,000) [13]. Similarly to the difference-of-means estimation above, the treatment indicator T_i is used as an instrument for X_i. The coefficient of our primary interest is β_1, which represents the complier average treatment effect at the threshold. In the main analysis in this article, models are estimated separately for each threshold with various bandwidths (5000, 10000, 15000, 20000, 25000). This is because simply estimating the equation (2) with the combined data from multiple thresholds will assume that the derivative of the regression functions in the limit at the threshold is the same for every threshold, which is not necessarily plausible [14].

Lastly, the global polynomial models are estimated. The first stage regression is the form:

$$X_i = \gamma + \psi_{30} Z_{30i} + \psi_{34} Z_{34i} + \cdots + \psi_{96} Z_{96i} + g(pop_i) + \eta_i \quad (3)$$

where X_i is the actual legislature size of municipality i, and $g(pop_i)$ is the function used to control for the population. Unlike the other approaches, dummy variables Z_{ji}, which indicate the population category of municipality i, are used as instruments for the size of legislature, X_i. For example, Z_{30i} takes the value of one if municipality i has population above 50,000 and below 99,999, and zero otherwise. Similarly, Z_{96i} takes the value one if municipality i has population above 2,500,000 and zero otherwise[15]. The assumption underlying the identification strategy is that once we control for the forcing variable (i.e. population), the instruments (i.e. Z_j) are considered exogenous. In other words, the population groups capture some portion of the variation in legislature size that is left unexplained by population, which provides the exogenous variation of legislature size that can be used to estimate the causal effect. Then the second stage regression is:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + f(pop_i) + \epsilon_i \quad (4)$$

where Y_i is the outcome variable, $\hat{X}_i$ is the predicted legislature size from the first stage, and $f(pop_i)$ is the function used to control for the population. The coefficient of our interest is β_1. In order to gain efficiency, all of the observations are used to estimate the model[16].

3 Results

This section presents the findings from the models described above. Table 3 shows the intention-to-treat (ITT) effect of the legislature size estimated using the difference-of-means models (Equation (1)). As for the governing parties' vote share, the effect does not reach the statistical significance at 0.1 level when the bandwidth is either 5,000 or 10,000. However, as the bandwidth gets larger, the results becomes highly statistically significant, while the size of the coefficients also becomes somewhat larger. For instance, when the bandwidth is 15,000, the predicted vote share of the governing parties is on average 1.4 percentage points (p.p. hereafter) lower in municipalities just above the thresholds than in municipalities just below the thresholds. Similarly, the effect of ITT on voter turnout becomes statistically more significant as the bandwidth becomes larger, and the substantive size of the coefficient also gets larger. When the bandwidth is 15,000, voter turnout is predicted to be 1.5 p.p. lower in the municipalities above the thresholds than in those below.

Table 3: Difference-of-means: Intention-to-treat effect

Bandwidth	5,000	10,000	15,000	20,000	25,000
dependent variable	(1)	(2)	(3)	(4)	(5)
Governing parties'	-.00165	-.00862	-.01405**	-.01492***	-.01748***
vote share	(.01069)	(.00807)	(.00632)	(.00518)	(.00477)
Turnout	-.00237	-.01358*	-.01507**	-.01925***	-.02092***
	(.01133)	(.00745)	(.00591)	(.00479)	(.00443)
Observations	121	255	370	518	606

Robust standard errors in parentheses

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Turning to the instrumental variable approach, Table 4 shows the results from the estimation of the additional legislator effects on the vote share and turnout. The coeffcients of the legislature size for both dependent variables reach the conventional level of statistical significance when the bandwidth

Table 4: Difference-of-means: IV estimation of additional legislator effect

Bandwidth	5,000	10,000	15,000	20,000	25,000
dependent variable	(1)	(2)	(3)	(4)	(5)
Governing parties' vote share	-.00132	-.00824	-.00912*	-.00896**	-.00893***
	(.00858)	(.00940)	(.00533)	(.00406)	(.00315)
First-stage F-test	2.26	3.51	10.95	17.97	28.73
Turnout	-.00189	-.01298	-.00978*	-.01157**	-.01068***
	(.00927)	(.01165)	(.00558)	(.00457)	(.00345)
First-stage F-test	2.26	3.51	11.95	17.97	28.73
Observations	121	255	370	518	606

Robust standard errors in parentheses

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

becomes large. Also, the first-stage F-test indicates that the instrument is reasonably strong when the bandwidth is relatively large, but it becomes weak when the bandwidth is 10,000 or narrower. The absolute size of the effect is highly stable with respect to the bandwidth choice. In the case of the 15,000 bandwidth, one additional legislator is associated with 0.91 p.p. decrease in the vote share and 1.3 p.p. decrease in turnout. In order to see the impact of these effects more substantively, the coefficient could be multiplied by the number of additional legislators when a municipality crosses a threshold. For instance, as shown in Table 1, a municipality can add four legislators if population increases from 99,999 to 100,000. In this hypothetical scenario, the municipality would, on average, experience 3.7 p.p. decrease in the governing parties' vote share and 3.9 p.p. decrease in voter turnout.

Next the results from local linear and polynomial models are presented (Equation (2)). Table 5 shows the estimated additional legislator effect at the 50,000 population threshold. All of the coefficients are far from statistically significant regardless of the specifications and the bandwidths. Also, the first-stage F-test suggests that the treatment indicator is a very weak instrument for the actual size of the legislature. Therefore it is difficult from the results

Table 5: Local linear/polynomial estimation: population 50,000 threshold

Specification/ bandwidth	Linear 10,000	Linear 20,000	Quad. 10,000	Quad. 20,000	Cubic 10,000	Cubic 20,000
	(1)	(2)	(3)	(4)	(5)	(6)
Governing parties'	-.00336	.00051	-.00556	-.00276	.02727	-.09043
vote share	(.02249)	(.01264)	(.03016)	(.02608)	(.10402)	(.63904)
First-stage F-test	0.00	1.28	1.10	0.29	0.00	0.00
Turnout	.01054	-.00037	.02046	.01612	.01834	.07551
	(.06758)	(.01152)	(.03599)	(.03658)	(.08342)	(.55249)
First-stage F-test	0.00	1.28	1.10	0.29	0.00	0.00
Observations	172	346	172	346	172	346

Robust standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

to argue that the treatment had some effects. Similarly, Table 6 and Table 7 show the results for population 100,000 threshold and population 200,000 threshold, respectively. The results are also statistically insignificant and the F-tests imply that the instrument is very weak. The results for the other population thresholds are not shown here because the number of observations is quite small and no statistically and substantively significant results are found.

Lastly, the results from the global polynomial model are presented (Equations (3) and (4)). Table 8 shows the results for the governing parties' vote share. The OLS specification in Column 1 indicates that there is a negative association between the legislature size and the governing parties' vote share without any controls, but the relationship disappears once municipality population is controlled (Column 2). This suggests that there is a negative relationship between municipality population and the vote share of the governing parties, which is not surprising because rural areas have traditionally been the LDP's stronghold. Turning to the RD specifications from Column 3 to 7, negative and highly statistically significant effects of the legislature size are found across various degrees of polynomials. The first-stage F-test suggests the instrument is sufficiently strong in models that use first, second, third, or

Table 6: Local linear/polynomial estimation: population 100,000 threshold

Specification/ bandwidth	Linear 10,000	Linear 20,000	Quad. 10,000	Quad. 20,000	Cubic 10,000	Cubic 20,000
	(1)	(2)	(3)	(4)	(5)	(6)
Governing parties'	-.00333	-.00385	.03496	.02751	-.01425	.03492
vote share	(.02021)	(.01092)	(.04789)	(.04307)	(.51905)	(.07697)
First-stage F-test	0.85	1.49	0.36	0.44	0.00	0.00
Turnout	.01270	.01006	.00035	-.00556	.00223	-.00468
	(.01277)	(.00790)	(.01323)	(.01914)	(.04497)	(.02658)
First-stage F-test	0.85	1.49	0.36	0.44	0.00	0.00
Observations	58	123	58	123	58	123

Robust standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 7: Local linear/polynomial estimation: population 200,000 threshold

Specification/ bandwidth	Linear 10,000	Linear 20,000	Quad. 10,000	Quad. 20,000	Cubic 10,000	Cubic 20,000
	(1)	(2)	(3)	(4)	(5)	(6)
Governing parties'	-.00327	.00950	-.00931	-.01109	-.07398	.00150
vote share	(.00758)	(.01693)	(.01950)	(.01174)	(.92589)	(.05685)
First-stage F-test	0.70	0.65	0.15	0.41	0.00	0.00
Turnout	.00076	.00352	-.00853	.00655	-.10639	.02150
	(.00693)	(.01035)	(.02436)	(.00874)	(1.45041)	(.09764)
First-stage F-test	0.70	0.65	0.15	0.41	0.00	0.00
Observations	17	28	17	28	17	28

Robust standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 8: Global polynomial estimation: governing parties' vote share

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS		RD				
Governing parties'	-.00099***	-.00016	-.00415***	-.00652***	-.00839***	-.01050***	-.00712**
vote share	(.00018)	(.00041)	(.00061)	(.00130)	(.00259)	(.00353)	(.00289)
Degree of polynomial in population	None	First	First	Second	Third	Fourth	Fifth
First-stage F-test			103.55	93.29	300.66	123.17	5.11
Hansen's J-test			20.821	15.659	9.852	9.157	7.823
χ^9 p-value			0.0135	0.0744	0.3626	0.4229	0.5521
Observations	804	804	804	804	804	804	804

Robust standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

fourth polynomials. The substantive size of the coefficient is not very different from those from the difference-of-means estimation in Table 4. The results for voter turnout are shown in Table 9. The OLS result in Column 1 shows that the legislature size is negatively associated with voter turnout, but the relationship disappears once population is controlled as shown in Column 2. The RD specifications from Column 3 to 7 indicate that additional legislators predict lower voter turnout. The effect reaches the statistical significance at 0.01 level, and the absolute size of the coefficient is relatively stable and similar to those from the difference-of-means estimation in Table 4.

In sum, negative and statistically significant results appear when either the simple difference-of-means or the global polynomials are used, but no significant result was found from the local linear and local polynomial estimations. One possible explanation for the divergent results is that the local linear and local polynomial estimations conducted here are inefficient because they are estimated separately for each threshold, thus making the number of observations smaller than those in the other approaches.

Another possible explanation is that the results from the difference-of-means and the global polynomial specifications could be biased. First, the simple difference-of-means is valid if units are truly exchangeable among the study group, i.e., there are no confounders. In other words, the rationale of using the simple difference-of-means is that, if we have a good natural experimental design, we need not control for the forcing variable (i.e. population) as well as other covariates, because in expectation they should not be different between the treatment and the control group (Dunning 2012). However, as Imbens and

Table 9: Global polynomial estimation: turnout

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS		RD				
Governing parties' vote share	-.00115***	-.00007	-.00511***	-.00765***	-.01007***	-.01206***	-.00912***
	(.00016)	(.00042)	(.00055)	(.00112)	(.00219)	(.00356)	(.00289)
Degree of polynomial in population	None	First	First	Second	Third	Fourth	Fifth
First-stage F-test			103.55	93.29	300.66	123.17	5.11
Hansen's J-test			25.031	14.703	10.185	7.753	8.536
χ^9 p-value			0.0029	0.0994	0.3357	0.5592	0.4811
Observations	804	804	804	804	804	804	804

Robust standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Lemieux (2008) point out, this assumption might not always hold in practice, and the derivative of the potential outcome function might not be zero in the limit even within a narrow band. In such cases, the simple difference-of-means can be biased as it does not deal with confounding. In the context of this article, population is negatively correlated with the governing parties' vote share and voter turnout, as the OLS results in Table 8 and 9 indicate. If there is a negative relationship between population and the outcome variables even within the bands, not controlling for population can lead to a spurious relationship between the legislature size and the outcome variables.

As for the global polynomials, the results might be biased by omitted confounders, as no covariate is included in the estimation. The strength of RD as a design-based empirical strategy is that the exogeneity assumption of a key independent variable becomes plausible near some arbitrary thresholds because of as-if randomness that they generate. Although the global polynomial specification still exploits the exogeneity provided by the arbitrary thresholds, it also gains statistical power from observations far from the thresholds, and relies on modeling assumptions more heavily than the other two approaches. Therefore collecting and controlling for covariates will be a next step for our future analysis.

4 Checking RD Assumptions

The assumption of our identification strategy is that municipalities close to the thresholds are assigned as-if randomly, so that the potential outcome function is smooth and does not present a jump at the thresholds. If, for example, municipalities just below the thresholds tried to over-report population so that the municipalities can have larger legislatures, then the RD assumption would be violated. In order to probe the possibility, Figure 2 implements an exercise suggested by McCrary (2008). If strategic manipulation has taken place, it would likely reflect in a jump close to the thresholds. If municipalities just below the thresholds actually over-reported population in order to cross the thresholds, then we would see surprisingly many municipalities just above the thresholds and surprisingly few below the thresholds. Figure 2 suggests that there is no clear evidence that indicates such manipulation, as the density function of population has no obvious discontinuities.

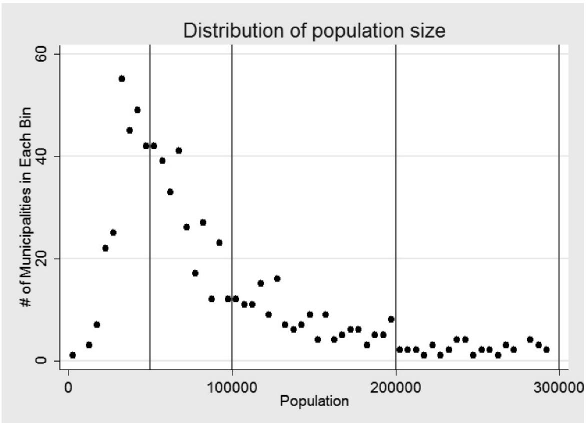

Figure 2: Distribution of municipality population

Next a placebo test is conducted to further examine the validity of the design. Here it is checked whether the number of eligible voters in municipalities[17], which should not be affected by the legislature size, is actually similar on each side of the key thresholds. Table 10 shows the results from difference-of-means models. The coefficients are substantively large and highly statistically significant. It might not be surprising, however, for population is not controlled in the difference-in-means specifications, whereas population and the number of eligible voters are highly correlated. Similarly, Table 11 shows the

Table 10: Difference-of-means: Number of eligible voters

Bandwidth	5,000	10,000	15,000	20,000	25,000
	(1)	(2)	(3)	(4)	(5)
Number of eligible voters	3141.117 (1955.954)	7765.384* (4020.425)	7419.263*** (2189.931)	9551.432*** (2233.738)	9582.445*** (1764.798)
First-stage F-test	2.26	3.51	10.95	17.97	28.73
Observations	121	255	370	518	606

Robust standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 11: Global polynomial estimation: Number of eligible voters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS		RD				
Number of eligible voters	16874.63*** (1888.701)	107.654* (56.23843)	234.6683 (156.5397)	-238.1935 (429.7906)	-1253.456*** (329.1423)	-842.1402 (736.9942)	-1404.975* (838.4072)
Degree of polynomial in population	None	First	First	Second	Third	Fourth	Fifth
First-stage F-test			103.55	93.29	300.66	123.17	5.11
Hansen's J-test			18.061	7.723	3.479	3.505	3.115
χ^9 p-value			0.0345	0.5622	0.9423	0.9409	0.9596
Observations	804	804	804	804	804	804	804

Robust standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

results from global polynomial specifications. Unsurprisingly, the OLS result without any control (Column 1) indicates that the number of eligible voters is highly and positively associated with the legislature size, which is positively associated with population. Thus, once population is taken into account, the coefficient becomes much smaller and statistically less significant (Column 2). Turning to RD models, the coefficients for the legislature size do not reach the conventional level of statistical significance when the degree of polynomials is either first, second, or fourth. When the degree is fifth, the first-stage F-test suggests that the instrument is very weak and unreliable. Also, the absolute size of the effect is substantively small across the model specifications. However, the coefficient reaches the statistical significance at 0.001 level when cubic polynomials are used. Although the substantive size of the coefficient is not very large, it suggests either that the estimation results might be biased for some reason, or that some form of manipulation has taken place. In any case, more covariates will need to be collected in order to further verify the validity of the RD design.

5 Conclusion

This article employed a variety of RD approaches to examine the causal effect of the number of local legislators on voter turnout and the governing parties' vote share in a Japanese national election. Contrary to our hypothesis, some of the results suggest that the size of legislature is negatively associated with the governing parties' vote share and turnout. However, the results are sensitive to the choice of estimation strategies. While both the most 'design-based' approach (difference-of-means) and the most 'model-based' approach (global polynomial regressions) detect statistically significant and substantively large effects, significant results are not found when the local linear and local polynomial regressions are used.

It is difficult to determine without an experimental benchmark which results are the closest to the truth, but some of the problems can be pointed out that might have made the results unstable. First, the national law that determines the maximum size of legislatures does not have a strong effect on the actual legislature size, as suggested by the weak instrument tests. That is, because municipalities have discretion over the actual size of their legislature as long as it is below the maximum stipulated by law, exogenous variation captured by the arbitrary thresholds could be small. This problem is analogous to the one a randomized controlled trial faces when the compliance rate is low. Second, the substantive magnitude of the intervention at the thresholds is not very large in the subject studied here. For example, the effect of adding four legislators in a municipality that already has thirty might not be substantively large enough to clearly detect the effect of local politicians. Investigating alternative research designs and novel data will be the next step so that results can be compared for more credible inferences.

References

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2), 533–575.

Dunning, T. (2012). *Natural Experiments in the Social Sciences*. Cambridge University Press.

Ferreira, F., & Gyourko, J. (2009). Do Political Parties Matter? Evidence from US Cities. *The Quarterly Journal of Economics*, 124(1), 399–422.

Fiorina, M. P. (1994). Divided Government in the American States: A Byproduct of Legislative Professionalism?. *American Political Science Review*, 304–316.

Fujiwara, T. (2011). A Regression Discontinuity Test of Strategic Voting and Duverger's Law. *Quarterly Journal of Political Science*, 6(3–4), 197–233.

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. WW Norton.

Gerber, E. R., & Hopkins, D. J. (2011). When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy. *American Journal of Political Science*, 55(2), 326–339.

Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., & Larimer, C. W. (2009). Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks. *Political Analysis*, 17(4), 400–417.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression- Discontinuity Design. *Econometrica*, 69(1), 201–209.

Horiuchi, Y., Saito, J. & Yamada, K. (2015). Removing Boundaries, Losing Connec- tions: Electoral Consequences of Local Government Reform in Japan. *Journal of East Asian Studies,* 15, 99–125.

Imbens, G. W., & Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2), 615–635.

Katsumata, H. (2016). Chuusenkyoku-sei no senkyo-kyousou ni okeru inga-kouka no suitei (Estimating the causal effect in electoral competition in the medium-sized constituency system) . Working paper presented at the 2016 Japan Association for

Com- parative Politics annual meeting. June 25, 2016.

Pettersson-Lidbom, P. (2012). Does the Size of the Legislature Affect the Size of Government? Evidence from Two Natural Experiments. *Journal of Public Economics*, 96(3), 269–278.

Saito, J. (2010). *Jiminto Choki Seiken No Seiji Keizaigaku [The Political Economy of the LDP Regime]*. Tokyo: Keiso Shobo.

Scheiner, E. (2005). *Democracy without Competition in Japan: Opposition Failure in a One-party Dominant State*. Cambridge University Press.

Stokes, S. C. (2005). Perverse Accountability: A Formal Model of Machine Politics with Evidence from Argentina. *American Political Science Review*, 99(3), 315.

---

1) This term is used here because the LDP has formed a coalition with the Komeito Party since 1999.

2) It is important to note, however, that municipalities can determine the actual size of their legislatures as long as it is below the maximum stipulated in the national law. Therefore, as explained later in detail, the national law provides “fuzzy” discontinuities.

3) This article was originally written and presented in PLSC508 at Yale University in May 2013. Readers are suggested to examine Katsumata (2016) as well, which is a study on the same topic using different data and modified methodology with knowledge of this article.

4) The author is grateful to Kyohei Yamada for sharing the election data.

5) As already noted, since the LDP formed coalition with Komeito since 1999, the combined vote share for the LDP and Komeito is used. Therefore, the governing parties' vote share is calculated by dividing the total number of votes for the LDP and Komeito by the total number of votes. Voter turnout is the total number of votes divided by the number of eligible voters.

6) The SNTV portion could also be used because district (i.e. prefecture) specific characteristics should not matter as long as the RD assumptions hold. Assessing the robustness of the findings using the SNTV election results will be a next step for future research.

7) One of the (non-strict) criteria for towns and villages to be a city is to have population of more than 30,000.

8) See Fujiwara (2011), Ferreira and Gyourko (2007), and Gerber and Hopkins (2011) as

some of the applications.

9) In contrast, we would have a "sharp" discontinuity if a forcing variable deterministically assigns the treatment status.

10) In order to for the IV estimation to be valid, we have to assume that there is no defiers (monotonicity assumption). That is, no municipality just above a threshold would have a larger legislature if the municipality was just below the threshold.

11) B_1 is used as the baseline category.

12) In general, the estimator from a block randomized experiment is not equivalent to the mean difference among all subjects in treatment and control, which can be biased. This occurs when the probability of being assigned to the treatment group varies by block (Gerber and Green 2012).

13) Population is thus centered at the threshold so that the interpretation of the result will be simple, because the terms other than β_1, which represents the causal effect, will disappear at the threshold.

14) This assumption can be relaxed by using interaction terms for every threshold, though the specification will be more complicated.

15) Municipalities of population below 49,999 are used as the baseline category.

16) Since the possibility of omitted variable bias is high as observations far from thresholds are used, it would be desirable to control for other observable confounders as well. It will be a task for future research.

17) Every adult (20 years old or older) who lives in the municipality is eligible to vote.

(Accepted Date: 9 May 2018)