Master Thesis:

A Causal Test of the "Law of 1/n" and its Mechanisms

Written by: Florian Fox
Course of Study: M. Sc. Volkswirtschaftslehre/Economics
Number of Words: 16,082

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Some of the abbreviations have only been used in tables or figures.

ATE	average treatment effect
BW	Baden-Württemberg
BY	Bavaria
CDU	Christlich Demokratische Union
CER	coverage error rate
CSU	Christlich-Soziale Union
Destatis	Federal Statistical Office
diff-in-disc	difference-in-discontinuities
E.	election
FE	fixed effects
IV	instrumental variable
HE	Hesse
het.	heteroskedastic
HHI	Herfindahl-Hirschman index
mcp.	municipality
MSE	mean squared error
MV	Mecklenburg-Vorpommern
NW	North Rhine-Westphalia
Obs.	observations
OLS	ordinary least squares
p. c.	per capita
RD	regression discontinuity
RDD	regression discontinuity design
RP	Rhineland-Palatinate
s.	since
SL	Saarland
SPD	Sozialdemokratische Partei Deutschlands
ST	Saxony-Anhalt
SE	standard error
SH	Schleswig-Holstein
SN	Saxony

SSC	Boston College Statistical Software Components
TE	treatment effect
TH	Thuringia
thr.	threshold
U. S.	United States
w/	with
w/o	without
yr.	year
ZEW	Leibniz Centre for European Economic Research

1 Introduction

This paper expands the empirical investigation of the effect of legislature size on public spending ("Law of 1/n") using regression discontinuity design (RDD) to a 13-year panel of German municipalities, without finding such an effect. Additionally, it offers insights into potential drivers of the effect: The peculiarities of the German local political system – a strong mayoral position and at-large council elections – may limit the size of the effect. Previously, no paper had tested a variety of potential mechanisms.

Standard theory (Weingast et al., 1981) and many "early" empirical papers relying on cross-sectional variation long hold that public spending increases if council size increases. Self-interested legislators are assumed to maximize their constituencies' net welfare by securing funding paid for by other constituencies, which creates an externality. This conjecture has repeatedly been questioned by more credible empirical methods such as RDDs. I give more evidence by extending the time dimension of the data sample over the ones previously reported for Germany, which focused on single states or short-running panels, yielding mixed results (Egger & Koethenbuerger, 2010; Höhmann, 2017; Holzmann & Zaddach, 2019). I use long-running data from 11 German states.

In Germany, council size depends on population-based thresholds, at which council size increases. The general idea of the RDD in this setting is to compare municipalities whose population figures fall just short of the council size increasing threshold with those that barely pass the cutoff. Only the latter are now "treated" in that their council size is increased but the two groups are not fundamentally different except for the treatment administration. The validity of these population-threshold based RDDs has recently been questioned (Eggers et al., 2018a). If units manipulate whether they want to be assigned to the treatment or the control group, the research design does not allow for a causal interpretation. Using my expanded sample, I do not detect a general sorting pattern. However, I do find indications for large cities and municipalities whose population is older to sort to the right side of the cutoffs. Since many thresholds are used to determine multiple population-threshold based policies, it is important to distinguish these confounded thresholds from non-confounded ones. Only the latter allow for a clean identification of the treatment effect. Doing so, I detect a number of mistakes in Höhmann's (2017) classification of confounded thresholds that might be one driver of his finding of a negative effect of legislature size on public expenditure in the sample of non-confounded cutoffs.

Investigating heterogeneous treatment effects gives researchers a more nuanced understanding of the treatment effect (Reguly, 2021a, p. 2). For instance, if the effects of a

treatment exactly cancel out in two subgroups, treatment does have an effect that, however, cannot be detected in the average treatment effect (ATE). In addition, important policy considerations can be derived from such insights. Say, for example, that a strong position of mayors in the local political system is found to curb the "Law of 1/n". Then policymakers may use that knowledge to reduce the externality.

Examining a number of potential mechanisms, I show that most factors related to the political system, spending categories as well as councilors' and municipal characteristics do not exhibit significant patterns in their variation for the "Law's" treatment effect. However, the state-specific treatment effects are different for some states in both magnitudes and signs. Factors that have been argued to limit the "Law's" effect such as at-large elections and mayors with strong positions in the local political system may indeed lead to the overall null finding, albeit not convincing across methods. Moreover, mayors nominated by conservative parties are found to reduce spending at the cutoffs more than their counterparts nominated by different parties.

The remainder of this paper is organized as follows: I will start by presenting the theory behind the "Law" in section 2. Section 3 discusses the "conventional" empirical evidence, followed by the mechanisms behind the effect mentioned in the literature. Having derived the hypotheses to be tested later on in the empirical sections, the institutional background is presented in section 4. Potential concerns regarding the regression discontinuity (RD) validity will be tackled in section 6. The RDD results are provided in section 7, along with RDD Robustness Checks, and the investigation of potential mechanisms in section 8. Section 9 concludes.

2 Theoretical Considerations Behind the "Law of 1/n"

Weingast et al.'s (1981) model is the standard model for claims on (the direction of) the council-size effect on government expenditure. Being solely accountable to and elected by their own constituency, they seek to maximize their district's net welfare. Say, for instance, that the parliament decides on a project $P_j(x)$ in district j in which x is the size of the project to be determined by the lawmakers. In j, utility $b(x)$ is generated from the project, $\frac{\partial b}{\partial x} > 0, \frac{\partial^2 b}{\partial x^2} < 0$. As a consequence, the benefits are reaped purely locally. $c_i(x)$, $i = 1, 2, 3$, are the cost components, with $\frac{\partial c_i}{\partial x} > 0$, $\frac{\partial^2 c_i}{\partial x^2} \geq 0$. c_1 describes resource expenditures spent within the constituency, c_2 defines costs occurring outside the constituency, and finally external costs within the district are called c_3. The total cost sums up to $c(x) = \sum_{i=1}^{3} c_i(x)$. A tax $T(x) = c_1(x) + c_2(x)$ is levied in all n

constituencies to finance the total "visible" cost of $P_j(x)$, of which constituency i bears a share $t_i \geq 0$, with $\sum_{i=1}^{n} t_i = 1$ (pp. 644-645).

First, the authors use a benevolent lawmaker who maximizes total net utility as the benchmark (p. 646):

$$\max_x E(x) = b(x) - c(x). \tag{1}$$

Differentiating with respect to x yields:

$$\frac{\partial b}{\partial x} - \frac{\partial c}{\partial x} = 0. \tag{2}$$

When $\frac{\partial b}{\partial x} = \frac{\partial c}{\partial x}$, a project of size x^E is realized.

Upon politicization, x is no longer socially optimal (pp. 651-652) as politicians overspend. Maximizing their constituency's net benefit in order to augment reelection probability, councilors aim to secure funding for projects in their constituency which are paid for by taxes levied in all other constituencies. Formally, for legislator j:

$$\max_x N_j(x) = b(x) + c_1(x) - t_j\, T(x) - c_3(x) \tag{3}$$

Differentiating with respect to x:

$$\frac{\partial b}{\partial x} + \frac{\partial c_1}{\partial x} = t_j \frac{\partial T}{\partial x} + \frac{\partial c_3}{\partial x} \tag{4}$$

Equation 4 states the optimality condition of a project in j's constituency with size x^n (pp. 652-653). Typically (more exactly, for $\frac{\partial c_1}{\partial x} > t_j \frac{\partial T}{\partial x}$), $x^N > x^E$ (pp. 650-653) so in a politicized setting, project sizes and thus expenditures are larger than optimal ("pork barrel").

Regarding the financing side, let us assume that taxes are equally levied by all n electoral districts ("generalized taxation"), $t_i = 1/n$. All legislative proposals are passed in an omnibus bill ("universalism", Primo & Snyder, 2008, p. 477). This is what Weingast et al. (1981, p. 654) coin the "Law of 1/n".[1] This "Law" describes an externality based on legislators aiming to set up projects in their own constituency that other constituencies pay for by taxes ("common pool"). As the tax share t_i typically

---

[1]Hereinafter, I will use different terms such as "Law of 1/n" or "pork-barrel spending" to describe the effect that council size has on public expenditure.

decreases in the number of districts n, inefficiency $x^N - x^E$ increases in n. A range of other theoretical models, some of them also including legislative decision-making, predicts similar effects (Baron, 1991; Buchanan & Yoon, 2002; Chari & Cole, 1995). If every constituency is represented by an equal number of lawmakers, an increase in constituencies (i. e. a rise in the number of representatives) should result in larger government size.

Primo and Snyder (2008) expand upon the Weingast et al. (1981) model in that spending does not solely depend on one factor (council size) but on five different factors, one of which is council size. In their model, the effect sign of council size is not clear a priori and may even result in a reverse "Law of 1/n" in which more lawmakers translate into less expenditure.

Since electoral systems and setups vary widely, it is important to note that the Weingast et al. (1981) theory exclusively refers to ward elections. In these kinds of elections, the whole jurisdiction is sliced into a number of different constituencies, each of which is a separate electoral battle ground. Voters only vote in their own ward and their votes cannot (directly) affect elections in a different ward. In at-large elections, the opposite to ward (or district) elections, the whole jurisdiction votes on the members of parliament whose number is typically fixed in advance. In between, there are mixed systems combining both elements.

This gives us no clear sign prediction for at-large and mixed systems with respect to the "Law of 1/n". If lawmakers treat the whole jurisdiction as their constituency, no "pork barrel" seeking behavior should be expected. However, Baqir (2002, p. 1342) argues that even at-large elected representatives have "home bases" (Baqir, 2002, p. 1342) which they favor over other constituencies. These "home bases" need not be characterized by geographical but also cultural, ethnic, or economic factors.

Hypothesis 1: *The "Law of 1/n": Public expenditure increases in council size.*

Papers that empirically examine the "Law of 1/n" commonly refer to the Weingast et al. (1981) model as their theoretical prediction despite some of them being located in mixed or at-large electoral systems. These empirical papers are presented in the subsequent section, section 3.

An obvious question, assuming the "Law of 1/n" exists, is how municipalities finance extra spending through additional councilors: by taxes or by new debt (or both)? As an arbitrary choice, H2 assumes:

Hypothesis 2: *The "Law of 1/n" is financed via additional taxes (as opposed to additional debt).*

And closely related: If H2 holds, which taxes are used for funding the extra expenditure?

3 Empirical Evidence in the Literature

Let us now turn to the preexisting empirical evidence. The Weingast et al. (1981) model has been tested in various ways. Nevertheless, consensus on its validity has yet to emerge. In this literature review, I will focus on the effect that an increase in representatives has on public spending as this is what will be examined empirically further below, with a special emphasis on the methodology employed to arrive at these conclusions. Subsequently, I will present theoretical considerations and empirical evidence on potential mechanisms.

3.1 The "Law of 1/n"

3.1.1 "Conventional" Empirical Literature

The "conventional" literature has used ordinary least squares (OLS) regressions or variations of it to ascertain the causal effects of the "Law of 1/n". A basic regression equation that includes covariates reads as follows:

$$Y_{it} = \alpha + \beta CS_{it} + \gamma X'_{it} + \varepsilon_{it} \tag{5}$$

with Y_{it} as public spending of municipality i in year t, CS_{it} as i's council size in t, and X'_{it} as the vector of covariates.

From an econometric point of view, however, two key problems arise. The first is bias from omitted variables. If for an omitted variable $w_{it} \notin X'_{it}$:

$$cov(w_{it}, Y_{it}) \neq 0 \text{ and } cov(w_{it}, CS_{it}) \neq 0, \tag{6}$$

then $\beta \neq \hat{\beta}$, i. e. the estimated coefficient from the regression $\hat{\beta}$ is not equal to its true but unknown parameter β: It is biased, even in large samples. This problem is not necessarily sufficiently addressed by a large number of control variables (Höhmann, 2017, p. 347). It is aggravated by the fact that the direction of the bias is not known ex ante; it depends on the covariances in equation 6. If, for instance, a constituency's voters prefer both a larger council size and more government spending, the correlation

between public budget and number of representatives might be spurious (Egger & Koethenbuerger, 2010, pp. 201, 204), giving us $\hat{\beta} > \beta$, that is an overestimate.

The second econometric complication lies in reverse causality. As Pettersson-Lidbom (2012, p. 269) argues, an expansion of the public budget may result in a more complex budget bill negotiation, thus requiring more lawmakers. Hence, not only equation 5 holds but also:

$$CS_{it} = \zeta + \eta Y_{it} + \theta X'_{it} + \iota_{it} \qquad (7)$$

with $\eta \neq 0$.

Inserting equation 5 in 7 yields:

$$CS_{it} = \frac{\zeta + \eta \alpha}{1 - \eta \beta} + \frac{(\eta \gamma + \theta) X'_{it}}{1 - \eta \beta} + \frac{\eta \varepsilon_{it} + \iota_{it}}{1 - \eta \beta} \qquad (8)$$

Regression 5 can no longer be causally interpreted because the treatment CS_{it} and the error term ε_{it} are correlated, as shown in equation 8.

Overcoming these econometric issues, which may even bias in different directions, is not a trivial task. In this section, I will discuss papers that predominantly ignore these issues. Thereafter, I will discuss empirical study whose settings can be expected to be more credible.

Bradbury and Stephenson (2003) investigate the "Law of 1/n" in counties in Georgia, United States (U. S.). Exploiting cross-sectional variation, they find a positive effect of the number of commissioners on public spending in the 4-5% range. The effect is stronger for expenditure on public health as well as public safety, and less for public infrastructure. They do not observe any differences between ward and at-large electoral districts though.

Gilligan and Matsusaka (1995, 2001) innovate upon the "Law of 1/n" literature strand by distinguishing between upper and lower legislative chambers. The former paper concerns government size in 48 U. S. states from 1960 to 1990. In a fixed effects (FE) regression, the authors note a positive association of council size on spending by about 10$ per capita in spending for a one-seat increase of the upper chamber. The effect of the lower chamber, however, is not statistically significantly distinguishable from zero. The effect is evident across spending categories (capital and non-capital on the one hand and welfare, education, and highway spending on the other). Gilligan and Matsusaka (2001) expand their bicameral model with a follow-up study in which they extend their

data set by analyzing data from the first half of the twentieth century, finding effects (and magnitudes) similar to their previous study. They add that the number of seats in the upper house (but not the lower one) is also linked to higher revenues. The authors are not able to give reasons for the null effect of the lower chamber.

In a similar vein, Bradbury and Crain (2001) leverage 1971-1989 panel data on the country level to compare the "Law of 1/n" in unicameral and bicameral systems. In spite of Gilligan and Matsusaka's (1995, 2001) reverse results, they find that the size of the lower house is positively related to public spending, whereas the sign of the upper house is varying, depending on specification. For the lower chamber, an increase in parliament size of one percent is expected to raise government budget by 0.24 to 0.35%. As bicameral systems require the two houses to negotiate over the budget, they are found to have a smaller "Law of 1/n" effect than unicameral systems.

Baqir (2002) takes the "Law" to a test in U. S. municipalities. Exploiting cross-sectional variation, he observes a rise of government spending per capita of approximately 1.6% for one additional councilor. The estimates do not differ for districting and at-large electoral systems. The paper adds to the literature the first test of reverse causality as laid out in the beginning of this section. First, he argues that due to the high cost of altering council size (avoidance of redistricting, time-consuming legislative process), it is hardly ever increased or decreased (pp. 1324-1328), limiting concerns of reverse causality. Second, employing an instrumental variable (IV) regression, he finds no evidence of reverse causality. He uses council size thirty years ago as the instrument. Contrary to the reverse-causality hypothesis, he finds IV coefficients to be larger than the OLS ones, hinting at an underestimation of the true causal effect in OLS regressions (pp. 1332-1338) and slightly alleviating concerns of simultaneity bias.

In order to reduce concerns of omitted-variable bias, MacDonald (2008) leverages the nature of her panel data set to account for municipality and time-specific effects. Once she does, the previously statistically significantly positive sign of council size disappears. The data stem from U. S. municipalities over the 1980-2000 time period. In order to obtain a valid estimate of the "Law of 1/n", it is thus important to keep omitted-variable bias in mind and adjust a pending study accordingly.

The evidence has so far dealt with settings in the U. S. but there is also evidence from other countries.

As for Italy, Fiorino and Ricciuti (2007) report a positive effect as well which is in line with standard theory. Employing a data set of Italian regions ranging from 1980 to 2000, they observe an increase of about 1.2% for a one-percent increase in legislature size in OLS and IV models.

For the Australian state of Victoria, Drew and Dollery (2017) investigate the "Law". Though they do not observe a statistically significant effect for council size, they do note a positive effect for the number of districts of a constituency. One additional ward can be expected to increase per capita spending by 3.4%. Furthermore, they find the ward-structure system to have 10% more public spending than municipalities whose councils are elected at-large.

Bel et al. (2018a) fail to find support for the "Law of 1/n" in Portuguese municipalities employing data from 2009-2013. With mayors being elected in a mixed system and a FE approach, they neither find an effect of council size nor of the number of directly elected lawmakers (Bel et al., 2018a, p. 45; Bel et al., 2018b, Appendix A).

The general idea of larger decision-making bodies leading to higher public spending extends beyond parliaments. Schaltegger and Feld (2009) show that public spending increases in the number of ministers in OLS regressions in Swiss cantons from 1980 to 1998. In FE regressions, the significance vanishes and the magnitude of the effect falls markedly. The sign on the coalition size, however, is ambiguous: It is significantly positive for the OLS methods but significantly negative in the FE method. Baskaran (2013) analyzes a similar question, albeit on the level of German states over a 35-year time frame. He finds the number of ministers to have a positive effect on government expenditure but coalition governments have none. The conclusions do not change when applying an IV approach with the number of parties in parliament as an instrument. Similar results are reported in earlier papers by Perotti and Kontopoulos (2002) and Roubini and Sachs (1989a, 1989b).

To conclude, "conventional" methods rarely find non-negative effect signs, with most of them being positive, particularly those using cross-sectional variation only. This holds even for some more advanced methods such as Baqir's (2002) IV regression.

3.1.2 Quasi-Experimental Methods

Next, I turn to more advanced empirical methods. In order for researchers to find the causal effect of the "Law of 1/n", they often primarily exploit cross-sectional variation in the data (Egger & Koethenbuerger, 2010, p. 201) as there are few changes in council size within one constituency. For instance, in my data sample described below, correlation of today's Bavarian councils with their historical counterparts 24 years ago is $r = 0.96$, as measured by the Pearson correlation coefficient r. This is why FE with respect to municipalities may be hardly applicable. Even if frequent changes in council size occur, this does not completely alleviate the endogeneity issue as there might still be reverse

causality or omitted-variable bias arising from factors not specific to municipalities or regions at play.

Egger and Koethenbuerger (2010) and Pettersson-Lidbom (2012) brought about a new idea by first reporting an RD examination of the "Law of 1/n". RDDs are a research design that requires a continuous running variable (also: forcing variable or score) with a cutoff that determines whether or not an observation is given treatment (Cattaneo et al., 2019, p. 5). Since in several countries council size is a function of population, this allows us to investigate the "Law" in an RDD. In this application, the running variable is the population of i in t, and treatment refers to (an increase in) council size. For simplicity, council size can be thought of as a deterministic function of population: Once a population cutoff is passed, council size increases. Councils are presumed to be unable to control whether they are slightly to the left or the right of the population cutoff. Given this random arrangement along the running variable in the immediate vicinity of the cutoff, cases on both sides of the threshold can reasonably be expected to be similar in terms of both observable and unobservable characteristics, with the sole exception of being assigned to the treatment group: Only those above the cutoff are part of the treated group, i. e. received a raise in council size. Hence, we have identified the causal effect in a quasi-experiment.

For the approach to be valid, we need the following assumptions to hold:

- Potential outcomes continuity: There is no jump in the potential-outcomes function.

- No sorting, i. e. no manipulation of the forcing variable: Observations do not control selection into treatment or control group.

- Thresholds are not confounded by other policy changes: Thresholds exclusively determine the treatment of interest.

Two extensions to the so-called sharp RDD are of relevance to this literature review. First, if the probability of treatment assignment at the cutoff does not jump discontinuously from 0% on the left to 100% on the right of the discontinuity (i. e. not deterministic but stochastic treatment assignment), we deal with a fuzzy RDD. For this, we still need the treatment assignment probability to jump at the cutoff. In this case, an IV approach with a dummy based on the forcing variable (either above or below the cutoff) as the instrument is applied. Second, this setting typically is a multiple-cutoff design as there are not one but several population thresholds inducing increases in council size. Then, the simplest approach is to normalize the forcing variable to zero

(see equation 9). RDDs, especially its extensions, will be explained in more depth in section 5.2.

Leveraging such RDDs, Pettersson-Lidbom (2012) found council size to decrease public expenditures at the threshold in study of Finnish and Swedish municipal councils, contradicting standard theory. In both countries, councils are elected at-large which might be a reason for this result. In the Finnish case, the author used the sharp RDD design whereas in the Swedish case he had to employ a fuzzy design.

For Italy, de Benedetto (2018) leverages a data set of Italian municipalities over the 2001-2007 period, using a sharp RDD. With total spending per capita as the dependent variable, he notes a decline of the local budget by 0.5% for an increase in one council seat.

Using Japanese municipalities over a six-year span in the early 2000s as a testing ground, Hirota and Yunoue (2012) do find the "Law of 1/n" in a sharp RDD. The elasticity of government size with respect to council size is roughly 1.8%, with the effect being particularly large for agricultural and civil engineering work spending. The additional budget is primarily financed by debt, whose elasticity is estimated to be 2.6% at the cutoff, which is larger than the spending elasticity.

A number of (working) papers investigates the "Law" in Brazil using RDDs. Kresch et al. (2020, pp. 12-13, 24) investigate, among other things, changes in public expenditure. Although spending on administration increases by around 30%, it is particularly social expenditure that increases at the thresholds (education, housing, social assistance) by roughly 25-60%. In addition, they observe an increase in taxes at the discontinuity, suggesting that additional spending is financed by increased taxation. Correa and Madeira (2014, pp. 22-23, 39) and de Britto and Fiorin (2016, pp. 11-12, 24–25), in contrast, neither find an effect on spending nor on revenue. Possible reasons include the low share that taxes make up in the budget and a cap on debt requiring more revenues than spending (de Britto & Fiorin, 2016, p. 11). de Britto and Fiorin (2016, p. 12) do find an effect for what they consider "potentially clientelistic" spending in that it can be targeted toward specific groups of people.

Finally, Lewis (2019) makes use of Indonesian laws, finding a negative effect of council size on spending at the discontinuities. The decrease in total spending amounts to 17% for five additional councilors. At the cutoff, no effect on own-resource revenues could be found.

Three papers exploit the German council size assignment rules: Egger and Koethen-buerger (2010), Höhmann (2017), and Holzmann and Zaddach (2019).

As the size of Bavarian councils is a deterministic function of population, Egger and Koethenbuerger (2010) run sharp RDDs to estimate differences on both sides of the cutoffs. Examining municipalities in the German state of Bavaria over a 21-year span, they are able to confirm the "Law of 1/n". The size of the effect at the threshold is in the 10-12% range, even though elections take place in an at-large system with strong mayors. Those factors have been argued to limit the pork-barrel problem but they do not in Bavarian municipalities. Additional expenditure at the cutoffs is financed by raised property taxes and is spent primarily for current expenditure which is more directly visible for voters.

Höhmann (2017) follows up on "Law's" effects by including eleven of 16 German states in his analysis but restricting the sample to the 2008-2010 time frame. Since the discontinuity is not deterministically determined in every state, he runs fuzzy RDDs. Contrary to conventional theory and Egger and Koethenbuerger's (2010) findings, he observes a reverse "Law of 1/n". One additional lawmaker causally decreases expenditures by 1%. Not only does the effect sign change, compared to Egger and Koethenbuerger (2010), but also the size of the effect shrinks drastically. Of note, the data set is limited to a time frame of public-finance turmoil due to the global financial crisis at that time.

Höhmann (2017) adds to the specific German case that cutoffs might be confounded, i. e. that a number of other policies are also altered at these particular thresholds, e. g. the mayor's salary. One of the RDD assumptions, in contrast, requires that nothing else be changed at the cutoffs as this may confound the results. Confounded thresholds can indeed be found in the German case. According to Höhmann (2017), for instance, in Bavaria only 60% of all cutoffs are not confounded. He re-runs the analysis excluding confounded cutoffs (64 of the 178 discontinuities excluded) but finds only slightly changed results. The issue of confounded thresholds had been completely neglected in the Egger and Koethenbuerger (2010) paper.

While the other two papers do find statistically significant results, Holzmann and Zaddach (2019) find null effects in the German state of Lower Saxony employing a multi-cutoff RDD. In Lower Saxony, voters elect councilors on an at-large basis and in a sharp design, making it a comparable case to Bavaria. More specifically, the paper is concerned with whether the RDD assumptions are fulfilled. Emphasizing the need to solely use non-confounded thresholds for identification, they only maintain ten out of 31 thresholds in their sample. Once they include all thresholds, however, they do note a positive "Law of 1/n" effect of the same order of magnitude as Egger and Koethenbuerger's (2010). This may indicate that the effects at the cutoffs found by

Egger and Koethenbuerger (2010) are primarily driven by other factors that change at these thresholds.

Summing up, the more advanced empirical methods challenge the "Law of 1/n" as they yield much more diverse results than the "conventional" methods. The latter consistently confirmed the Weingast et al. (1981) model, or at least did not provide evidence for the opposite. The conclusion of more diverse results is well reflected in the German case with three studies and three different effect signs.

3.2 Mechanisms

Apart from the unsettled question of the empirical basis of the "Law of 1/n", the literature also lacks systematic tests of the potential mechanisms behind it. Hence, proposed mechanisms in the literature are the topic of the subsequent section.

3.2.1 Differences between States

One reason for the "Law" effect to differ between German states (e. g. Egger & Koethenbuerger, 2010; Holzmann & Zaddach, 2019) might be that it actually differs, for instance owing to varying local political systems as determined by state law. Testing for differences between states also allows us to draw comparisons to the existing literature.

Hypothesis 3: *The "Law of 1/n" effect differs between states.*

3.2.2 Different Political Systems

Traditionally, the Weingast et al. (1981) model only relates to district systems. In a cross-section of U. S. cities, Dalenberg and Duffy-Deno (1991) note a one-percent raise in public capital stocks for district systems, compared to at-large ones. Southwick (1997) observes higher spending, taxes, as well as debt in ward municipalities. Drew and Dollery (2017) report spending elevated by approximately 10% in district systems. In the U. S. setting of his study, however, Baqir (2002) argues that the "Law" may apply to at-large electoral systems as well. Even though councilors in at-large systems can be voted for by the entire constituency, they too have "home bases" (Baqir, 2002, p. 1342), e. g. characterized by ethnic, cultural, social, or economic properties (as opposed to geographical electoral districts in ward systems). Indeed, he finds no differences between these two electoral systems. Farnham (1990) agrees as he does not find an effect of at-large elected councils.

Hypothesis 4: *The effect size of the "Law of 1/n" varies depending on the electoral system (at-large or mixed). The effect is smaller in at-large elected councils.*

Höhmann (2017) who finds a negative "Law of 1/n" cites the rather strong role of German mayors as a curbing factor in spending but does not run any empirical tests. For example, German mayors are directly elected and hold the power to block council decisions, possibly altering budget bills. With their strong positions, mayors are well equipped to internalize the "pork barrel" externality as they are accountable to the whole constituency. In the U. S., Baqir (2002, pp. 1347-1351) theorizes that systems with strong mayors (i. e. mayor-council systems, compared to mayor-manager ones) are better able to cope with the cost externality owing to their strong role. He reports that this is indeed the case empirically.

Hypothesis 5: *The stronger the mayor's position, the lower the effect of council size on spending.*

Furthermore, there is an ongoing debate on whether mayor characteristics in general (Brollo & Troiano, 2016; Chattopadhyay & Duflo, 2004) and party ideology (de Benedictis-Kessner & Warshaw, 2016; Dippel, 2022; Ferreira & Gyourko, 2009; Gamalerio, 2020; Gerber & Hopkins, 2011) in particular affect policy outcomes. Since some studies do find partisan effects, it makes sense to also investigate those in my setting.

Hypothesis 6: *The mayor's party affiliation affects the "Law of 1/n".*

A larger council might allow for a different composition of the legislation, which in turn alters public spending. A larger number of parties in parliament might be associated with higher negotiation cost when aiming to find a majority for council decisions such as the annual budget plan. Mukherjee (2003), for instance, hypothesizes that parliaments with more parties require more parties in (coalition) governments, raising government size. He leverages a country-level data set consisting of 110 nations and ranging from 1980 to 1996 to confirm his argumentation: Short of important thresholds (majority or supermajority), coalitions/parties need to attract further parties to form a government by spending more on the new partner's projects, creating a non-linear effect. Once the cutoffs are passed, public expenditure decreases.

Turning back to the local level, Kresch et al. (2020) do not only analyze the "pork barrel" but also diversity of the parties in the respective legislations. They find that a one-seat increase in council size leads to 1.3 more "effective parties" (Kresch et al., 2020, p. 9-10), i. e. larger political representation. Measuring the number of parties and their

competition, de Britto and Fiorin (2016, pp. 13, 27) use the Herfindahl (1950) and Hirschman (1964) index (HHI) but find no differences between municipalities left and right of the cutoff.

Hypothesis 7: *The size of the "Law of 1/n" effect is a function of the heterogeneity of the parties in parliament.*

A follow-up mechanism derived from Mukherjee's (2003) argumentation is to examine the effect of the 50% threshold. Single-party majorities, in theory, facilitate decision-making markedly. Negotiation cost, e. g. in budget consultations, decreases significantly.

Hypothesis 8: *An absolute majority by a single party curbs the "Law of 1/n".*

And similar to H6 for the council instead of the mayor:

Hypothesis 9: *The party that has an absolute majority in the council affects the "Law of 1/n".*

This is in a similar vein to the mayor's influence over the budgetary process: Fiscally more conservative parties might be more inclined to curb spending when council size increases, given they have the means, i. e. an absolute majority. The reverse might hold true for the fiscally more liberal parties.

3.2.3 Different Types of Expenditure

Multiple papers examine which types of public spending are primarily affected by the "Law of 1/n".

Pettersson-Lidbom (2012) found a reverse "Law of 1/n". As for potential mechanisms, he proposed a bureaucrat-council agency conflict in which bureaucrats maximize budgets and lawmakers, catering to their voters' demands, prefer smaller budgets, creating a conflict of interest. Councilors face time constraints, especially those at the local level that choose politics as their side job. Hence, they hand over some of their decision-making power to the administration and their officials, giving the latter substantive discretion over political decisions. Legislators then scramble to monitor the administrations actions, which again consumes plenty of time. In this situation, more lawmakers would allow them to increase monitoring quality and thus a raise in legislators would lead to less public spending, reducing the bureaucrat-council conflict (p. 270). While testing this hypothesis is not directly possible, Pettersson-Lidbom (2012, p. 270) reverts

to investigating expenditure categories that are assumed to be more directly under the lawmakers' control for a "Law of 1/n", that is public employment and operating expenditure. Indeed, Pettersson-Lidbom (2012, p. 277) reports a reverse "Law of 1/n" in these spending items of similar magnitude to the effect previously found in the main specification, in line with his theoretical prediction.

Egger and Koethenbuerger (2010, pp. 209-210) report results on three different spending categories: investment, material, and personnel. The latter two have a rather large size and significance at the thresholds (about a 14-17% raise), whereas the investment budget is only increased by 10%. They conclude that politicians favor more directly visible spending.

Kresch et al. (2020, pp. 12-13, 24), using data from Brazil, observe a focus on social spending categories. Spending on education, housing, and social assistance is markedly raised by 25-60%. Out of the four social expenditure types, only a single one of them (health) proves non-significant. As the cost categories that the authors consider administrative, neither legislative nor total labor or security expenditures are statistically significant, but expenditure on administration is.

As the scope of the de Britto and Fiorin (2016) paper concerns corruption in particular, the authors additionally check whether spending in what they consider clientelistic categories increases. While that is not the case individually, it is significant for a composite index. Their point estimate amounts to a 33% raise in clientelistic spending.

Owing to a lack of comparable data, I am not able to formulate hypotheses accordingly and test them empirically with the data set at hand. Notwithstanding that, I will check the expenditure categories at my disposal for suspicious patterns, primarily between administrative and capital expenditure.

Hypothesis 10: *The "Law's" effect varies by the expenditure categories at my disposal.*

3.2.4 Effects of Council Characteristics

As a general consideration, a larger number of councilors might increase diversity of representation, hence increasing the number of projects, which possibly results in higher spending overall.

Having more female representation might change the council's preferences since women tend to have different policy preferences. Regarding public expenditure, woman have been shown to favor spending on social issues (Abrams & Settle, 1999; Aidt & Dallal, 2008; Lott & Kenny, 1999). Similarly, an increase in female legislators induces higher

social public spending (Besley & Case, 2003, pp. 45-46; Svaleryd, 2009; Chattopadhyay & Duflo, 2004).

Correa and Madeira (2014) analyze the thresholds with respect to female representation, reporting that an increase of council seats at the threshold elevates the number of male candidates but does not change their vote totals. The female vote share as well as the number of female candidates is left unaltered. Since the competition solely for male candidates increases, this translates into more women in parliament and an increase in the likelihood of having at least one woman as a councilor. de Britto and Fiorin (2016) do not observe any changes in councilors' characteristics except for the the probability of having at least one female legislator (in line with what Correa & Madeira, 2014, found) and an increased mean age.

Hypothesis 11: *Larger female representation increases the "Law of 1/n" effect.*

3.2.5 Municipality Characteristics

Finally, it is fairly straightforward to investigate which municipalities primarily drive the effect. Hirota and Yunoue (2012) do so by looking at which municipalities in terms of population drive the "pork barrel" problem, and conclude that it is mainly small municipalities that face this problem.

Hypothesis 12: *The size of the municipalities by population determines the "Law of 1/n".*

Additionally, typical municipalities are fundamentally different from independent large cities ("*kreisfreie Städte*"). The latter, which do not belong to a county as a typical municipality does, assume some of the responsibilities of the counties. Hence, the "Law of 1/n" might differ between both entities. This is why Höhmann (2017, p. 352) excluded independent large cities from his analysis.

Hypothesis 13: *There is a difference between independent large cities, normal cities and municipalities when it comes to the "pork barrel" problem.*

4 Institutional Background

This section will present the background of the German municipal system to give the reader a basic understanding.

Germany is organized as a federal state with $10,796$ municipalities as of December 31, 2020 (Statistisches Bundesamt, 2021). They are the lowest level of the German administrative hierarchy, with counties and 16 states (“*Bundesländer*”) above them. There are different types of municipalities: Independent large cities, towns, and municipalities. Towns are generally larger than municipalities and both belong to counties while independent large cities do not. Municipalities’ tasks consist of self-government affairs on the one hand and delegated responsibilities on the other hand (Bogumil & Jann, 2020, pp. 121-122; Rudzio, 2015, p. 358). The former are voluntary responsibilities that are up to the respective council to decide upon, including theaters, museums, sports facilities, economic development, construction planning, building supervision, schools, and waste disposal. The latter consists of executing the matters commissioned to them by the upper levels of the political hierarchy, that is state and federal laws: Execution of elections, maintenance of the fire department, social welfare, natural conservation, and epidemic protection (Rudzio, 2015, p. 358).

As for public finance, revenue from taxes makes up 40% of total revenue (Rudzio, 2015, p. 356). The local business tax accounts for 14% of total revenue, property tax for an additional 5%. Most of the municipalities’ revenue comes from a share of the income tax that is levied by the federal government and distributed to the municipalities. Accounting for 17% of total revenue, municipalities receive grants from the respective state and the federal government (Deutscher Städtetag, 2021, p. 10). On the spending side, personnel, material expenditure, social welfare payments account for close to 25% each. (Deutscher Städtetag, 2021, p. 10).

The exact design of local constitutions is subject to state law and hence differs between states, not so much between municipalities within states. They all have in common that legislation takes place in local councils elected by the citizens (Rudzio, 2015, p. 369). Of the 16 German states, three are so-called city states (“*Stadtstaaten*”) whose local structure differs substantially from the others. All states have laws in place that govern local constitutions. For instance, council elections take place every five or six years, depending on state regulations. Council size is a function of population size: Larger municipalities have larger councils. In several states, the respective council is free to differ from the designated number of councilors by changing local statutes. In addition, some states allow for “overhang seats” (“*Überhangmandate*”), i. e. more seats than designated in case one party wins more wards than seats it would have in parliaments based on total votes, increasing the number of legislators. Hence, council size *stochastically* depends on population. In Rhineland-Palatinate, Bavaria, Saarland, Saxony-Anhalt, and Thuringia, council size is deterministic. Details on thresholds and local constitutions are provided in tables 9, 10 and 11 in the Appendix.

As for H4, in all but two states elections take place at-large. Only in North Rhine-Westphalia and Schleswig-Holstein, half of the (nominal) council size is elected in wards, with the rest coming from closed party lists (Wahlrecht.de, n. d.), creating a mixed electoral system. No state allows municipalities to hold pure ward elections.

Regarding the precise definition of the mayor's role (H5 and H6), German local constitutions, determined by state regulations, have converged within the last 30 years into constitutions that all foresee a strong position of the mayors who are directly elected and have a privileged position towards both the council and the administration (Ipsen, 2007, p. 654; Naßmacher & Naßmacher, 2007, pp. 201-208; Rudzio, 2015, pp. 370-374). However, there are still some degrees of variation between states. The "Southern German mayoral constitution" ("*Süddeutsche Bürgermeisterverfassung*"), in which the mayor acts as both the head of administration and council (Ehlers, 2007, p. 524), and holds the right to veto council decisions, is the most widespread system. The opposite is the Hessian "magistrate constitution" ("*Magistratsverfassung*") in which the mayor is not the head of the administration. In the following, Schleswig-Holstein, North Rhine-Westphalia, Rhineland-Palatinate, Baden-Württemberg, Bavaria, Saarland, and Saxony are thought of as having a "Southern German mayoral constitution" (Schubert & Klein, 2018), even though this classification is rather blurred than clear-cut.

With mayors commonly having a strong role and at-large council elections, Germany appears well equipped to curb the "Law of 1/n". Both factors have been argued to lower the effect (see section 3.2).

The effect of party affiliation in H6 and H9 can be solely tested with respect to the center-right/conservative *Christlich Demokratische Union (CDU)* and its counterpart *Christlich-Soziale Union (CSU)* – the latter only active in Bavaria – and center-left *Sozialdemokratische Partei Deutschlands (SPD)* mayors. The former are assumed to be more fiscally conservative, whereas the latter might be more fiscally liberal. Local parties often offer joint lists as well as uncommon coalitions and are generally thought of as less ideological than their state and federal-level counterparts. Voter groups, represented in many local councils, are local party-like associations that do not fall under the definition of federal legislation of what constitutes a party. Typically, they claim to base their decisions on pragmatic aspects rather than ideological ones.

As for H8, it would have been interesting to investigate different council compositions by analyzing different majorities or parties in parliament by ideology. However, German local elections are multi-party systems and its electoral system favors these, e. g. through proportional representation. This system makes it hard for researchers to quantify the power of each political bloc. Hence, I leveraged absolute majorities only as these facilitate the interpretation of the majority's party ideology.

5 Methodology

The following section deals with the data used in the empirical analysis as well as the methodology.

5.1 Data

Apart from the three city states, Lower Saxony and Brandenburg are removed from the analysis owing to missing data. For Mecklenburg-Vorpommern and Saxony-Anhalt I only have data for some years for independent large cities.

An RD framework requires data only on three dimensions: the forcing variable, the treatment indicator, and the outcome variable.

For the (fiscal) outcome variable(s) and several covariates, I was provided a data set by the Leibniz Centre for European Economic Research (ZEW) running from 2002 to 2019. Asatryan et al. (2017) were first to use it. It has been updated internally even after publication. Their data is available via two different sources: Unemployment data is downloaded from the Federal Employment Agency ("*Bundesagentur für Arbeit*"), the remainder of the variables originate from the "*Regionaldatenbank*" database of the Federal Statistical Office (Destatis). The main dependent variable, the natural log of gross expenditure per capita (p. c.), is in 2005 Euros. I code all fiscal variables in this fashion.

For the treatment variable, I collected election data including the total number of (actual) council seats from the 13 statistical offices of the states, along with further information (mayoral party affiliation, council-party seats, and more, if available) that come in handy for the mechanism tests.

Though the ZEW data set contains population data as of December 31 of every year, these data typically have not been used to determine council size. Instead, all states mandate different cutoff dates, with one state (Lower Saxony) even having a deviating definition of what constitutes an inhabitant. Since these data are not necessarily centrally collected (and in some cases, not even freely and publicly available), they have been compiled from the respective state-statistical offices. In spite of statements to the contrary (Höhmann, 2017, p. 353, footnote 4), Höhmann declined to hand out data sets as those were restricted by confidentiality agreements with the statistical agencies. Höhmann (2017) told me in personal conversation that he received his population data used to determine council size directly from the state-statistical offices. However, when I contacted the statistical offices, only one was able to provide a single data set

containing the exact population date cutoffs and corresponding council size, with some of them not being able to hand out exact population data at all. Hence, I conducted extensive legal research to identify the exact population cutoff dates. This research included policy changes at the same population thresholds as council size changes (see table 9 and 10). In cases in which the relevant authorities were unable to detail population cutoff dates, I also wrote to a sample of municipalities seeking clearance on this matter which turned out rather unsuccessful. Details on state provisions are shown in table 11 in the Appendix. The difference in dates between the actual election day and the deadline for the population data has a mean of 455 days (median: 363 days, minimum: 237, maximum: 876).

To apply the pooling-and-normalization approach for the running variable, I followed Egger and Koethenbuerger (2010, p. 205) in applying the following formula to analyze the set of cutoffs as a single one:

$$pop_{it}^{norm} = ln\left(pop_{it}/pop_0\right) \tag{9}$$

with pop_{it}^{norm} as the normalized population size (and forcing variable) and pop_0 as the closest cutoff point so that $pop_{it}^{norm} = 0$ at the thresholds.

The data run from 2002 to 2014, when the Destatis database discontinued publishing municipal fiscal data. In my data sample, one observation is a municipality-year. One election produces a council size that is constant for the entire legislative period until the next election. Election re-runs (e. g. due to illegal election manipulation in the first run) and special/extraordinary elections (e. g. because of municipal mergers) have been ignored, if possible, as they are not typically part of the election data sets of the state statistical agencies. In addition, in the latter case sometimes extra rules are put in place to adequately take into account the interests of all merged municipalities.

Descriptive statistics are shown in Table 1. Details on the code used for this analysis is available in the short Documentation in the Appendix.

5.2 Methods

In order to compare results, I will start off the results section 7 by running standard OLS and FE models while also briefly discussing potential IV models. The main method, however, is the RDD.

Table 1: Descriptive Univariate Statistics

Variable	Min.	1st Quar.	Median	3rd Quar.	Max.	Mean	SD	n
RDD without controls								
Ln of gross expenditure p. c.	0.14	6.90	7.19	7.45	13.94	7.19	0.42	103,618
Council size	0	9	14	19	94	15.58	9.02	103,618
Above cutoff (dummy, "instrument")	0	0	1	1	1	0.53	0.50	103,618
Running variable (inhabitants relative to thresholds)	-3.76	-0.15	0.02	0.19	0.86	-0.02	0.36	103,618
Year	2002	2005	2007	2011	2014	2007.66	3.63	103,618
Council size population thresholds	70	500	2,000	5,000	700,000	6,019.66	21,825.06	103,618
RDD with controls								
Ln of gross expenditure p. c.	0.14	6.91	7.20	7.45	13.80	7.19	0.41	101,409
Council size	0	10	14	19	94	15.78	8.99	101,409
Above cutoff (dummy, "instrument")	0	0	1	1	1	0.54	0.50	101,409
Running variable (inhabitants relative to thresholds)	-3.31	-0.14	0.03	0.19	0.86	-0.01	0.32	101,409
Year	2002	2005	2008	2011	2014	2007.69	3.63	101,409
Council size population thresholds	70	750	2,000	5,000	700,000	6,130.97	22,022.54	101,409
Population (as of 12-31)	8	683	1,929	5,155	995,420	6,403.66	24,387.28	101,409
Share of working-age population	0.34	0.60	0.61	0.63	0.82	0.61	0.03	101,409
Share of population aged > 65 years	0.02	0.17	0.19	0.21	0.47	0.19	0.04	101,409
Unemployment rate	0.00	0.03	0.04	0.06	0.30	0.05	0.03	101,409
Total area in ha	39	712	1,580	3,372	40,516	2,566.77	2,851.21	101,409
Independent large city (dummy)	0	0	0	0	1	0.01	0.10	101,409
Town (dummy)	0	0	0	0	1	0.16	0.37	101,409
Years since last election	0	1	2	3	5	2.13	1.56	101,409
Potentially heterogeneity-inducing variables/Potential mechanisms								
Ln of gross revenue p. c.	-1.29	6.89	7.19	7.45	13.91	7.18	0.42	103,610
Ln of net expenditure p. c.	-0.24	6.49	6.85	7.16	13.94	6.81	0.55	103,555
Ward elections (dummy)	0	0	0	0	1	0.14	0.34	103,618
Southern-German type local constitution (dummy)	0	1	1	1	1	0.87	0.34	103,618
Magistrate local constitution (dummy)	0	0	0	0	1	0.05	0.22	103,618
CDU/CSU mayor (dummy, BY)	0	0	0	1	1	0.44	0.50	26,702
SPD mayor (dummy, BY)	0	0	0	0	1	0.14	0.34	26,702
Number of parties in council (NW, BW, BY, SL, MV, SN)	1	2	3	4	9	2.84	1.27	48,489
Number of parties in council by HHI (NW, BW, BY, SL, MV, SN)	0.18	0.35	0.43	0.57	1.00	0.52	0.24	48,489
Absolute majority (dummy, NW, BW, BY, SL, MV, SN)	0	0	1	1	1	0.51	0.50	48,489
Absolute CDU/CSU majority (dummy, NW, BW, BY, SL, MV, SN)	0	0	0	0	1	0.12	0.33	48,489
Absolute SPD majority (dummy, NW, BW, BY, SL, MV, SN)	0	0	0	0	1	0.00	0.06	48,489
Ln of administrative expenditure p. c.	-0.23	6.61	6.85	7.08	13.80	6.86	0.36	103,616
Ln of personnel expenditure p. c.	-3.34	4.32	5.45	5.83	9.67	5.11	0.95	103,589
Ln of operating expenditure p. c.	-2.22	4.85	5.26	5.60	9.81	5.22	0.57	103,602
Ln of capital expenditure p. c.	-5.21	5.10	5.77	6.32	13.26	5.62	1.09	102,754
Ln of loan expenditure p. c.	-5.30	2.92	3.65	4.34	9.70	3.58	1.22	86,517
Ln of property investment expenditure p. c.	-5.14	4.48	5.30	5.90	10.85	5.04	1.31	100,061
Female share in council (SH, BW, BY, ST, TH)	0.00	0.11	0.17	0.25	0.67	0.18	0.10	56,964
Female share in largest council party (BW, BY)	0.00	0.10	0.17	0.25	0.71	0.17	0.12	41,007

5.2.1 RDD

Usage of RDDs in economics[2] has steadily increased over recent years (Cunningham, 2021, chapter 6.1.1, figure 6.1). They are now also widely used in empirical political economy.[3] After the brief introduction in section 3.1.2, RDDs will be explored in more depth in this chapter.

A sharp global-parametric RDD regression equation with a binary treatment indicator without covariates X_{it}' may be:

$$
\begin{aligned}
Y_{it} =& \alpha + \beta \times \mathbb{1}(pop_{it}^{norm} \geq pop_0) + \gamma \times f(pop_{it}^{norm}) \\
& + \delta \times f(pop_{it}^{norm}) \times \mathbb{1}(pop_{it}^{norm} \geq pop_0) + \varepsilon_{it}
\end{aligned}
\tag{10}
$$

with pop_0 as the threshold value (here: $pop_0 = 0$), $\mathbb{1}(pop_{it}^{norm} \geq pop_0)$ as the RDD dummy, and $f(pop_{it}^{norm})$ as a (potentially non-linear) function of the forcing variable, along with an interaction with the $\mathbb{1}(pop_{it}^{norm} \geq pop_0)$ dummy. β would then be considered the RD treatment effect.

In practice, RDDs are not estimated using a parametric form over the entire data set (Cattaneo & Vazquez-Bare, 2017, pp. 137-138). Rather, only observations in close proximity to the threshold are retained in the data set (Gelman & Imbens, 2019, p. 456) and local regressions whose polynomials may vary between the two sides of the cutoff are employed (Cattaneo et al., 2019, p. 33). First proposed by Imbens and Kalyanaraman (2012), bandwidths are nowadays selected in a data-driven manner instead of an ad-hoc fashion commonly employed earlier (Cattaneo & Vazquez-Bare, 2017, pp. 137-138). Essentially, more precision through larger sample size (and lower polynomial degrees) is traded off against less effort required to get the degree of polynomials right (but a smaller sample size), reducing the potential for misspecification bias.

If compliance with the treatment assignment rule ($CS_{it} \geq CS_0$) is not perfect, we speak of a fuzzy RDD. In this case, not everyone who is assigned to the treatment group ($CS_{it} \geq CS_0$) actually receives treatment, and vice versa for the control group. Still, the treatment probability jumps significantly at the threshold, allowing us to employ an RDD. This setting is a rather rare application of the fuzzy RDD as the treatment

---

[2]The work of Angrist and Lavy (1999), Black (1999), Cook (2008), Hahn et al. (2001), Imbens and Lemieux (2008), Lee (2008), Lee and Lemieux (2010), and Thistlethwaite and Campbell (1960) has been seminal for the RD method.

[3]See e. g. Bagues and Campa (2021), Brollo et al. (2013), Casas-Arce and Saiz (2015), Eggers et al. (2015), Ferreira and Gyourko (2009), Hopkins (2011), Lee (2008), and Pettersson-Lidbom (2008, 2012).

variable, council size, is not binary but continuous. This is different from Höhmann (2017, pp. 353-354), who apparently found a way to form a binary treatment indicator.[4]

Another extension to the RD framework is the multiple-cutoff setting. The normalization and pooling of the running variable as discussed in section 5.1 (equation 9) yields an estimator that is simply the weighted average of the respective RD effects at all cutoffs (Cattaneo, Idrobo, et al., 2018, p. 92; Cattaneo et al., 2016).

RDDs also come with a downside: While their internal validity can be fairly high because omitted-variable and simultaneity bias are avoided, their external validity is rather low (Imbens & Lemieux, 2008, pp. 621-622). In an RDD setting, one can only draw conclusions from the changes *at the cutoff* (i. e. locally) but not about general trends. Once we additionally assume heterogeneous treatment effects in the fuzzy setting, this issue aggravates as the effect is now solely identified for compliers (Imbens & Angrist, 1994), i. e. units that switch into treatment due to the instrument. In that sense, it is "double-local".

From a technical point of view, I use the R software (R Core Team, 2022; RStudio Team, 2022) for analysis, along with libraries from the Calonico et al. (2022) RDD package universe, mainly the rdrobust package (Calonico et al., 2014b).[5] I employ triangular kernel weights and the bandwidth is chosen based on a data-driven mean squared error (MSE) optimal bandwidth selection algorithm (Cattaneo et al., 2019, pp. 39-44). The bandwidth is the largest value of the running variable from the cutoff to be included in the estimation. In this case, the bandwidth is similar on both sides of the thresholds. The point estimator is estimated using a local linear polynomial ($p = 1$), the bias correction is calculated using a quadratic local polynomial ($q = 2$). All of these arguments are the rdrobust default options. The standard MSE-optimal method used to construct the point estimator is not statistically valid for hypothesis testing, as it overrejects the null hypothesis. This makes the bias correction necessary for statistical inference (Cattaneo et al., 2019, pp. 57-62). Since I use *standard* coefficients along with *robust bias-corrected* confidence intervals that have good statistical coverage properties (Cattaneo et al., 2019, pp. 62-67) the point estimate is not necessarily centered within the confidence interval. In fact, in some more extreme cases they do not even overlap. Standard errors are clustered at the threshold level. Otherwise, on the first stage an

---

[4]He does not say so explicitly. He interprets the treatment effect of -0.03, along with an average increase of council size of 3.1 seats, as a one-percent increase in spending for each additional councilor. If he had a continuous treatment indicator, the interpretation would have been different. Then the treatment effect would have already been divided by the first-stage jump (of about three councilors) instead of the treatment-administration probability jump in the binary treatment-indicator case, given the definition of the Wald estimator.

[5]As for statistical details, refer to Calonico et al. (2018, 2020), Calonico et al. (2019). Regarding the technical implementation, see Calonico et al. (2017) and Calonico et al. (2014a, 2015), Cattaneo, Jansson, et al. (2018).

observation close to a 500 cutoff would be considered identical to a unit at the $500,000$ cutoff, if both have the same value of the pooled forcing variable. For computational reasons, I refrain from using municipality FE in the non-parametric regressions and employ state FE instead, along with year and threshold FE.

For OLS, FE, and IV models, I leverage the R `fixest` package (Bergé, 2018, 2020). Control variables include the total number of inhabitants, the share of the population aged 65 and older, the unemployment rate, total area, the share of working-age population as well as dummies for towns, independent large cities, and year distance to the last election. Though adding covariates to the estimated model is not necessarily a requirement if the RD assumptions hold, it generally allows for more precision in the estimates (Calonico et al., 2019, p. 443). The inclusion of covariates, however, is problematic if the controls are affected by treatment administration (Calonico et al., 2019, p. 443) which is possible given the panel nature of my data set. Hence, I show both the RD regressions with and without controls, with the former serving as the baseline model. This issue does not only affect the RDD but also the "conventional" regressions for which this is known as "bad controls", potentially biasing the estimates even beyond what was discussed in section 3.1.1.

For the interpretation of both the "conventional" and the RD results I will use the log approximation of the log-lin model. Hence, the point estimates and confidence intervals in figures are the coefficients from the regression output. They are not exponentiated.

5.2.2 The RDD's Validity

While population thresholds have been frequently used in causal assessments[6], the identifying assumptions in population-cutoff RDDs have repeatedly been questioned on the grounds of potential sorting and confounding.

First, sorting has been shown to occur in Brazil (Litschig, 2012) and three European countries, among them Germany (Eggers et al., 2018a). Sorting violates the continuity assumption. Theoretically, population-number calculations are carried out independently by statistical agencies, avoiding any conflict from potentially interested parties such as municipal representatives trying to "push" a municipality over a threshold. If entities were free to choose if they like to be above or below a discontinuity in an RDD, we would have to expect the entities on both sides of the cutoff to be quite different on

---

[6]See RDD discussion of the effects of council size above as well as Arnold and Freier (2015), Barone and de Blasio (2013), Brollo et al. (2013), Eggers (2015), Gagliarducci and Nannini (2013), Harvey (2020), Litschig (2012), Litschig and Morrison (2013), Sanz (2020), and Tyrefors Hinnerich and Pettersson-Lidbom (2014). Closely related, other running variables include registered voters (Fujiwara, 2011) and the number of non-native speakers (Hopkins, 2011).

a number of covariates, which determined their decision in the first place. Remedy is comparably simple: Plot the density of the running variable surrounding the cutoff, in its formal form also know as the McCrary (2008) test.

Furthermore, Eggers et al. (2018a) found evidence for confounded thresholds. If a threshold is used to determine multiple policies, the causal effect of council size (in my case) is no longer identified since it could also be the other policy that causes Y_{it} to change. Eggers et al. (2018a, pp. 212-214) demonstrate that for the three European countries in question, the same cutoffs are used to determine multiple policies. In France and Italy, every cutoff determining council size is confounded in that at each threshold, other policies change, too. The authors also provide extensive research on population-based thresholds in Germany (Eggers et al., 2018b, pp. 15-16, table 6), concluding that out of 759 policy-changing population cutoffs, only 94 are not confounded (Eggers et al., 2018a, pp. 212). Detecting confounded thresholds thus requires detailed knowledge of the political and legal system. As a remedy, they propose a difference-in-discontinuities (diff-in-disc) design (Grembi et al., 2016): At one threshold, multiple policies change. Now, the policy of interest is introduced or abolished at that particular cutoff, allowing us to compare effect sizes before and after the policy change and, under certain conditions such as local parallel trends, giving us a causal estimate (Grembi et al., 2016, pp. 2-3, 8–12). The diff-in-disc method is beyond the scope of this paper as it would require changes in council-size thresholds, of which there are few. These changes took place toward the end of my data set in municipalities not contained in my data set.[7] In addition, the routine is neither implemented in widely used software packages in economics such as `Stata`[8] or `R` nor is it well applicable to my case owing to the wide range (across 11 states with 11 different legal settings) of the data set. Alternatively, one can use interaction terms in parametric regressions or try to re-run regressions on a subset of non-confounded thresholds, similar to Höhmann (2017, pp. 355-356), which, however, does not allow for inference on whether the effect of confounded threshold drives the effect, and merely gives an indication.

To detect confounded thresholds, I extended the legal research to other policy changes at cutoffs at which council size is changed, in the spirit of Höhmann (2017, pp. 355-356). Tracking down policy changes that existed in the early years of the data set but were eliminated over the course of the panel are hard to find, if possible at all. Hence, some

---

[7]Three states changed thresholds: Schleswig-Holstein in 2012 ($2,000$ to $2,500$), with no election after that in my data set; Brandenburg in 2007 (several changes to municipalities with less than 700 inhabitants), with no data at all in my data set; Saxony-Anhalt in 2014 (several changes to municipalities with less than $1,000$ inhabitants), with no data on small municipalities in my data set.

[8]Giambona and Ribas (2018) and Ribas (2016, n. d.) have created a `Stata` command which, however, is in its beta version and not yet available on Boston College Statistical Software Components (SSC).

of the thresholds indicated as non-confounded might in fact be confounded. Research results are shown in tables 9 and 10 of the Appendix.

There are several differences in the classification of confounded thresholds between Höhmann's (2017, pp. 360-364, table 5) and my categorization. In total, there are 28 thresholds (24%) classified differently that, when weighted by municipalities close to these thresholds, account for 56% of all observations. Three states stand out with a more than 50% share of observations classified differently by Höhmann and me: Rhineland-Palatinate with 98%, Bavaria with 69%, and Thuringia with 60%. Four states are exactly identically classified (Baden-Württemberg, Saarland, Mecklenburg-Vorpommern, Saxony-Anhalt). In total, I find 81 confounded cutoffs and only 25 non-confounded ones, and Höhmann finds 60 confounded and 47 non-confounded ones. Hence, my classification is more "conservative" in that I look for more changes. Some of those may not matter much for the outcome variable but do not allow for the clean identification of the treatment effect, such as an increase in the required number of signatories mayoral candidates need to enter the contest or the status of a municipality's officials. However, Höhmann also had more mistakes: In Bavaria, he left out citizen petitions, citizen referenda, and some mayoral wage increases, which are included for other states. This mistake leads to all but one (instead of 6) non-confounded thresholds. In Saxony-Anhalt (5,000 threshold) and Thuringia (500, 1,000, and 2,000 threshold), he lists policy changes that apparently refer to the population of municipal districts, i. e. subordinations of the municipalities. For Hesse he lists several confounding policies that were not introduced until after his sample ended (full-time mayor at 5,000 since 2016, quota requirements for citizen petitions and referenda at 50,000 since 2011 and 2016, respectively). Similar patterns are visible for Thuringia and Schleswig-Holstein. Perhaps he looked into confounded policies significantly after his sample ended but ignored that the confounding policies might have been included (or changed or abolished) in the meantime. When asked for comment, Höhmann replied that he concentrated on "institutional changes [. . .] that may have significant impact on the municipalities' and towns' budgets" (citation from private conversation). He acknowledged that one could be more precise when looking for other confounded thresholds.

These arguments give more than sufficient reason to test the identifying assumptions of the RDD in my data set. Whether they are met will be examined in depth below.

5.2.3 The Mechanisms

To evaluate heterogeneity in the RD framework, two standard approaches are commonly used, both of which I utilize as well: The split-sample approach and the interaction-term method. For the former, the sample is divided by the values of the potentially heterogeneity-inducing variable. Then, I run standard non-parametric local regressions. For the interaction-term method, I use the parametric IV RDD approach coupled with an interaction term of the treatment effect and the heterogeneity variable. The interaction-term method is known to over-reject if models are misspecified, which is more likely in the parametric case than in the non-parametric one (Gelman & Imbens, 2019, p. 456; Hsu & Shen, 2019, p. 469). However, unlike the non-parametric procedure, it does allow for statistical inference on *differences* between two subgroups. The subsample method, in contrast, is non-parametric but aggravates the multiple-testing issue as we run more tests than in the interaction-term method. Furthermore, if the heterogeneity-inducing variable is continuous, this forces the researcher to split the variable into arbitrarily discretized groups.

A sharp regression equation with a binary treatment indicator testing for treatment-effect heterogeneity in a continuous covariate X_{it} may look like this:

$$
\begin{aligned}
Y_{it} = {} & \alpha + \beta \times f(pop_{it}^{norm}) + \gamma \times \mathbb{1}(pop_{it}^{norm} \geq pop_0) + \delta \times X_{it} \\
& + \zeta \times f(pop_{it}^{norm}) \times \mathbb{1}(pop_{it}^{norm} \geq pop_0) + \eta \times f(pop_{it}^{norm}) \times X_{it} \\
& + \theta \times \mathbb{1}(pop_{it}^{norm} \geq pop_0) \times X_{it} \\
& + \iota \times f(pop_{it}^{norm}) \times \mathbb{1}(pop_{it}^{norm} \geq pop_0) \times X_{it} + \varepsilon_{it}
\end{aligned}
\tag{11}
$$

with γ as the baseline treatment effect and θ as the treatment effect at the cutoff conditional on X_{it}.

The standard approaches, however, lead to a couple of statistical problems. When testing multiple null hypotheses, all of which are true in reality, and rejecting each if the empirical test statistic yields $p \leq 0.05$, we would be expected to mistakenly reject one out of twenty (i. e. 5%) hypotheses, the so-called type I error, by the very design of hypothesis testing. While there exist methods to adjust for multiple testing (e. g. Anderson, 2008), this section is aimed at giving preliminary evidence. As a small remedy, I lower the significance level α necessary to reject the null hypothesis to $\alpha = 0.01$. This does not directly address the issue but makes the conclusions more careful. However, it leads to the results laid out below to merely give an indication of

what *might* be driving the effects. They should not be considered causal (or statistically valid, for that matter) in any way.

While many authors have put forward ideas on detecting RD treatment heterogeneity (e. g. Becker et al., 2013; Bertanha, 2020; Hsu & Shen, 2019; Reguly, 2021a), none of them have been packaged into software libraries. The only test to have publicly accessible code (Hsu & Shen, 2018, 2019) in R or Stata does not allow for fixed effects to the best of my knowledge. Without fixed-effects, estimating the first stage is a questionable endeavor as it is impossible to distinguish between observations that are to the left of the cutoff but at a higher population threshold and observations to the right but on lower population threshold, given that all cutoffs have been normalized to zero in this study. Furthermore, the Hsu and Shen (2019) test does not allow for testing more than one heterogeneity covariate at once.

Reguly (2021a) developed a causal machine learning algorithm to detect treatment-effect heterogeneity in the RD setting without invalidating statistical inference by multiple testing and without the researcher being required to select potential heterogeneity variables. His code (Reguly, 2021b), however, is only available for the commercial statistics software Matlab.

Apart from these inference problems, the coefficients may also be biased. If a variable that induces heterogeneity is in fact a result of the treatment, econometricians fear "bad controls". However, due to the RD research design, this should not pose a problem but it would in other designs, e. g. fixed effects regressions.

6 Validity Checks

In order for the RDD to be valid, several assumptions need to hold. Since the RDD assumptions are rather well testable, compared for instance to an IV framework, I will attempt to test them as thoroughly as possible to convince the reader of the (in)validity of the setup. In this section, I show results from discontinuity and balance tests. I additionally test whether confounded thresholds drive the effect in the RDD Robustness Checks section as this is methodologically more similar to the RDD results.

6.1 Discontinuities in the Running Variable

First of all, there is a general incentive and potential for municipalities to sort. Municipalities thrive for larger councils as they associate it with higher prestige (discussed in

the subsequent paragraph). As the number of inhabitants is recorded in a local resident registration, the administrations are likely to have a good idea on where they are with respect to the cutoff, giving raise to potential issues with sorting. One test is to check for formal or institutionalized mechanisms for units (municipalities) to appeal the value of the running variable (Cattaneo et al., 2019, pp. 77-79) and hence, assignment into treatment or control group. However, to the best of my knowledge, no such ways exist as the determination of the population data is carried out independently by the statistical offices.

In conversation with state ministries, one issue casts doubt on the RDD's validity. In multiple (if not all) states (for North Rhine-Westphalia or Thuringia, this has been confirmed by the statistical agencies or ministries), municipalities autonomously interpret the legal documents that codify council size and it is not the ministries' task to exercise legal supervision of the municipalities' decisions. The North Rhine-Westphalia Ministry of the Interior, in personal conversation, pointed out that since (particularly large independent) cities would like to spare themselves the "embarrassment" of loosing council seats when local population declines, municipalities in North Rhine-Westphalia at times use monthly population data – despite state law mandating that municipalities use biannual data – or even population projections which are not even stipulated in law. If this was indeed the case, I would expect municipalities to sort on the right side of the cutoff, again motivating a thorough investigation of potential RD validity pitfalls. This scenario is consistent with Eggers et al.'s (2018a, p. 219, table 2) findings of sorting in German municipalities to the right of thresholds that is especially large at higher cutoffs.

To check for evidence on sorting, the McCrary (2008) density test has for years been the go-to test. It tests for a jump in the density of the running variable at the cutoff, with the null hypothesis of no discontinuity. Of note, maintaining the null is no evidence of no sorting. If sorting occurs on both sides, for example some municipalities want larger councils for prestige while others try to keep their council size small (e. g. for cost reasons), this would not necessarily result in evidence for sorting in the McCrary test (or similar other tests). For my data set, the density test does not reject the null of a continuous running variable at conventional significance levels ($p = 0.21$)[9], with visual support in figure 1.

---

[9]Test details: Bin size: 0.004, bandwidth: 0.243, log jump at cutoff: 0.038 with SE $= 0.030$. Testing the variables individually, no year reaches significant jumps, while Hesse and Thuringia have a positive jump that is significant at the $\alpha = 10\%$ level, which probably is to be expected for multiple testing. Of the 21 population cutoffs for which the test could be run individually, two proved to be (positively) significant at the $\alpha = 10\%$ level or less.

Figure 1: McCrary (2008) Density Plot

Running variable (population relative to thresholds)
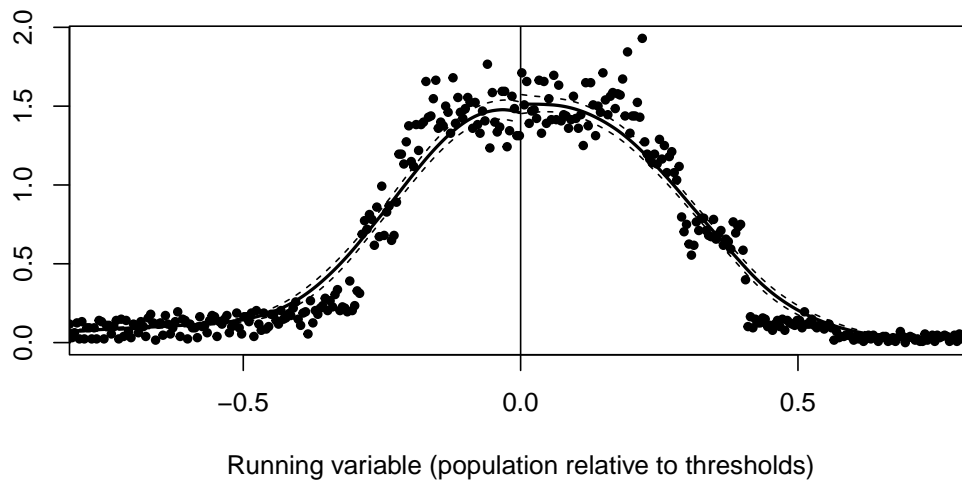
Cattaneo, Jansson, et al. (2018, 2020) develop and implement a density estimator that, unlike the McCrary (2008) one, does not require transformations of the data such as pre-binning. In addition, Eggers et al. (2018a, pp. 218-220) show that the McCrary (2008) test is biased when using percentage deviations of discrete running variables from thresholds, motivating the use of the Cattaneo, Jansson, et al. (2018, 2020) density estimator. The result of this test is once again a positive jump that is not statistically distinguishable from zero ($p = 0.34$).[10] We cannot assume the alternative hypothesis of sorting. The test is shown visually in figure 2.

To conclude, both methods provide no evidence of sorting whatsoever, consistent with Höhmann (2017, pp. 355-357) but contesting what Eggers et al. (2018a, pp. 217-219) find. One potential reason may be that Eggers et al. (2018a) most likely use annual data, while Höhmann (2017) and I employ the exact population data used to determine council size, which deviates at times as laid out in section 5.1.

6.2 Balance Tests

In real experiments (as opposed to quasi-experiments) researchers typically run balance tests to convince their reader that the randomization strategy worked. A similar approach is feasible in RDDs: One can run tests to detect whether units to the left and the right of the cutoff are similar with respect to observable characteristics. If so, it may be

---

[10]Tested individually, Baden-Württemberg and Saxony-Anhalt, two (2003 & 2004) out of eight election years, and one single cutoff show signs of sorting at the 10% significance level, once to the right and once to the left of the cutoffs, respectively.

Figure 2: Cattaneo, Jansson, et al. (2018, 2020) Density Plot
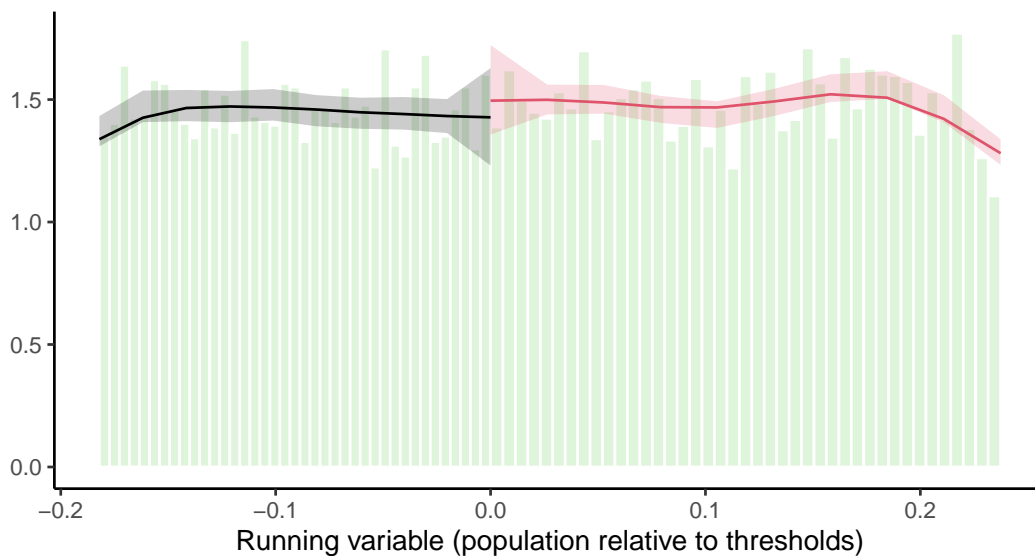
plausible that units on both side have similar unobservable properties, implying that what we have is indeed a randomized (quasi-)experiment.

To do so, I run the main fuzzy RD specification described in the subsequent section with covariates as the dependent variable, including lagged treatment as proposed by Eggers et al. (2018a, pp. 223-225). If a jump can be observed at the threshold for lagged council size, municipalities whose council sizes were slightly above the threshold in the previous electoral cycle are more frequently again to the right than their counterparts immediately at the left of the cutoff. This imbalance would potentially imply that official population numbers are manipulated in favor of municipalities above the threshold (Eggers et al., 2018a, p. 224). Following Cattaneo et al.'s (2019, p. 84) recommendation, I use a different bandwidth selection procedure than before, one that is coverage error rate (CER) optimal (instead of the MSE-optimal one used in section 7).

Similar to the Mechanisms section 5.2.3, the statistical inference is not correct when running multiple balance tests. Fusejima et al. (2022) provide one viable avenue for RDD balance tests by testing a single joint hypothesis. Due to the comparably low number of covariates, I refrain from using more advanced approaches and employ the easier solution of multiple tests instead.

Of the seven covariates used (plus lagged treatment), three turn out statistically significantly different from zero at the five percent-significance level, a surprisingly high number. They are reported in table 2, the insignificant ones in table 12 in the Appendix. Two of these control variables are obviously highly collinear as a higher share

Table 2: Local-Polynomial RDD – Balance Tests: Statistically Significant Covariates

Dependent variable	Independent large city (dummy)	Share of working-age population	Share of population aged > 65 years
RD effect	0.0077	-0.0016	0.0024
	(0.0015)	(0.0007)	(0.0012)
	$p < 0.001^{***}$	$p = 0.024^{*}$	$p = 0.039^{*}$
Obs.	103,618	103,618	103,618
Bandwidth	0.0509	0.0433	0.0377
Obs. below	7,430	6,253	5,587
Obs. above	8,100	6,934	5,982

Notes: Results of local-polynomial RD balance tests on significant covariates. For the non-significant covariates, see table 12. Model includes threshold, year and state FE. Robust bias-corrected standard errors clustered at the threshold level in parenthesis. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

of working-age people goes along with fewer people aged 65 and more ($r = -0.59$). However, neither of these two variables is more than weakly correlated to the independent large cities dummy ($|r| < 0.1$), which thus poses a different question. Nevertheless, the Eggers et al. (2018a, pp. 223-225) hypothesis of lagged treatment leading to imbalance could not be confirmed ($p = 0.53$). Conclusions are similar for the MSE-optimal bandwidth.

To sum up, covariates appear imbalanced between the two sides of the cutoff. Independent large cities are more often on the right side, municipalities whose population is younger more often on the left side. This imbalance does not invalidate the RDD per se: If all imbalances are accounted for in the RD regression, the result will nevertheless be valid. However, with selection-on-observables possibly taking place, one may justifiably question whether selection-on-*un*observables might have taken place as well, which would invalidate the research design.

After finding weak evidence that the local-randomization approach might not have completely worked, I now turn to the results.

7 The "Law of 1/n" Revisited Empirically

This section briefly discusses the results of the more conventional OLS, FE, and IV methods. After that, the results from the RD approach are presented, including the first stage, and several robustness checks.

7.1 Conventional Results

For the sake of comparison, I am starting with the "conventional" methods. As table 3 shows, there is a positive association of council size with gross spending at least at the 10% significance level across all models, albeit of small magnitudes. An increase in council size by one seat (non-causally) raises municipal expenditures by about 0.2% to 0.8%. The OLS model exploits cross-sectional variation only. The second and third model include FE. Leveraging only within-municipality and within-year variation in model (2) dampens the effect, compared to model (1), by about three quarters. Adding covariates in model (3) hardly changes the council size coefficient, implying very weak correlation of the covariates with the regressor of interest.[11]

I also run IV models, applying Baqir's (2002, pp. 1334-1338) idea to German municipalities. He used council size 30 years ago as an instrument Z_{it} for today's council size CS_{it}. Bavarian data (one of the few states to offer easily accessible long-running data on council seats) lend themselves to council size *24* years ago as the instrument. Again using year and municipality FE, the first stage reveals a small coefficient (point estimate 0.005, 95% confidence interval: $[0.025, -0.0157]$, overall F statistic $F = 1.2$) and, on the second stage, a rather implausible estimate with a 95% confidence interval of $[4.7, -3.0]$ for the log-linear model. The reason may be few overall changes in council size within municipalities over time. Indeed, the correlation between the instrument and the endogenous regressor is quite strong: $r = 0.96$. Dropping municipality FE yields a more plausible first-stage 95% confidence interval on the instrument ($[0.993, 0.977]$, overall $F = 120$) and also on the second stage of $[0.0009, -0.002]$, displayed as model (4) in table 3. Given hardly any changes in Bavarian regulations of assigning council seats and few large changes in population, the first-stage coefficients should be expected to be close to one. Comparing the coefficient of model (4) to the OLS and FE ones, it shrank even further and lost statistical significance, now even having a negative sign. For the IV results to be valid, however, the IV assumptions need to hold. As for the relevance condition, $cov(CS_{it}, Z_{it}) \neq 0$ is plausibly fulfilled in the second IV regression, which becomes clear when looking at the F statistic. The exclusion restriction, however, is not likely to be valid, as can be seen from the high correlation of the instrument Z_{it} and the endogenous regressor.

To conclude, all these models are seriously flawed. Model (1) from table 3 potentially suffers from both omitted-variable and simultaneity bias. Models (2) and (3) are more credible in terms of omitted variables specific to fixed effects within years and

---

[11]This could in principle also be the result of either no correlation of the covariates with the dependent variable or of mutually offsetting correlations. However, some of the regressors do affect the dependent variable statistically significantly, and when adding controls in a step-wise manner to the FE model, the coefficient remains unaltered as well (not shown). Hence, this scenario appears rather unlikely.

Table 3: Results from the "Conventional" Regression Methods

Dependent Variable:	Ln of gross expenditure p. c.			
Model	OLS	FE bivariate	FE multivariate	IV multivariate
	(1)	(2)	(3)	(4)
Council size	0.0081***	0.0023$'$	0.0025*	-0.0006
	(0.0002)	(0.0011)	(0.0011)	(0.0007)
Controls	Yes	No	Yes	Yes
Year FE	No	Yes	Yes	Yes
Municipality FE	No	Yes	Yes	No
SE	het.	at yr. & mcp.	at yr. & mcp.	at yr.
Data sample	All data	All data	All data	BY only
Observations	105,186	107,515	105,186	26,639
R^2	0.2001	0.7561	0.7616	0.2069
Within R^2		0.0001	0.0015	0.0447

Notes: Results from the "conventional" regression methods. Controls, FE and standard error (SE)s as indicated in the respective columns. Controls, if included: population, share of population over 65 years, unemployment rate, total area, share of working-age population, and dummies for independent large cities, cities, as well as year in the electoral cycle. Standard errors in parentheses: heteroskedastic (het.); if SE starts with "at", they are clustered at the year (yr.) and/or municipality (mcp.) level. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

municipalities. However, both biases, especially the reverse causality one, may still be present in all three models, as explained in section 3.1.2. The IV approach, in principle able to overcome both biases, suffers from potential violations of its specific premises, leaving us with no credible causal estimate.

7.2 The RDD First Stage Descriptively

This shortcoming motivates the RDD. In the section at hand, I will attempt to convince the reader of the relevance of the "instrument" pop_{it} (or more precisely, the forcing variable) for the endogenous regressor CS_{it}, and more specifically will demonstrate its relevance *at the threshold*.

First of all, the population data gathered here appear to be extremely close to those used by the municipalities and/or states to assign "standard" council size as prescribed by law. A few states have a comparatively weak correlation between the predicted number of council seats based on my exact population data (Baden-Württemberg, North Rhine-Westphalia, Saxony), starting at $r = 0.93$, while the rest of them has an even higher correlation. Four states – among them two states with a large number of municipalities, Schleswig-Holstein and Bavaria – have a correlation coefficient of $r > 0.99$. The small

Saarland and the handful of independent large cities of Saxony-Anhalt exhibit perfect correlation. Overall correlation of predicted and actual seats is $r = 0.98$. Correlation close to the thresholds (10% window) is lower, as municipalities apparently tend to deviate more often from what is prescribed by state law, but in general still quite high: $r = 0.97$. To summarize, municipal councils predominantly stick to their "assigned" council size.

A weaker correlation between predicted and actual council size goes hand in hand with a more "messy" assignment mechanism, as demonstrated graphically in figure 3. The states with comparatively low correlations mentioned in the previous paragraph are those where jumps at the cutoff are not as clear-cut as in other states. This may be due to specific state rules, e. g. allowing municipalities to autonomously change council size and/or "overhang seats". For the vast majority of states, a jump at the threshold is well visible. Upon further inspection, this holds even for the more "messy" states: In Baden-Württemberg, the lower bound of council size increases for each council size "bin" at least at certain cutoffs; in North Rhine-Westphalia, the presence of optional council size increases and "overhang seats" is clearly visible but nevertheless, the number of council seats jumps at the cutoff.

This section showed descriptively that a jump in treatment assignment probability takes place at the cutoffs, motivating the fuzzy design. The first stage of the parametric and non-parametric RDD models will also be discussed in the subsequent sections.

7.3 RDD Results

For the main test of this paper, the Weingast et al. (1981) "Law of 1/n" model, I first briefly present the results of a parametric polynomial IV RDD estimation.

As can be seen in table 4 for the FE specifications, the estimate on council size is negative but not statistically significantly different from zero. I use a ten-percent window around the cutoff, that is the ten percent of the observations that are closest to the cutoff. The nonsignificance extends throughout all IV RDD estimations, no matter the polynomial of the running variable (up to the sixth polynomial), the addition of covariates or different ad-hoc chosen bandwidths (the latter not shown). The coefficients in the FE specifications always indicate a single-digit percent decrease of council size at the threshold that is not significant. This stands in stark contrast to the "conventional" regression results from section 3 that suggested positive council size effects. Note, however, that that section looked at all observations while here, we focus on the

Figure 3: The Council Size Assignment Rule in German States
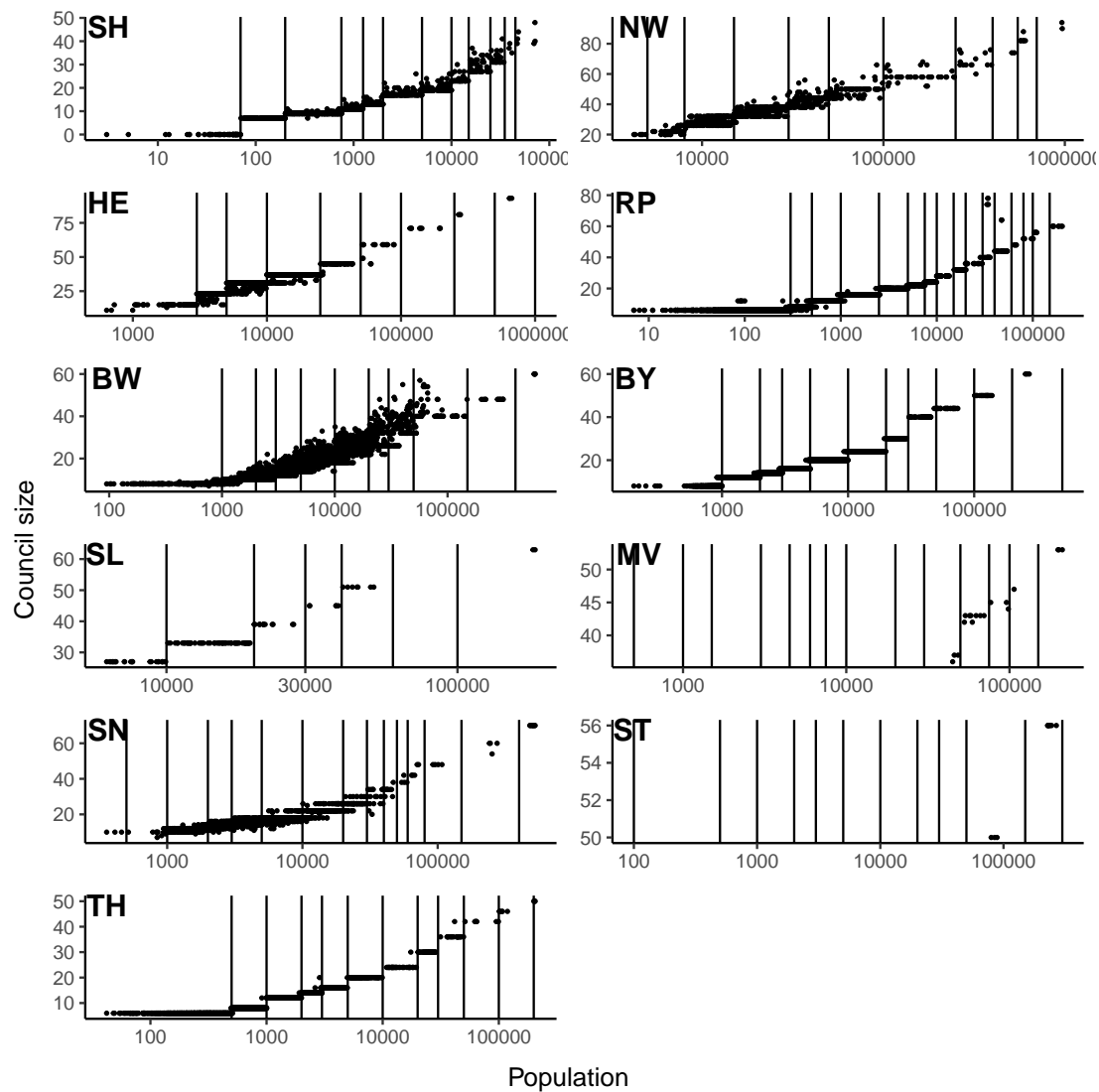
Plots show council size assignment mechanism. Vertical lines indicate thresholds prescribed by state law. For Mecklenburg-Vorpommern and Saxony-Anhalt, independent large cities only. For state abbreviations, refer to the List of Abbreviations.

Table 4: Results of the IV RDD

Dependent Variable:	Ln of gross expenditure p. c.			
Model	w/o covariates & w/o FE	w/o covariates & w/ FE	w/ covariates & w/o FE	w/ covariates & w/ FE
	(1)	(2)	(3)	(4)
RD effect	0.0125	-0.0176	0.0092	-0.0128
	(0.0256)	(0.0150)	(0.0243)	(0.0118)
Controls	No	No	Yes	Yes
Threshold FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
Municipality FE	No	Yes	No	Yes
SE	at thr.	at mcp., yr. & thr.	at thr.	at mcp., yr. & thr.
Obs.	103,618	103,618	101,409	101,409
Window around the threshold	10%	10%	10%	10%
Obs. below	5,073	5,073	4,933	4,933
Obs. above	5,288	5,288	5,207	5,207
Adjusted R^2	0.2336	0.7924	0.3094	0.7990
Within R^2	-0.0533	-0.0133	0.0445	-0.0246

Notes: Results from the IV RDD, in a ten-percent window around the cutoffs. Controls, FE and SE as indicated in the respective columns. Controls, if included: population, share of population over 65 years, unemployment rate, total area, share of working-age population, and dummies for independent large cities, cities, as well as year in the electoral cycle. Standard errors clustered at the threshold, year and municipality level in parentheses. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

cutoffs only. So potentially (but, given the biases outlined above, rather unlikely) both conclusions might be true.

In general, non-parametric models are considered more credible because they are more flexible and put higher weight on the observations that matter the most, i. e. those closest to the threshold. Thus, they will be discussed in more depth than the IV approach.

As for the first stage, that is the jump at the threshold of the endogenous regressor council size at the population cutoffs, the point estimates range from an increase in council size of 1 to 1.4 seats. This result is always statistically significant at least at $\alpha = 0.1$, as shown in table 5. In three of the four models, it is significant at the $\alpha = 0.05$ level, and for the two FE models, it is even highly significant at $\alpha = 0.001$. The first-stage jump of council size is thus much smaller than the three-councilor increase reported by Höhmann (2017, p. 354, table 2). This may be the result of different

Table 5: Results from Local-Polynomial RDD – *First* Stage

Dependent variable:	Council size			
Model	RDD w/o covariates (1)	FE RDD w/o covariates (2)	RDD w/ covariates (3)	FE RDD w/ covariates (4)
RD effect of running variable	1.4107 (0.6163) $p = 0.068'$	1.242 (0.12) $p < 0.001^{***}$	0.9992 (0.3839) $p = 0.039^*$	1.3921 (0.1165) $p < 0.001^{***}$
Controls	No	No	Yes	Yes
Threshold FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
State FE	No	Yes	No	Yes

Notes: Results from the first stage of a local-polynomial RDD. Controls and FE as indicated in the respective column. Robust bias-corrected standard errors clustered at the threshold level in parenthesis. For control variables, number of observations, and bandwidths, refer to table 6. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

estimation strategies as I think Höhmann's treatment indicator is binary, based on his interpretation (p. 353; see footnote 4 in this paper).

Results from the second stage are shown in table 6. The coefficients fluctuate around zero, with some of them being positive and some negative. They never deviate from zero in a statistically significant fashion, as indicated by the *p* values, squaring up nicely with the results from the global-polynomial IV RDD results in table 4. The inclusion of covariates and FE do not alter the coefficient in a statistically meaningful way either. Again, the magnitude of the effect is quite small, ranging from about -1.1% to 1.4%. Evaluated at the mean, spending might have been increased by 19 euros per capita in the baseline model (2). Model (1) is also displayed visually in figure 15 in the Appendix.

This finding contradicts both Egger and Koethenbuerger (2010) and Höhmann (2017) who found large positive and small negative results, respectively. I did neither, in line with Holzmann and Zaddach's (2019) results for Lower Saxony. The absolute magnitude of 1.4% of the baseline specification, model (2), is similar to Höhmann's (2017), whose study is best comparable to mine, yet the sign and statistical significance are entirely different.

To conclude, H1 of a "Law of 1/n" is rejected.

Table 6: Results from Local-Polynomial RDD – *Second* Stage

| Dependent variable | Ln of gross expenditure p. c. | | | |
Model	RDD w/o covariates (1)	FE RDD w/o covariates (2)	RDD w/ covariates (3)	FE RDD w/ covariates (4)
RD effect	-0.0001	0.0142	-0.0114	0.0119
	(0.0236)	(0.0121)	(0.0311)	(0.0098)
	$p = 0.99$	$p = 0.22$	$p = 0.59$	$p = 0.20$
Controls	No	No	Yes	Yes
Threshold FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
State FE	No	Yes	No	Yes
Obs.	103,618	103,618	101,409	101,409
Bandwidth	0.1351	0.0496	0.0796	0.0659
Obs. below	20,465	7,238	11,809	9,680
Obs. above	20,947	7,893	12,431	10,371

Notes: Results of the second stage a local-polynomial RDD. Controls, FE and SE as indicated in the respective columns. Controls, if included: population, share of population over 65 years, unemployment rate, total area, share of working-age population, and dummies for independent large cities, cities, as well as year in the electoral cycle. Robust bias-corrected standard errors clustered at the threshold level in parenthesis. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

7.4 RDD Robustness Checks

In the upcoming section, I will further check the validity of the RD approach and test different specifications. The non-parametric RDD FE model without covariates (model (2) in table 6) will be regarded as the baseline model because it contains a number of FEs and is less extensive computationally than model (4), all while avoiding the "bad control" issue.

The following specification changes to the "default" RD model are run in addition, all of which are not shown to preserve space. The software package used allows for different variance-covariance matrix estimators, none of which visibly change the results presented previously, and neither do the kernels: As expected (Cattaneo et al., 2019, p. 47), Epanechnikov and uniform kernels yield similar coefficients and confidence bands. Varying the degree of the polynomials used to estimate the coefficient and the degree of polynomials used for bias correction does not yield visibly altered coefficients or statistically significant p values.

7.4.1 Confounded Thresholds

I will apply the two approaches presented in section 5.2.3 for the mechanisms: the parametric interaction-term IV RDD on the one hand and non-parametric local-linear subsample regression on the other.

Distinguishing between confounded and non-confounded thresholds, I find no significant treatment effect on government expenditure for both groups in the non-parametric regression. Both coefficients have positive signs without any indication for statistical significance at $p = 0.2$ and $p = 0.5$, respectively. The results from the parametric IV regression with interaction terms are similar: Table 7, model (1), shows that both the baseline treatment effect for non-confounded thresholds ($p = 1$) and the difference from that baseline effect for confounded thresholds (i. e. the interaction term; $p = 0.88$) are not significantly different from zero. Both can be treated as nonexistent. In model (2) of table 7, however, we see results based on my data and Höhmann's (2017, pp. 360-364) categorization of confounded thresholds. As laid out in section 5.2.2, I consider several of his categorizations questionable. The split-sample approach for the confounder variable as defined by Höhmann yields results similar to the ones before in terms of coefficient sign and magnitude as well as p values. Both null hypotheses are not rejected individually at $\alpha = 0.1$. The parametric approach, however, finds a negative coefficient of approximately -2.1% ($p < 0.001$) for the non-confounded thresholds, a magnitude similar to Höhmann's (2017) finding of a reverse "Law of 1/n". The interaction term is significantly positive at $p = 0.01$, indicating about a 2.1% raise of spending at confounded thresholds. The positive effect for confounded thresholds would hint at the confounding policy changes on average increasing spending at the thresholds.

To put it in a nutshell, I do neither find an effect in the preceding analysis over the entire sample, including confounded thresholds, nor for the distinction between confounded and non-confounded thresholds. This finding is different from Höhmann because I do find an effect in the interaction-term approach using his probably incorrect definition of confounded cutoffs. The magnitude of the non-confounded treatment effect is in line with what Höhmann (2017, p. 356, table 3) finds for these cutoffs.

7.4.2 Placebo Tests

Running placebo tests to look for changes elsewhere in the running variable allows us to test the continuity assumption. If there are several jumps elsewhere in the running variable and one at the threshold, we may not causally attribute the jump at the threshold to the RDD as the general underlying function appears to be "jumpy". Technically, this

Table 7: Results of the interacted IV RDDs to investigate (non)confounded thresholds

| Dependent Variable: | Ln of gross expenditure p. c. | |
Model	Confounded thresholds based on my research (1)	Confounded thresholds based on Höhmann's (2017) research (2)
Council size baseline RD TE for non-confounded thresholds	-0.00001 (0.0409)	-0.0215*** (0.0038)
Additional RD TE for confounded thresholds	0.0377 (0.2345)	0.0428* (0.0130)
Observations	103,589	103,618
Window around the threshold	10%	10%
Observations below	5,073	5,073
Observations above	5,288	5,288
Adjusted R^2	-1.0634	0.3281
Within R^2	-2.6511	-0.1842

Notes: Results from the IV RDD, in a ten-percent window around the cutoffs. No control variables but threshold, year, and municipality FE. Standard errors clustered at the threshold, year and municipality level in parentheses. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

is simply a re-run of the RDD with a cutoff not at zero. For the arbitrary placebo values, I choose the $[-0.15, 0.18]$ interval which is essentially identical to the interquartile range of the running variable (see table 1) and run the RDD in steps of 0.03 units of the forcing variable. Results are presented visually in figure 4: The function does not appear to be "jumpy" in general, underlining the general validity of the RD setting. Only one out of twelve thresholds (with the one at the "real" cutoff already reported in model (2) of table 6) is statistically significant at 5%, as would be expected for multiple testing. Results look similar if one splits the sample by the instrument, i. e. whether they are above or below the cutoff, and then runs placebo regressions, as recommended by Cattaneo et al. (2019, pp. 89-90): Two out of 24 placebo cutoffs turn out statistically significant at $\alpha = 0.05$.

7.4.3 "Donut RDD"

Barreca et al. (2011) and Barreca et al. (2016) were first to use "donut RDDs". As suggested by the name, in "donut RDDs" the observations closest to the cutoff are excluded from the analysis. This allows researchers to investigate the sensitivity of the RDD to the units in the immediate vicinity of the threshold (Cattaneo et al., 2019, pp. 92-94). A second reason is that sorting, if suspected, can be expected primarily close to the cutoff. To avoid including potentially sorted observations, "donut RDDs" are one

Figure 4: Local-Polynomial RDD Placebo Thresholds
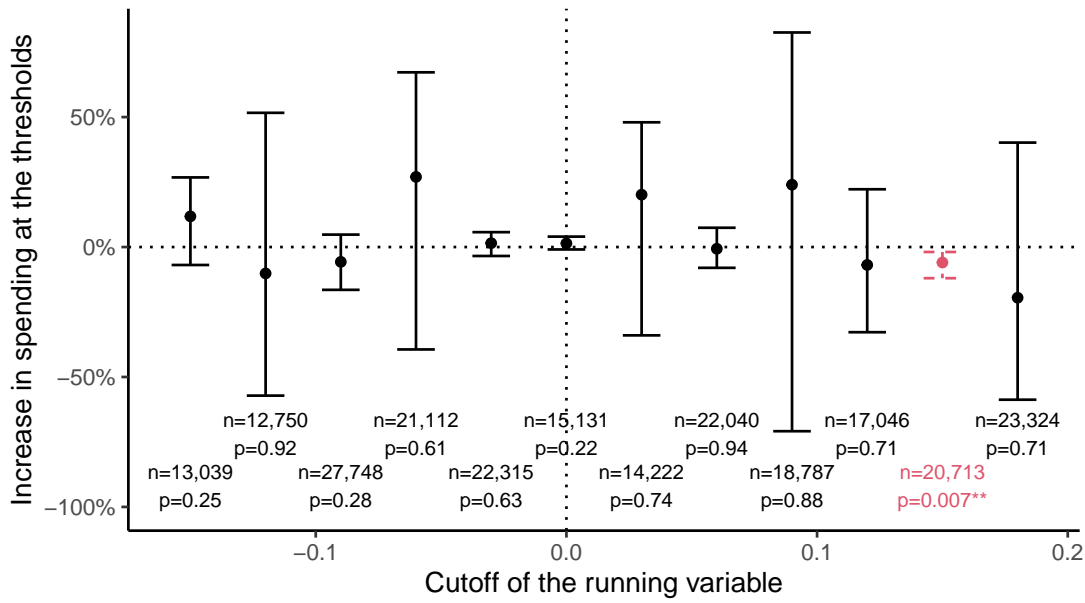
Notes: Plot shows standard coefficients, robust bias-corrected confidence interval, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from fuzzy local-polynomial RD estimations for different placebo thresholds. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.05$.

remedy. While sorting is not a major concern given the reassuring results from section 6.1, this approach serves as a sensitivity check with respect to the most "influential" data points, given that triangular (and also Epanechnikov) kernels put higher weights on near-threshold observations.

Results are shown graphically figure 5. Unlike previously in the main specification, we do find a positive jump at the threshold for many specifications. The regressions to the right of the plot exclude a large number of the closest observations, almost 7.3% of the sample for the rightmost model, which is a lot. For comparison, in the main parametric IV RDD regressions in table 4, only the closest ten percent of the observations are included. Hence, the regressions to the left of plot 5 are likely to be more credible. They, too, provide evidence for a jump. This leads me to conclude that the null effects found above in section 7.3 are driven by units very narrow to the cutoff.

By the way, the first-stage effects are much higher here. For example, the estimation in which the $[-0.015, 0.015]$ interval is left out has a first-stage confidence interval of $[1.28, 1.77]$. This indicates that the rather small jumps in council size at the cutoffs in the main specification in section 7.3 are due to municipalities close to the thresholds voluntarily switching into treatment or dropping out of treatment.

Figure 5: Local-Polynomial "Donut RDD"

Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected p values, the number of excluded observations $n_{dropped}$, and the number of observations within the respective bandwidths from fuzzy local-polynomial RD estimations for different sizes of "dropped intervals" close to the thresholds. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.05$.

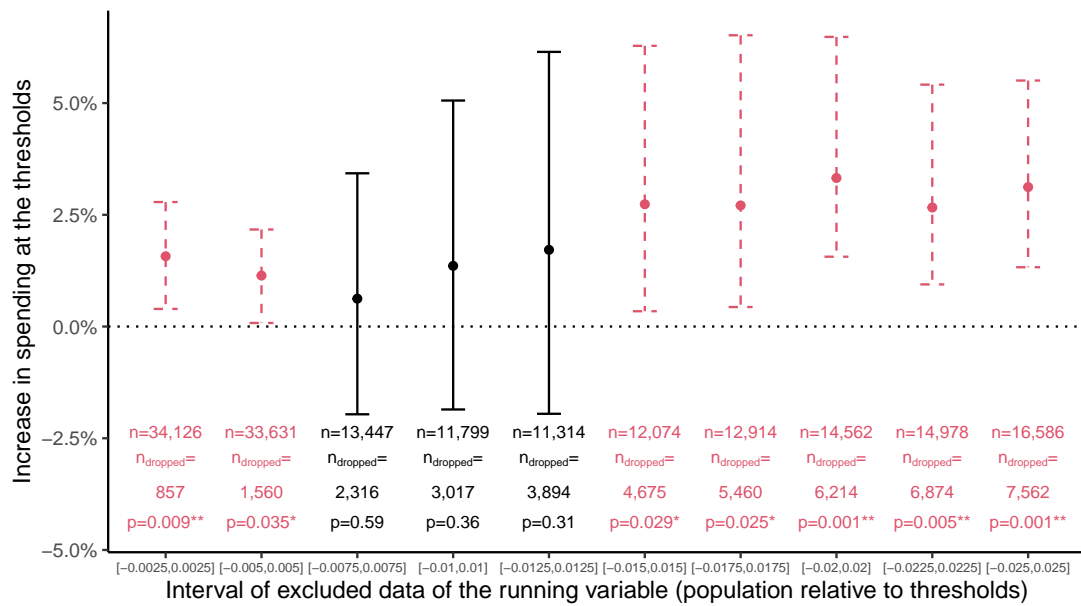### 7.4.4 Sensitivity with Respect to Bandwidth Choice

The sensitivity analysis shows the bandwidth for different windows around the cutoffs. The 10% window estimation, for instance, is based on the 10% of the observations closest to the cutoff, no matter what side of the cutoff they are on. This procedure is an extension to Höhmann's (2017, p. 356) approach. Figure 6, however, shows no deviation from the conclusions previously drawn because the vast majority of windows show no statistically significant effect at the cutoff. For the smaller windows, a few confidence intervals are statistically significantly different from zero. The bias correction, which by the way results in some point estimates (which are not bias-corrected) not overlapping with their accompanying confidence intervals, leads to some negative coefficients. On the one hand, even the smallest window at 1% comprises more than 1,000 observations and can thus not be considered truly small. On the other hand, these appear to be outliers when looking e. g. at the 5% window whose observations can still be considered to be in the narrow surroundings of the cutoff, and hence not (yet) suffer from endogeneity issues that occur when choosing a "too large" bandwidth.

When replacing this "ad-hoc" approach of choosing essentially arbitrary bandwidths by the data-driven bandwidth selection algorithms implemented in the `rdrobust` package,

Figure 6: Local-Polynomial RDD Bandwidth Sensitivity
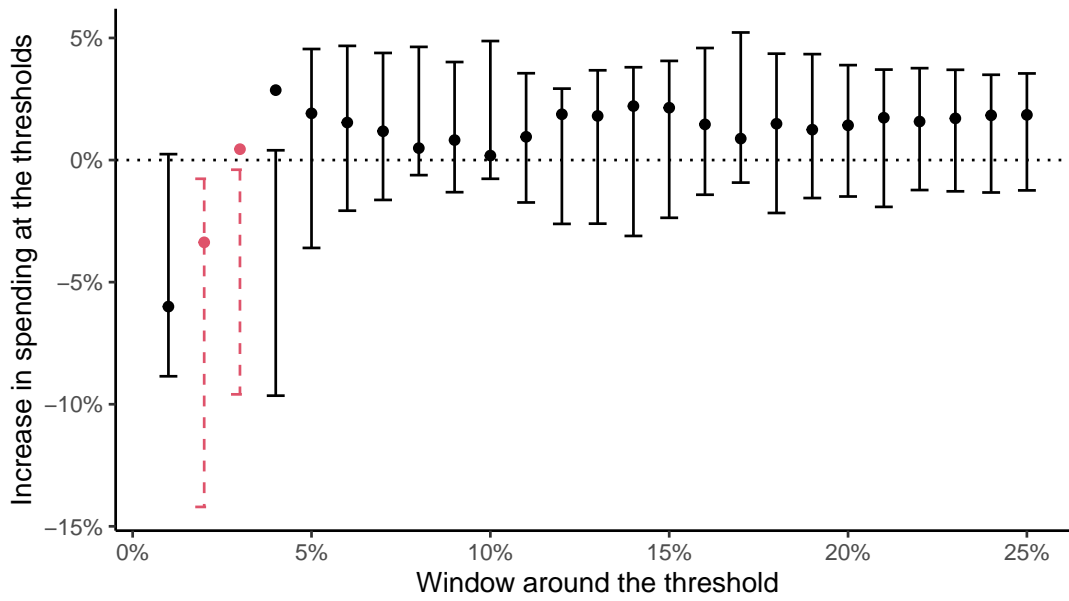
Notes: Plot shows coefficients and bias-corrected confidence intervals from local-polynomial RD estimations for different windows, similar to Höhmann (2017). The 10% window, for example, contains the 10% of the observations closest to the thresholds. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.05$. Coefficients and confidence intervals need not overlap as the first are standard coefficients while the latter are robust-corrected confidence intervals.

the RD treatment effects are very consistent in their point estimates and confidence intervals (not shown).

The last two subsections taken together, it looks like the RD treatment effect dummy is negative for smaller bandwidths (see plot 6) but positive for samples whose observations are further away from the cutoffs (see plot 5).

8 Mechanisms Driving the "Law of 1/n"

After empirically investigating the validity of the general RD setting and discussing the main results, I turn to the drivers behind the effect.

For the *average* treatment effect estimated above to be meaningful, it requires either assuming homogeneous treatment effects across all thresholds and covariates or simply ignoring potential heterogeneity from a researcher's point of view (Bertanha, 2020, pp. 216-217). In order for us not to ignore such heterogeneity, I now investigate it.

Table 8: Results from Local-Polynomial RDD – Financing

Dependent variable: Model	Ln of gross revenue p. c. FE RDD w/o covariates (1)	Ln of net expenditure p. c. FE RDD w/o covariates (2)
RD effect	0.0175 (0.0125) $p = 0.14$	0.0194 (0.0148) $p = 0.17$
Observations	103,613	103,555
Bandwidth	0.0502	0.0496
Observations below	7,295	7,228
Observations above	7,966	7,886

Notes: Results from a local-polynomial fuzzy RDD. Threshold, year and state FE are included, controls are not, similar to model (2) from table 6. Standard errors clustered at the threshold level in parenthesis. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

8.1 Financing the "Law of 1/n"

Had I found a "Law of 1/n", it would have been particularly interesting to see whether it was financed by additional tax revenue or debt. Given that there was no such effect, the H2 of the "Law" being financed by additional taxes might well be close to zero, too. This is closely aligned to the results from table 8 which neither show a significant effect for gross revenue p. c. nor for net expenditure p. c. The follow-up question asking for the taxes used to finance the "Law of 1/n" is hence rendered irrelevant. In stark contrast, Egger and Koethenbuerger (2010, pp. 210-211) found Bavarian municipalities to increase both debt and taxes at the thresholds.

8.2 Differences between States

Next, I investigate H3 stating that the treatment effect differs between states. Figure 7 (a) shows the increase in council size at the thresholds, i. e. the first stage. These changes differ widely between states, in line with the purely descriptive figure 3. The effect for Saxony is not significant, most likely owing to only small increases for municipalities with 10,000 inhabitants and fewer, concerning more than 60% of Saxon municipalities. More surprisingly, the effect for Saarland is not significant either, potentially because of the small sample size: The data-driven bandwidth algorithm selects only 18 close municipalities. As Saarland's confidence bands cover a wide space, they are not shown in the first plot of figure 7 (similarly those for Saxony on the second stage). For all other states, the effect is individually significant at least at the 1% level, as expected. States with lots of deviation, such as Baden-Württemberg, or with many

small increases, e. g. Rhineland-Palatinate, have only small jumps in council size at the cutoffs, whereas others such as Schleswig-Holstein, North Rhine-Westphalia, Hesse, Bavaria, or Thuringia have jumps of two councilors or more.

On the second stage displayed in figure 7 (b), the change in public expenditure is analyzed. North Rhine-Westphalia and Hesse show positive jumps of 3.6 and 1.6%, respectively, while Bavaria has a -2.7% decrease in spending at the thresholds. This stands in contrast to Egger and Koethenbuerger's (2010) results, with respect to both sign and magnitude, as they found a ten percent (or more) increase of spending at Bavarian thresholds. However, the results are non-significant for most of the states. The treatment-effect heterogeneity between states shown here might hint at positive and negative results canceling each other out when pooled as done in section 7.3.

The conclusions might be slightly altered in light of the sharp RDDs possible for a subset of states: Rhineland-Palatinate, Bavaria, Saarland, and Thuringia. Saxony-Anhalt does not have enough observations for a separate regression. While the sharp setting is immediately visible for Bavaria and Saarland in figure 3, it is essentially the norm for Rhineland-Palatinate and Thuringia as well, as determined by my legal research.[12] In this case, depicted in figure 16 in the Appendix, Rhineland-Palatinate and Thuringia remain insignificant. Bavaria, negatively significant in the fuzzy setting, now turns insignificant as well, while Saarland is positive across all conventional significance levels, with a large effect of around 23% (99% confidence interval: $[0.087, 0.371]$). As already indicated by the large confidence band, this might be a small-sample issue: The total sample contains 676 time-municipality observations from Saarland, of which $n = 145$ are selected by the bandwidth algorithm.

8.3 Different Political Systems

This section will look at how different aspects of the local constitutions shape the "Law".

H4, claiming that the electoral system divided into at-large and ward elections matters, is examined first. In the absence of pure ward elections, Schleswig-Holstein and North Rhine-Westphalia are the group holding mixed electoral systems, whereas the remaining states have at-large electoral systems. In the subsample approach shown in figure 8, the districting states are in line with the overall "Law of 1/n" result as there is no effect (point estimate: 0.9%, $p = 0.47$). The mixed group, however, proves statistically

---

[12]Saarland did not introduce the option for municipalities to change council size from the number prescribed by law until 2020. Thuringia allows for municipal changes to council size but only in cases of municipality mergers.

Figure 7: Local-Polynomial Fuzzy RDDs for Each State Separately
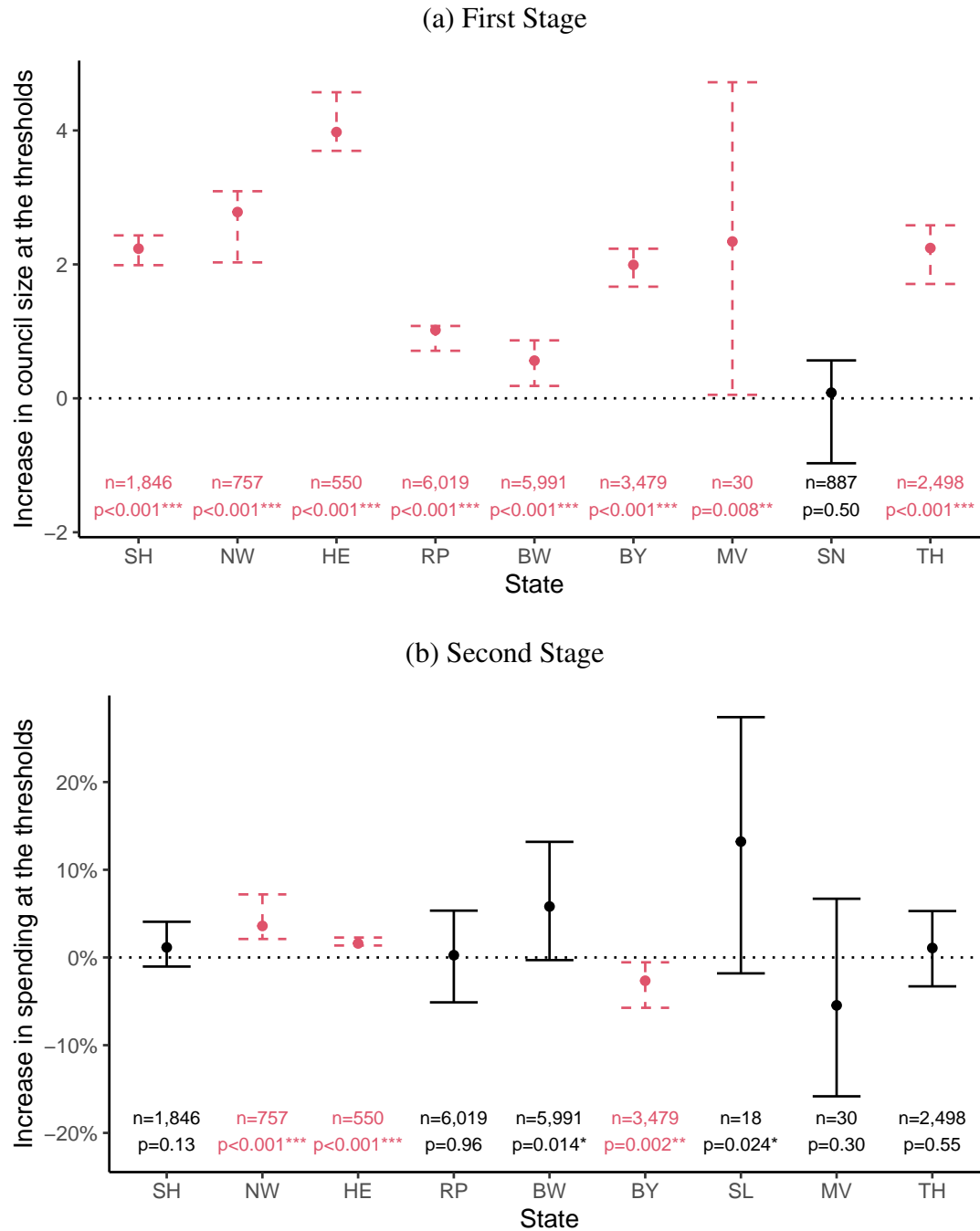
(a) First Stage

(b) Second Stage

Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from separate (split-sample) fuzzy local-polynomial RD estimations for each state, including those with sharp cutoffs. Saarland has been removed from the first plot owing to large confidence intervals, Saxony from the second for the same reason. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

Figure 8: Separate Local-Polynomial RDDs for At-Large and Mixed Elected Councils
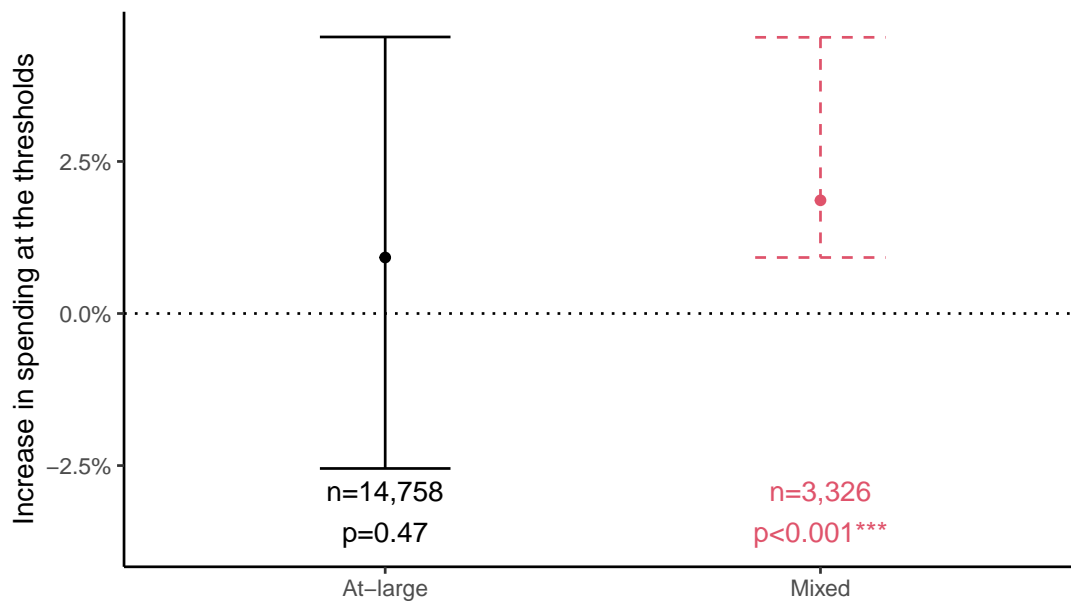
Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from separate (split-sample) fuzzy local-polynomial RD estimations for councils that are elected on an at-large basis and those elected in mixed systems, i. e. approximately half in wards and half at-large. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

significant with a 1.9% increase at the cutoff ($p < 0.001$). At first glace, this might hint at the partly mixed states indeed spending more at the cutoffs, as standard theory would expect. However, upon running the interaction-term IV RDD regression, the differences disappear as there is no statistically significant *difference* between these two values. This is not surprising given that the confidence bands of the subsample approach overlap for large parts of the interval. In addition, it is hard to distinguish potential state-specific treatment effects from ward vs. at-large treatment effects, especially since there are only two states in the mixed group.

To conclude with respect to H4, there appear to be no differences between these two electoral systems with respect to the "Law of 1/n", contrary to what Baqir (2002) argued. However, I do find municipalities whose councils are elected partly in wards to exhibit higher spending at the threshold.

According to H5, mayors with particularly strong roles can be expected to dampen the effect of the "Law of 1/n". Classifying only Hesse as a "magistrate constitution" (where the mayor has a rather weak position), this obviously yields a significantly positive effect (point estimate 1.6%, $p < 0.001$), since the Hessian treatment effect is rather large (see section 8.2). The effect for the remainder of states is not significant, with a point estimate of 1.4% ($p = 0.21$). Testing for differences between these two estimates

in the parametric interaction-term IV RDD yields no significant result either ($p = 0.84$). The result from the subsample approach is hard to interpret: Is the positive treatment effect found for Hesse a result exactly of the "magistrate constitution" and its weak mayoral role? Or has Hesse a positive treatment effect for different reasons but the coincidence with the "magistrate constitution" leads us to believe that the local political system is the reason for the effect? This question cannot definitively be answered in this analysis.

As a second approach, I classify a number of "core" "Southern German council type" states as such (SH, NW, RP, BW, BY, SL, SN), following Schubert and Klein (2018). The local-polynomial split-sample approach gives positive point estimates for both samples yet fails to reject the H0 with $p > 0.15$ in both cases. Similarly, the IV RDD gives a slightly positive estimate for "Southern German council type" states and a slightly negative one for those that are not, in any case non-significant with p values $p > 0.8$. To sum up, I cannot provide a definite answer to H5.

We have seen weak evidence for a positive "Law 1/n" effects in municipalities whose councilors are at least partly elected in wards and for municipalities whose mayors have a rather weak position. If the results were confirmed, they would give evidence for factors that have long been speculated to reduce the "Law's" effect (Baqir, 2002, pp. 1347-1351; Egger & Koethenbuerger, 2010, pp. 201-202; Höhmann, 2017, p. 347): Strong mayors and at-large elections. Nonetheless, I only found significant results employing one method but not the other, and the local constitutional types depend on state regulations, making it hard to disentangle potential state-specific treatment effects from these factors.

Hypothesis 6 predicted that the party affiliation of the mayor affects the treatment effect. Data on mayors in my data set was only collected for Bavaria. This means that as the Bavarian ATE we can only consider the negative effect of -2.3% ($p = 0.09$) from the sharp state RDDs shown in figure 16. Results of the sample-split method in a sharp RDD are shown in figure 9. A mayor nominated by the center-right CDU/CSU is found to dampen the "Law of 1/n" (-6.6%, $p = 0.002$) whereas both SPD and other mayors have a negative sign but lack statistical significance ($p > 0.3$). This finding is partially confirmed by the FE RD regression. Despite its size of -6.3%, the effect of having a CDU/CSU mayor is not statistically distinguishable from zero any more ($p = 0.33$). However, the joint F test on the two interaction coefficients, both of which are positive but individually insignificant, is statistically significant at $p = 0.001$, indicating that the effect is different from the CDU/CSU one, in the ballpark of zero percent and thus in line with the non-parametric regressions. To conclude, the party affiliation of Bavarian mayors matters for the "Law of 1/n", confirming H6.

Figure 9: Local-Polynomial Sharp RDDs for Bavaria, Sample Split by Mayoral Party
            Affiliation
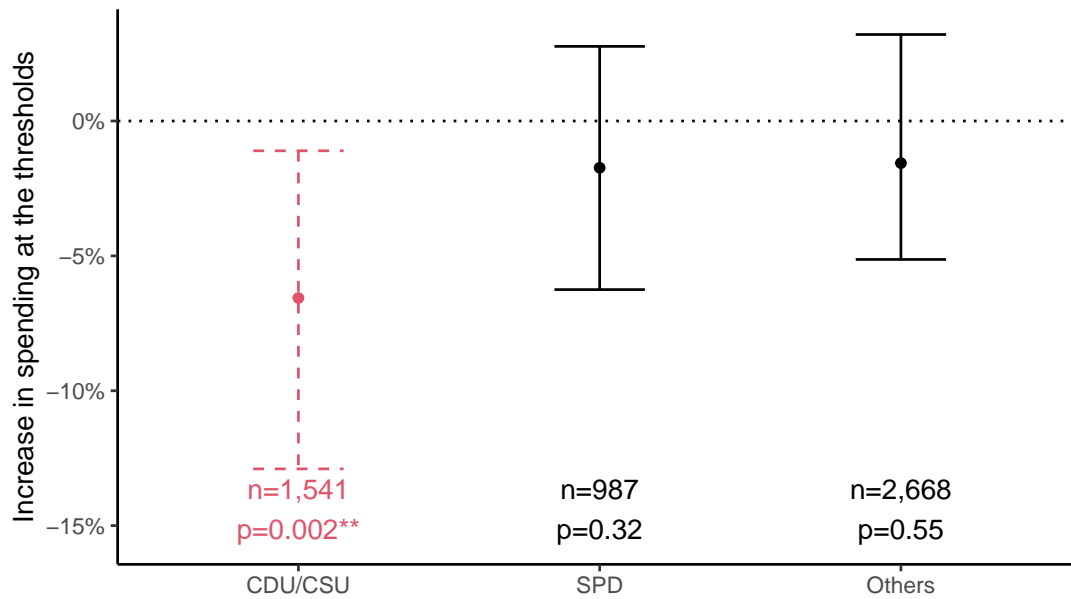
Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected
p values, and the number of observations within the respective bandwidths from separate (split-sample)
sharp local-polynomial RD estimations for Bavarian municipalities with samples split by the mayor's
party affiliation. Red dashed confidence intervals indicate an effect statistically significantly different
from zero at $\alpha = 0.01$.

H7 suggests that councils with more parties face a larger "common-pool" problem.
Nevertheless, the definition and counting of parties is not trivial. Statistical agencies at
times lump either small parties or voter groups together. This might lead to undercount-
ing of these types of parties, potentially causing measurement errors of this particular
covariate. Methodologically, one could think of several ways to measure the number
of parties and their parliamentary power. To simplify things, I calculate the (probably
undercounted) number of parties as given in the data set as well as the HHI to determine
the "market concentration" of these parties. Examining six states I have party seats
data on (NW, BW, BY, BY, SL, MV, SN), I find no evidence for any of these theories.
Starting with the (discrete) number of parties in parliament and splitting the sample by
this number (one to seven), one non-parametric regressions out of seven is statistically
significantly different from zero, the one at six parliamentary parties ($p = 0.002$), with
a negative coefficient sign. This result is in the ballpark of what I would have expected.
In the HHI split-sample approach, I split the continuous variable into four quartiles,
none of which are statistically significant at $\alpha = 0.01$. For the interaction-term method
of both the "total number" and the HHI approach, none of the regressors are significant
even at the $\alpha = 0.1$ level, neither individually nor jointly, neither as several categorical
variables (in the "total number" case) nor as one continuous variable. To sum up, there
appears to be no effect of the size of parliament on the "Law of 1/n".

Now I turn to the council composition and H8 predicting that absolute single-party majorities reduce the "Law's" effect. In the same six states as H7, I see no reason to pursue this theory further, as evidenced by the left-hand plot in figure 10. The subsample method reveals very weak evidence for councils without absolute majorities to exhibit the "Law" (plus 1.6% at the threshold, $p = 0.07$) whereas absolute-majority councils have no effect ($p = 0.68$). The sign and magnitude are in line with theoretical predictions but lack statistical significance. The parametric method does not find any evidence for this effect either, with no treatment-effect difference among the two groups ($p = 0.7$ on the interaction term).

H9 points to the ideology of the party holding the absolute council majority to affect the "Law". Owing to data and estimation limitations, however, I needed to lump together non-conservative parties (SPD and independent voter groups). Hence, this group might be rather heterogeneous in their ideology. Empirically, I do not find evidence for any of these hypotheses, as demonstrated in figure 10 on the right-hand side. The only effect remotely close to statistical significance is the one for "No absolute majority". This regression has already been discussed in the previous paragraph for H8. Similarly, the parametric method does not reveal significant joint differences from the non-significant CDU/CSU baseline effect, $p = 0.38$. To conclude, I have to reject H9 as well: Party affiliation does not seem to matter.

8.4 Different Types of Expenditure

While ideally, in this section I would investigate different types of expenditures, this is not possible due to data limitations, as outlined in section 3.2.3. Instead, I analyze possible effects of the "Law" regarding administrative and capital cost, with results displayed in table 11. The lowest p value is for the administrative cost category at $p = 0.08$. A test statistic this low must be expected for multiple testing. Since regressions are not even individually significant at the 5% level, this leads me to assume that there are no further differences between spending categories. H10, predicting that the treatment effect varies by category, is rejected.

8.5 Effects of Council Characteristics

In the penultimate mechanism subsection, I test Hypothesis 11 asserting that larger female representation affects the "Law of 1/n". I collected information on the share of female council members from five states (SH, BW, BY, SN, ST), and for two states (Baden-Württemberg, Bavaria) the share of female councilors in the largest

Figure 10: Separate Local-Polynomial RDDs for Each Value of the Two Council Composition Variables
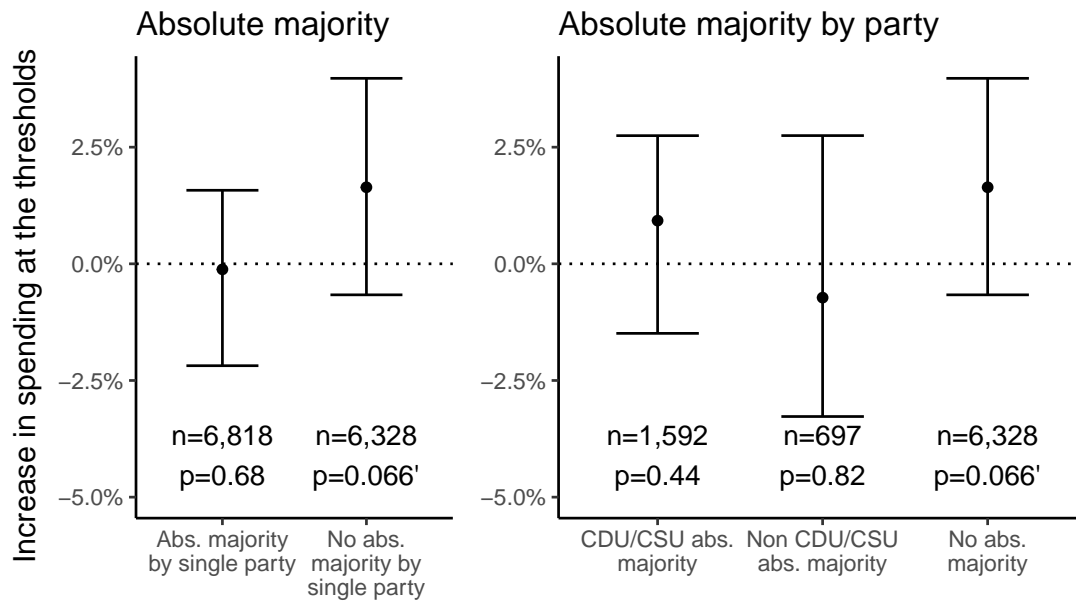
Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from separate (split-sample) fuzzy local-polynomial RD estimations for different values of the two variables describing council composition: The absolute-majority dummy and the absolute-majority dummy distinguishing by party ideology. The two regressions to the right of each plot are identical. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

Figure 11: Separate Local-Polynomial RDDs for Expenditure Subcategories
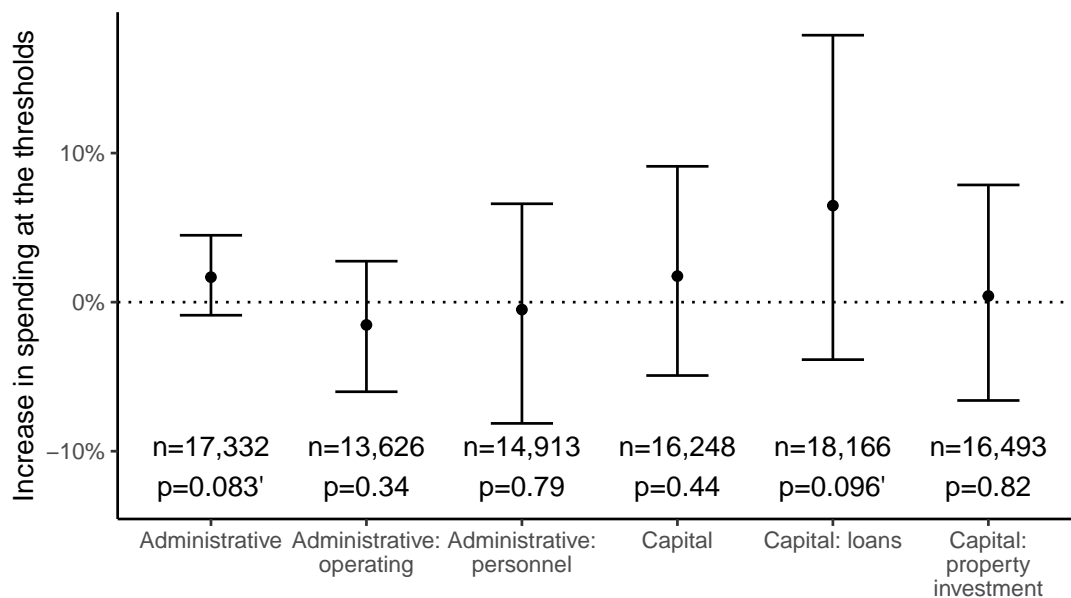
Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from separate fuzzy local-polynomial RD estimations for the six different categories of public spending. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

Figure 12: Separate Local-Polynomial RDDs for the Two "Women in Council" Variables
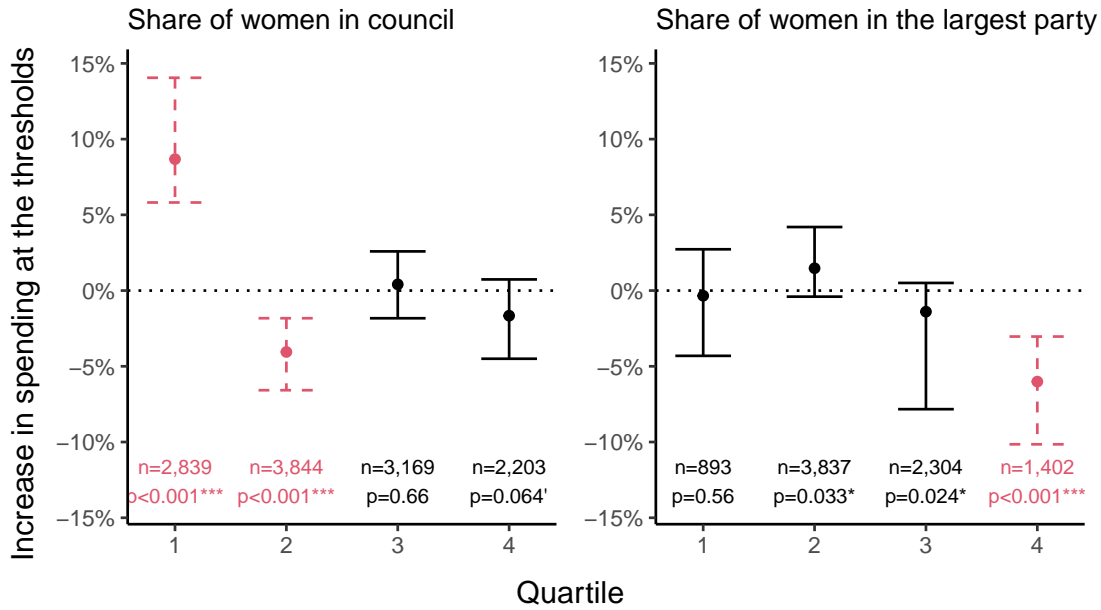
Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from separate (split-sample) fuzzy local-polynomial RD estimations for different subsets of the two variables based on their quartiles. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

party in council. Methodologically, for the split-sample approach the two continuous gender variables are split into four quartiles. I find ambiguous results shown in plot 12. Despite three out of eight coefficients being statistically significant at $\alpha = 0.001$ in the subsample method, the signs and significances remain mixed. Looking at possible patterns across quartiles, one hardly finds any. To exploit the continuous nature of the two covariates, it might be more informative to run interaction-term IV regressions. Their interaction terms, however, turn out non-significant at conventional levels, with $p = 0.37$ and $p = 0.57$, respectively: The treatment effect does not vary by the female share of councilors. To summarize, there does not appear to be any (clear) pattern between women in council or in the largest party and the "Law of 1/n". H11 cannot be confirmed.

8.6 Municipality Characteristics

H12 suggested that the effect might vary by the size of the municipalities, or more precisely by threshold. I ran the entire RDD setting at each cutoff separately, as revealed by figure 13. Each threshold split sample may include different states. The threshold at 10,000 inhabitants, for instance, is relevant in seven states, whereas a

Figure 13: Separate Local-Polynomial RDDs for Each threshold

Notes: Plot shows standard coefficients and robust bias-corrected confidence intervals from separate (split-sample) fuzzy local-polynomial RD estimations at each threshold. Several thresholds yield no results (e. g. due to few observations close to the threshold), the threshold at 100,000 inhabitants is dropped because its confidence interval is quite large. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

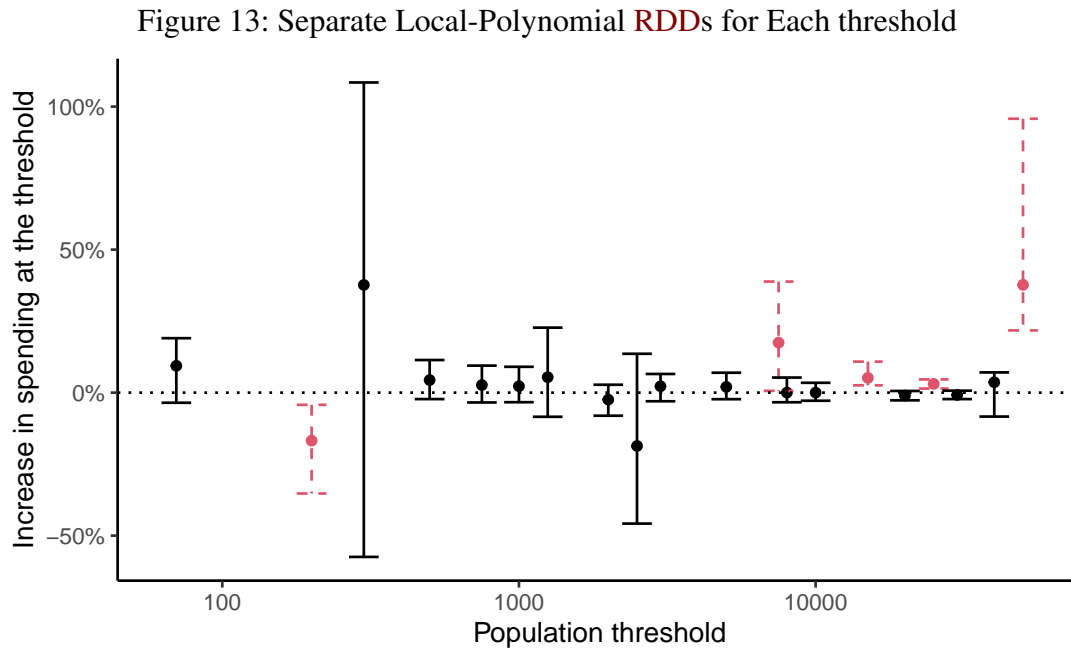large number of thresholds is relevant for one state only, e. g. the ones at 70 and 200 inhabitants (Schleswig-Holstein). If the treatment effect is indeed different between states, this might play a role here. For a number of cutoffs, the non-parametric RDD was unable to determine results, potentially owing to too few observations in close vicinity to the respective cutoffs: 13 cutoffs are excluded, the smallest of them at 45,000 inhabitants. As with numerous other threshold-heterogeneity checks, these do not reveal a clear pattern either. While the number of significant coefficients is rather large, their occurrence appears to be quite random. Thresholds whose subsample includes the states of Schleswig-Holstein, North Rhine-Westphalia, and Hesse appear to have lower *p* values, which is not surprising given that these to states have positive point estimates and two of some are highly statistically significant individually. This is again confirmed in the parametric regression whose interaction-term coefficient was slightly positive but, at $p = 0.95$, is far from any commonly accepted significance level.

The final hypothesis 13 suggested that the effect might differ by municipality type. For Höhmann (2017, p. 352), this was the reason to exclude independent large cities from his analysis. I, however, do not find evidence for different treatment effects, neither in parametric nor in non-parametric regressions. Figure 14 shows non-parametric results, None of the non-parametric split-sample regressions are individually significant at the $\alpha = 0.1$ significance level. The joint *F* test for difference in the parametric IV approach

Figure 14: Separate Local-Polynomial RDDs for Each Municipality Type
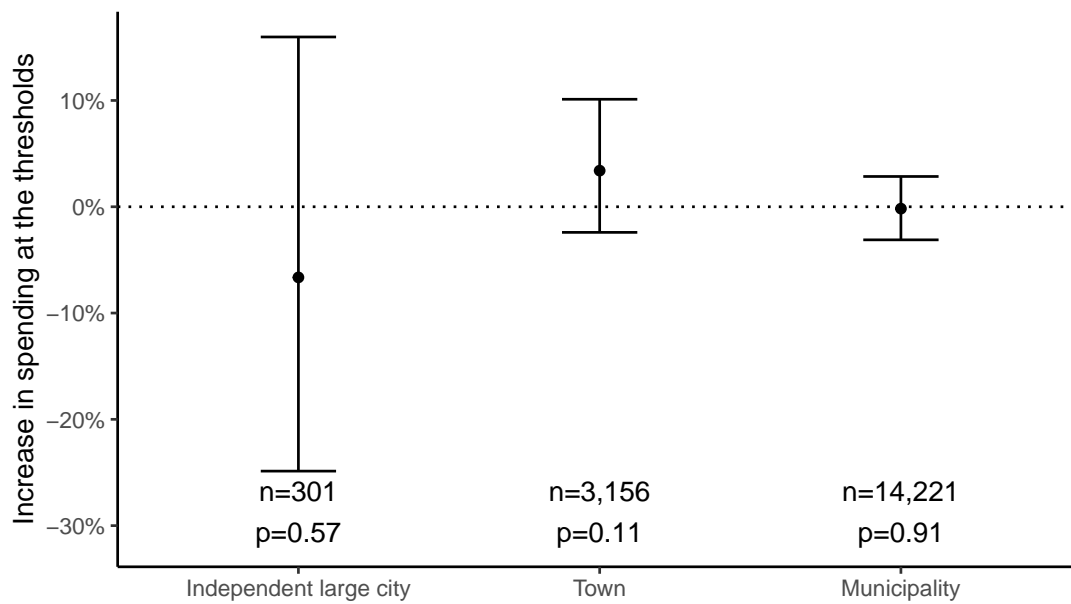
Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from separate (split-sample) fuzzy local-polynomial RD estimations, with samples split by municipality type. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

is not significantly different from the baseline effect (independent large city) either. To conclude the final hypothesis, H13 is rejected as well.

9 Conclusion

To briefly sum up the paper, I took the "Law of 1/n" asserting that public spending increases in legislature size to an empirical test using RDD. I find no evidence for sorting in population figures but evidence for covariate imbalance: Independent large cities and older municipalities sort into treatment more often. Regarding the effect of "Law", I do not find any effects on public spending (nor for net expenditure or revenue, for that matter). This might potentially be driven by differing treatment effects across states canceling each other out. The difference in treatment effects between states might also explain why papers on the "Law of 1/n" in Germany diverge as much as they do. Hesse, with a significantly positive "Law" effect, happens to be the only state to have a local constitution in which the mayor plays a comparably unimportant role. This might hint at stronger mayors being able to curb the "pork-barrel" externality. Similarly, states whose councils are partly elected in wards have a significant effect. These features, which have been argued to reduce the "Law's" effect, play a rather large role in German local politics and hence, might explain the null findings. However, the evidence is

rather weak as many of the deductions from the heterogeneity section could not be confirmed across methods, especially given the multiple-testing issue. The only effect robust across both methods were conservative mayors reducing spending in council size increases in Bavaria, compared to their counterparts affiliated with other parties. In general, however, I was unable to provide evidence for many of the hypothesized mechanisms.

The RD method comes with the downside of low external validity. One might think of other methods having more external validity while maintaining high internal validity. For instance, if someone came up with a valid instrument, this might lead to a credible IV estimate of the "Law of 1/n". Nevertheless, I was unable to do so and hence, employed the credible yet limited RD approach.

Finally, the RDD literature is evolving rapidly. The diff-in-disc method (Grembi et al., 2016), to my knowledge, has not been applied to the "Law of 1/n" in any country. Given the large number of thresholds and some changes to them, Germany might provide a good testing ground for causal diff-in-disc estimates. As for the examination of the mechanisms, I have sketched multiple avenues for further investigation. Hsu and Shen's (2019) test allows researchers to evaluate one heterogeneity-inducing variable at a time. Reguly's (2021a) machine-learning approach grows causal decision trees that even allow for a set of potential mechanism variables to be tested. Both approaches mark progress over previous practice in that their statistical inference is completely valid, given their assumptions are fulfilled. In addition, their code is completely and freely available online (Hsu & Shen, 2018; Reguly, 2021b).

References

Abrams, B. A., & Settle, R. F. (1999). Women's suffrage and the growth of the welfare state. *Public Choice*, *100*(3/4), 289–300. https://doi.org/10.1023/A:1018312829025

Aidt, T. S., & Dallal, B. (2008). Female voting power: the contribution of women's suffrage to the growth of social spending in Western Europe (1869–1960). *Public Choice*, *134*(3-4), 391–417. https://doi.org/10.1007/s11127-007-9234-1

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495. https://doi.org/10.1198/016214508000000841

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. https://doi.org/10.1162/003355399556061

Arnold, F., & Freier, R. (2015). Signature requirements and citizen initiatives: Quasi-experimental evidence from Germany. *Public Choice*, *162*(1-2), 43–56. https://doi.org/10.1007/s11127-014-0189-8

Asatryan, Z., Havlik, A., & Streif, F. (2017). Vetoing and inaugurating policy like others do: evidence on spatial interactions in voter initiatives. *Public Choice*, *172*(3-4), 525–544. https://doi.org/10.1007/s11127-017-0460-x

Bagues, M., & Campa, P. (2021). Can gender quotas in candidate lists empower women? Evidence from a regression discontinuity design. *Journal of Public Economics*, *194*, 104315. https://doi.org/10.1016/j.jpubeco.2020.104315

Baqir, R. (2002). Districting and Government Overspending. *Journal of Political Economy*, *110*(6), 1318–1354. https://doi.org/10.1086/342804

Baron, D. P. (1991). Majoritarian Incentives, Pork Barrel Programs, and Procedural Control. *American Journal of Political Science*, *35*(1), 57. https://doi.org/10.2307/2111438

Barone, G., & de Blasio, G. (2013). Electoral rules and voter turnout. *International Review of Law and Economics*, *36*, 25–35. https://doi.org/10.1016/j.irle.2013.04.001

Barreca, A. I., Guldi, M., Lindo, J. M., & Waddell, G. R. (2011). Saving babies? Revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, *126*(4), 2117–1223. https://doi.org/10.1093/qje/qjr042

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. *Economic Inquiry*, *54*(1), 268–293. https://doi.org/10.1111/ecin.12225

Baskaran, T. (2013). Coalition governments, cabinet size, and the common pool problem: Evidence from the German states. *European Journal of Political Economy*, *32*, 356–376. https://doi.org/10.1016/j.ejpoleco.2013.09.009

Becker, S. O., Egger, P. H., & von Ehrlich, M. (2013). Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects. *American Economic Journal: Economic Policy*, *5*(4), 29–77. https://doi.org/10.1257/pol.5.4.29

Bel, G., Raudla, R., Rodrigues, M., & Tavares, A. F. (2018a). These rules are made for spending: testing and extending the law of 1/n. *Public Choice*, *174*(1-2), 41–60. https://doi.org/10.1007/s11127-017-0488-y

Bel, G., Raudla, R., Rodrigues, M., & Tavares, A. F. (2018b). These rules are made for spending: testing and extending the law of 1/n: Supplementary material 1. *Public Choice*, *174*(1-2), 41–60. Retrieved January 30, 2022, from https://static-content.springer.com/esm/art%3A10.1007%2Fs11127-017-0488-y/MediaObjects/11127_2017_488_MOESM1_ESM.doc

Bergé, L. R. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA Discussion Papers*, (13). Retrieved September 13, 2022, from https://wwwfr.uni.lu/content/download/110162/1299525/file/2018_13%20Efficient%20estimation%20of%20maximum%20likelihood%20models%20with%20multiple%20fixed-effects%20-%20the%20R%20package%20FENmlm.pdf

Bergé, L. R. (2020). fixest: Fast and user-friendly fixed-effects estimation. Retrieved September 13, 2022, from https://lrberge.github.io/fixest/index.html

Bertanha, M. (2020). Regression discontinuity design with many thresholds. *Journal of Econometrics*, *218*(1), 216–241. https://doi.org/10.1016/j.jeconom.2019.09.010

Besley, T., & Case, A. (2003). Political Institutions and Policy Choices: Evidence from the United States. *Journal of Economic Literature*, *41*(1), 7–73. https://doi.org/10.1257/002205103321544693

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, *114*(2), 577–599. https://doi.org/10.1162/003355399556070

Bogumil, J., & Jann, W. (2020). *Verwaltung und Verwaltungswissenschaft in Deutschland*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-28408-4

Bradbury, J. C., & Crain, W. M. (2001). Legislative organization and government spending: cross-country evidence. *Journal of Public Economics*, *82*(3), 309–325. https://doi.org/10.1016/S0047-2727(00)00150-X

Bradbury, J. C., & Stephenson, E. F. (2003). Local Government Structure and Public Expenditures. *Public Choice*, *115*(1/2), 185–198. https://doi.org/10.1023/A:1022857028836

Brollo, F., Nannicini, T., Perotti, R., & Tabellini, G. (2013). The Political Resource Curse. *American Economic Review*, *103*(5), 1759–1796. https://doi.org/10.1257/aer.103.5.1759

Brollo, F., & Troiano, U. (2016). What happens when a woman wins an election? Evidence from close races in Brazil. *Journal of Development Economics*, *122*, 28–45. https://doi.org/10.1016/j.jdeveco.2016.04.003

Buchanan, J. M., & Yoon, Y. J. (2002). Universalism through Common Access: An Alternative Model of Distributive Majoritarian Politics. *Political Research Quarterly*, *55*(3), 503–519. https://doi.org/10.1177/106591290205500301

Calonico, S., Cattaneo, M. D., Chandak, R., Farrell, M. H., Jansson, M., Ma, X., Masini, R., Titiunik, R., & Vazquez-Bare, G. (2022). RD Packages: Regression Discontinuity Designs. Retrieved September 13, 2022, from https://rdpackages.github.io/

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2018). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, *113*(522), 767–779. https://doi.org/10.1080/01621459.2017.1285776

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *23*(2), 192–210. https://doi.org/10.1093/ectj/utz022

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for Regression-discontinuity Designs. *The Stata Journal*, *17*(2), 372–404. https://doi.org/10.1177/1536867X1701700208

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression Discontinuity Designs Using Covariates. *Review of Economics and Statistics*, *101*(3), 442–451. https://doi.org/10.1162/rest_a_00760

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014a). Robust Data-Driven Inference in the Regression-Discontinuity Design. *The Stata Journal*, *14*(4), 909–946. https://doi.org/10.1177/1536867X1401400413

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, *82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. *The R Journal*, *7*(1), 38. https://doi.org/10.32614/RJ-2015-004

Casas-Arce, P., & Saiz, A. (2015). Women and Power: Unpopular, Unwilling, or Held Back? *Journal of Political Economy*, *123*(3), 641–669. https://doi.org/10.1086/680686

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2018). *A Practical Introduction to Regression Discontinuity Designs: Volume II: Preliminary draft*. Retrieved July 24, 2022, from https://cattaneo.princeton.edu/books/Cattaneo-Idrobo-Titiunik_2018_CUP-Vol2.pdf

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A Practical Introduction to Regression Discontinuity Designs*. Cambridge University Press. https://doi.org/10.1017/9781108684606

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation Testing Based on Density Discontinuity. *The Stata Journal*, *18*(1), 234–261. https://doi.org/10.1177/1536867X1801800115

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, *115*(531), 1449–1455. https://doi.org/10.1080/01621459.2019.1635480

Cattaneo, M. D., Keele, L., Titiunik, R., & Vazquez-Bare, G. (2016). Interpreting Regression Discontinuity Designs with Multiple Cutoffs. *The Journal of Politics*, *78*(4), 1229–1248. https://doi.org/10.1086/686802

Cattaneo, M. D., & Vazquez-Bare, G. (2017). The Choice of Neighborhood in Regression Discontinuity Designs. *Observational Studies*, *3*(2), 134–146. https://doi.org/10.1353/obs.2017.0002

Chari, V. V., & Cole, H. L. (1995). A Contribution to the Theory of Pork Barrel Spending. Retrieved January 29, 2022, from http://minneapolisfed.org/research/sr/sr156.pdf

Chattopadhyay, R., & Duflo, E. (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*, *72*(5), 1409–1443. https://doi.org/10.1111/j.1468-0262.2004.00539.x

Cook, T. D. (2008). "Waiting for Life to Arrive": A history of the regression-discontinuity design in Psychology, Statistics and Economics. *Journal of Econometrics*, *142*(2), 636–654. https://doi.org/10.1016/j.jeconom.2007.05.002

Correa, G., & Madeira, R. A. (2014). The Size of Local Legislatures and Women's Political Representation. Evidence from Brazil. Retrieved February 12, 2022, from https://economia.uniandes.edu.co/sites/default/files/imagenes/eventos/Correa-Madeira2014.pdf

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press. https://doi.org/10.2307/j.ctv1c29t27

Dalenberg, D. R., & Duffy-Deno, K. T. (1991). At-Large versus Ward Elections: Implications for Public Infrastructure. *Public Choice*, *70*(3), 335–342. Retrieved February 1, 2022, from https://www.jstor.org/stable/30025474

de Benedetto, M. A. (2018). The effect of council size on municipal expenditures: evidence from Italian municipalities. Retrieved February 11, 2022, from https://eprints.bbk.ac.uk/id/eprint/26853/1/26853.pdf

de Benedictis-Kessner, J., & Warshaw, C. (2016). Mayoral Partisanship and Municipal Fiscal Policy. *The Journal of Politics*, *78*(4), 1124–1138. https://doi.org/10.1086/686308

de Britto, D. G. C., & Fiorin, S. (2016). Corruption and Legislature Size: Evidence from Brazil. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2859940

Deutscher Städtetag. (2021). *Stadtfinanzen 2021: Schlaglichter des Deutschen Städtetages* (Vol. 118). Deutscher Städtetag.

Dippel, C. (2022). Political Parties Do Matter in US Cities . . . for Their Unfunded Pensions. *American Economic Journal: Economic Policy*, *14*(3), 33–54. https://doi.org/10.1257/pol.20190480

Drew, J., & Dollery, B. (2017). The Price of Democracy? Political Representation Structure and Per Capita Expenditure in Victorian Local Government. *Urban Affairs Review*, *53*(3), 522–538. https://doi.org/10.1177/1078087416629806

Egger, P., & Koethenbuerger, M. (2010). Government Spending and Legislative Organization: Quasi-Experimental Evidence from Germany. *American Economic Journal: Applied Economics*, *2*(4), 200–212. https://doi.org/10.1257/app.2.4.200

Eggers, A. C. (2015). Proportionality and Turnout: Evidence from French Municipalities. *Comparative Political Studies*, *48*(2), 135–167. https://doi.org/10.1177/0010414014534199

Eggers, A. C., Fowler, A., Hainmueller, J., Hall, A. B., & Snyder, J. M. (2015). On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races. *American Journal of Political Science*, *59*(1), 259–274. https://doi.org/10.1111/ajps.12127

Eggers, A. C., Freier, R., Grembi, V., & Nannicini, T. (2018a). Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions. *American Journal of Political Science*, *62*(1), 210–229. https://doi.org/10.1111/ajps.12332

Eggers, A. C., Freier, R., Grembi, V., & Nannicini, T. (2018b). Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions: Supplementary information. Retrieved September 10, 2022, from https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fajps.12332&file=ajps12332-sup-0001-Appendix.pdf

Ehlers, D. (2007). Die Gemeindevertretung. In T. Mann & G. Püttner (Eds.), *Grundlagen und Kommunalverfassung* (pp. 459–534). Springer. https://doi.org/10.1007/978-3-540-68884-6_21

Farnham, P. G. (1990). The impact of citizen influence on local government expenditure. *Public Choice*, *64*(3), 201–212. https://doi.org/10.1007/BF00124366

Ferreira, F., & Gyourko, J. (2009). Do Political Parties Matter? Evidence from U.S. Cities. *The Quarterly Journal of Economics*, *124*(1), 399–422. https://doi.org/10.1162/qjec.2009.124.1.399

Fiorino, N., & Ricciuti, R. (2007). Legislature size and government spending in Italian regions: Forecasting the effects of a reform. *Public Choice*, *131*(1-2), 117–125. https://doi.org/10.1007/s11127-006-9108-y

Fujiwara, T. (2011). A Regression Discontinuity Test of Strategic Voting and Duverger's Law. *Quarterly Journal of Political Science*, *6*(3-4), 197–233. https://doi.org/10.1561/100.00010037

Fusejima, K., Ishihara, T., & Sawada, M. (2022). Joint diagnostic test of regression discontinuity designs: multiple testing problem. https://doi.org/10.48550/arXiv.2205.04345

Gagliarducci, S., & Nannicini, T. (2013). Do Better Paid Politicians Perform Better? Disentangling Incentives From Selection. *Journal of the European Economic Association*, *11*(2), 369–398. https://doi.org/10.1111/jeea.12002

Gamalerio, M. (2020). Do national political parties matter? Evidence from Italian municipalities. *European Journal of Political Economy*, *63*, 101862. https://doi.org/10.1016/j.ejpoleco.2020.101862

Gelman, A., & Imbens, G. W. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456. https://doi.org/10.1080/07350015.2017.1366909

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. https://doi.org/10.1111/j.1540-5907.2010.00499.x

Giambona, E., & Ribas, R. P. (2018). Unveiling the Price of Obscenity: Evidence from Closing Prostitution Windows in the Netherlands. *SSRN Journal*. https://doi.org/10.2139/ssrn.2994037

Gilligan, T. W., & Matsusaka, J. G. (1995). Deviations from constituent interests: The role of legistlative structure and political parties in the States. *Economic Inquiry*, *33*(3), 383–401. https://doi.org/10.1111/j.1465-7295.1995.tb01870.x

Gilligan, T. W., & Matsusaka, J. G. (2001). Fiscal Policy, Legislature Size, and Political Parties: Evidence from State and Local Governments in the First Half of the 20th Century. *National Tax Journal*, *54*(1), 57–82. Retrieved January 23, 2022, from https://www.jstor.org/stable/41789534

Grembi, V., Nannicini, T., & Troiano, U. (2016). Do Fiscal Rules Matter? *American Economic Journal: Applied Economics*, *8*(3), 1–30. https://doi.org/10.1257/app.20150076

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. Retrieved October 31, 2020, from http://www.jstor.org/stable/2692190

Harvey, A. (2020). Applying regression discontinuity designs to American political development. *Public Choice*, *185*(3-4), 377–399. https://doi.org/10.1007/s11127-019-00696-2

Herfindahl, O. (1950). *Concentration in the Steel Industry*. Retrieved October 6, 2022, from https://archive.org/details/herfindahl-concentration-in-the-steel-industry-1950-publish/mode/2up

Hirota, H., & Yunoue, H. (2012). Local government expenditure and council size: Quasi-experimental evidence from Japan. Retrieved February 11, 2022, from https://mpra.ub.uni-muenchen.de/42799/1/MPRA_paper_42799.pdf

Hirschman, A. O. (1964). The Paternity of an Index. *American Economic Review*, *54*(5), 761–762.

Höhmann, D. (2017). The effect of legislature size on public spending: evidence from a regression discontinuity design. *Public Choice*, *173*(3-4), 345–367. https://doi.org/10.1007/s11127-017-0484-2

Holzmann, C., & Zaddach, O. (2019). Legend of the Pork Barrel? The Causal Effect of Legislature Size on Public Spending. *FinanzArchiv*, *75*(1), 39. https://doi.org/10.1628/fa-2018-0024

Hopkins, D. J. (2011). Translating into Votes: The Electoral Impacts of Spanish-Language Ballots. *American Journal of Political Science*, *55*(4), 814–830. https://doi.org/10.1111/j.1540-5907.2011.00534.x

Hsu, Y.-.-C., & Shen, S. (2018). Finalcodes. Retrieved October 7, 2022, from https://www.dropbox.com/sh/hacenzqpop3o15s/AAA-vdgQtgxwkBbundCkL_0Sa

Hsu, Y.-.-C., & Shen, S. (2019). Testing treatment effect heterogeneity in regression discontinuity designs. *Journal of Econometrics*, *208*(2), 468–486. https://doi.org/10.1016/j.jeconom.2018.10.004

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467–475. Retrieved October 4, 2022, from https://www.jstor.org/stable/2951620

Imbens, G. W., & Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies*, *79*(3), 933–959. https://doi.org/10.1093/restud/rdr043

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Ipsen, J. (2007). Die Entwicklung der Kommunalverfassung in Deutschland. In T. Mann & G. Püttner (Eds.), *Grundlagen und Kommunalverfassung* (pp. 565–659). Springer. https://doi.org/10.1007/978-3-540-68884-6_24

Kresch, E. P., Schneider, R., Veras de Paiva Fonseca, Henrique, & Walker, M. (2020). Pull Up a Chair: Municipal Council Size and Local Taxes in Brazil. Retrieved January 27, 2022, from https://cdep.sipa.columbia.edu/sites/default/files/cdep/WP89Kresch.pdf

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, *142*(2), 675–697. https://doi.org/10.1016/j.jeconom.2007.05.004

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Lewis, B. D. (2019). Legislature Size, Local Government Expenditure and Taxation, and Public Service Access in Indonesia. *Studies in Comparative International Development*, *54*(2), 274–298. https://doi.org/10.1007/s12116-019-09278-1

Litschig, S. (2012). Are rules-based government programs shielded from special-interest politics? Evidence from revenue-sharing transfers in Brazil. *Journal of Public Economics*, *96*(11-12), 1047–1060. https://doi.org/10.1016/j.jpubeco.2012.08.010

Litschig, S., & Morrison, K. M. (2013). The Impact of Intergovernmental Transfers on Education Outcomes and Poverty Reduction. *American Economic Journal: Applied Economics*, *5*(4), 206–240. https://doi.org/10.1257/app.5.4.206

Lott, J. J. R., & Kenny, L. W. (1999). Did Women's Suffrage Change the Size and Scope of Government? *Journal of Political Economy*, *107*(6), 1163–1198. https://doi.org/10.1086/250093

MacDonald, L. (2008). The impact of government structure on local public expenditures. *Public Choice*, *136*(3-4), 457–473. https://doi.org/10.1007/s11127-008-9308-8

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2), 698–714. https://doi.org/10.1016/j.jeconom.2007.05.005

Mukherjee, B. (2003). Political Parties and the Size of Government in Multiparty Legislatures. *Comparative Political Studies*, *36*(6), 699–728. https://doi.org/10.1177/0010414003254240

Naßmacher, H., & Naßmacher, K.-H. (2007). *Kommunalpolitik in Deutschland* (2., völlig überarbeitete und aktualisierte Auflage). VS Verlag für Sozialwiss. Retrieved

March 15, 2022, from https://link.springer.com/book/10.1007/978-3-531-90702-4

Perotti, R., & Kontopoulos, Y. (2002). Fragmented fiscal policy. *Journal of Public Economics*, *86*(2), 191–222. https://doi.org/10.1016/S0047-2727(01)00146-3

Pettersson-Lidbom, P. (2008). Do Parties Matter for Economic Outcomes? A Regression-Discontinuity Approach. *Journal of the European Economic Association*, *6*(5), 1037–1056. https://doi.org/10.1162/JEEA.2008.6.5.1037

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*(3-4), 269–278. https://doi.org/10.1016/j.jpubeco.2011.07.004

Primo, D. M., & Snyder, J. M. (2008). Distributive Politics and the Law of 1/n. *The Journal of Politics*, *70*(2), 477–486. https://doi.org/10.1017/S0022381608080444

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Retrieved June 23, 2022, from https://www.R-project.org/

Reguly, Á. (2021a). Heterogeneous Treatment Effects in Regression Discontinuity Designs. https://doi.org/10.48550/arXiv.2106.11640

Reguly, Á. (2021b). Regression Discontinuity Trees. Retrieved September 17, 2022, from https://github.com/regulyagoston/RD_tree

Ribas, R. P. (2016). Multidimensional Regression Discontinuity and Regression Kink Designs with Difference-in-Differences. Retrieved September 9, 2022, from https://www.stata.com/meeting/chicago16/slides/chicago16_ribas.pdf

Ribas, R. P. (n. d.). Codes. Retrieved September 9, 2022, from https://sites.google.com/site/r4ribas/codes

Roubini, N., & Sachs, J. D. (1989a). Government Spending and Budget Deficits in the Industrial Countries. *Economic Policy*, *4*(8), 99. https://doi.org/10.2307/1344465

Roubini, N., & Sachs, J. D. (1989b). Political and economic determinants of budget deficits in the industrial democracies. *European Economic Review*, *33*(5), 903–933. https://doi.org/10.1016/0014-2921(89)90002-0

RStudio Team. (2022). RStudio: Integrated Development Environment for R. Retrieved September 12, 2022, from https://www.rstudio.com/

Rudzio, W. (2015). Die Kommunen: Zwischen Verwaltung und Politik. In W. Rudzio (Ed.), *Das politische System der Bundesrepublik Deutschland* (pp. 355–383). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-06231-6_12

Sanz, C. (2020). Direct democracy and government size: evidence from Spain. *Political Science Research and Methods*, *8*(4), 630–645. https://doi.org/10.1017/psrm.2018.65

Schaltegger, C. A., & Feld, L. P. (2009). Do large cabinets favor large governments? Evidence on the fiscal commons problem for Swiss Cantons. *Journal of Public Economics*, *93*(1-2), 35–47. https://doi.org/10.1016/j.jpubeco.2008.06.003

Schubert, K., & Klein, M. (2018). *Das Politiklexikon: Begriffe, Fakten, Zusammenhänge* (7., vollständig überarbeitete und erweiterte Auflage). Dietz. https://doi.org/120505

Southwick, L., Jr. (1997). Local Government Spending and At-Large Versus District Representation; Do Wards Result in More "Pork"? *Economics and Politics*, *9*(2), 173–203. https://doi.org/10.1111/1468-0343.00027

Statistisches Bundesamt. (2021). Gemeinden nach Bundesländern und Einwohner-größenklassen am 31.12.2020. Retrieved September 8, 2022, from https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/08-gemeinden-einwohner-groessen.html

Svaleryd, H. (2009). Women's representation and public spending. *European Journal of Political Economy*, *25*(2), 186–198. https://doi.org/10.1016/j.ejpoleco.2008.12.004

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317. https://doi.org/10.1037/h0044319

Tyrefors Hinnerich, B., & Pettersson-Lidbom, P. (2014). Democracy, Redistribution, and Political Participation: Evidence From Sweden 1919-1938. *Econometrica*, *82*(3), 961–993. https://doi.org/10.3982/ECTA9607

Wahlrecht.de. (n. d.). Übersicht über die Wahlsysteme bei Kommunalwahlen. Retrieved September 9, 2022, from https://www.wahlrecht.de/kommunal/

Weingast, B. R., Shepsle, K. A., & Johnsen, C. (1981). The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics. *Journal of Political Economy*, *89*(4), 642–664. https://doi.org/10.1086/260997

Appendix

Council Size Thresholds

Table 9: Coucil Size Thresholds and Confounding Policies, 2002-2014 – Part 1

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
Schleswig-Holstein – municipalities (without independent large cities)				
0-70	0			
71-200	7		✗	Below, no council election
201-750	9	✓	✓	
751-1250	11	✓	✓	
1251-2,500[13]	13	✓	✓	
2,501-5,000	17	✓	✗	Increase in constituencies
5,001-10,000	19	✓	✗	Increase in constituencies
10,001-15,000	23	✗	✗	Increase in constituencies
				Quota requirement for citizen request decreases (s. 2012)
				Quota requirement for citizen petition decreases (s. 2013)
				Quota requirement for referenda decreases (s. 2013)
				Municipality can have city status
				Wage of mayor increases (s. at least 2010)
				Different label on ballot sheet
15,001-25,000	27	✗	✗	Full-time equal opportunities officer (s. 2006)
				Wage of mayor increases
25,001-35,000	31	✓	✓	
35,001-45,000	35	✓	✓	
≥45,001	39	✓	✓	
Schleswig-Holstein – independent large cities				
≥15,0000	43			

continues on next page

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
		Table 9 – continued from previous page		
≥15,0001	49		✗	Quota requirement for citizen request decreases (s. 2012)
				Quota requirement for citizen petition decreases (s. 2013)
				Quota requirement for referenda decreases (s. 2013)
				Wage of mayor increases
North Rhine-Westphalia				
0-5,000	20			
5,001-8,000	26	✓	✗	Number of signatories for candidates increases
8,001-15,000	32	✓	✓	
15,001-30,000	38	✓	●	Status of a city must be changed (s. 2007)
30,001-50,000	44	✗	✗	Wage of (deputy) mayor increases
				Quota requirement for citizen petition decreases
50,001-100,000	50	✗	✗	Quota requirement for citizen petitions decreases
				Status of a larger city (s. 2007)
				Quota requirement for referenda decreases (s. 2011)
				Quota requirement for voting out the mayor decreases (s. 2011)
100,001-250,000	58	✗	✗	Wage of (deputy) mayor increases
				Quota requirement for citizen petition decreases
				Quota requirement for referenda decreases (s. 2011)
				Quota requirement for voting out the mayor decreases (s. 2011)
250,001-400,000	66	✗	✗	Wage of (deputy) mayor increases
400,001-550,000	74	✓	✓	
550,001-700,000	82	✓	✓	
≥700,001	90	✓	✓	
Hesse				
0-3,000	15			

continues on next page

XVIII

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
3,001-5,000	23	✓	✓	
5,001-10,000	31	✗	●	Wage of mayor increases (s. 2014)
10,001-25,000	37	✗	✗	Wage of deputy mayor increases
25,001-50,000	45	✓	✓	
50,001-100,000	59	✗	✗	Additional competencies for municipalities
				Quota requirement for citizen petition decreases (s. 2011)
				Wage of (deputy) mayor increases
				Accounting agency
100,001-250,000	71	✗	✗	Quota requirement for citizen petition decreases (s. 2011)
				Quota requirement for referenda decreases
				Wage of (deputy) mayor increases
250,001-500,000	81	✓	✗	Wage of (deputy) mayor increases
500,001-1,000,000	93	✗	✗	Wage of (deputy) mayor increases
≥1,000,001	105	✓	✓	
Rhineland-Palatinate				
0-300	6			
301-500	8	✓	✗	Below, municipalities may be merged with another on state order
501-1,000	12	✓	✗	Number of signatories for candidates increases
1001-2,500	16	✓	✗	Number of signatories for candidates increases
				Municipalities may have more than one voting district ("*Stimmbezirk*")
2,501-5,000	20	✓	✗	Number of signatories for candidates increases
5,001-7,500	22	✓	✗	Municipalities may have multiple constituencies
				Number of signatories for candidates increases
7,501-10,000	24	✓	✗	Number of signatories for candidates increases

<div align="center">Table 9 – continued from previous page</div>

continues on next page

		Table 9 – continued from previous page		
Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
10,001-15,000	28	✗	✗	Quota for citizen request decreases
				Number of constituencies increases
				Number of signatories for candidates increases
				Wage of mayor increases
15,001-20,000	32	✗	✗	Municipalities may opt for full-time deputy (s. 2010)
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases
20,001-30,000	36	✗	✗	In "*Verbandsgemeinden*", may opt for full-time deputy mayor (s. 2010)
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases
30,001-40,000	40	✗	✗	Number of signatories for candidates increases
				Wage of (deputy) mayor increases
40,001-60,000	44	✗	✗	Number of deputy mayors may increase
				Number of constituencies increases
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases
60,001-80,000	48	✗	✗	Number of signatories for candidates increases
				Wage of (deputy) mayor increases
80,001-100,000	52	✗	✗	Number of deputy mayors may increase
				Number of signatories for candidates increases
100,001-150,000	56	✗	✗	Quota for citizen petition decreases
				Municipalities may have administrative office for large districts
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
≥150,001	60	✓	✗	Number of signatories for candidates increases
				Wage of (deputy) mayor increases

Baden-Württemberg

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
0-1,000	8			
1,001-2,000	10	✗	✗	Size of advisory board for top secret issues increases
				Wage of mayor increases
				Reduced opening time on election day
				Mayoral candidates with less than five votes not mentioned by name
2,001-3,000	12	✗	✗	Full-time mayor
				Wage of mayor increases
				Slight change to accounting regulations in cases of partiality (§ 93 (3) GemO)
3,001-5,000	14	✗	✗	Number of signatories for candidates increases
5001-10000	18	✗	✗	Wage of mayor increases (s. 2011)
				Mayoral candidates with less than five votes not mentioned by name
10,001-20,000	22	✗	✗	Quota requirement for citizen assembly decreases (s. 2015)
				Quota requirement for citizen request decreases (s. 2015)
				Deputy mayors
				Size of advisory board for top secreat issues increases
				Wage of mayor increases
				Number of signatories for candidates increases
20,001-30,000	26	✗	✗	Status of a larger city
				Quota requirement for youth assembly decreases (s. 2015)
				Wage of (deputy) mayor increases
				Number of signatories for candidates increases

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
Table 9 – continued from previous page				
30,001-50,000	32	✗	✗	Size of advisory board for top secret issues increases
				Wage of (deputy) mayor increases
50,001-150,000	40	✗	✗	Quota requirement for youth assembly decreases (s. 2015)
				Wage of (deputy) mayor increase
				Number of signatories for candidates increases
				Quota for citizen assembly decreases
150,001-400,000	48	✓	✓	
≥400,001	60	✓	✓	
Bavaria				
0-1,000	8			
1,001-2,000	12	✓	✗	Number of signatories for candidates increases
				Wage of mayor increases
2,001-3,000	14	✓	✗	Spin-off of municipality possible
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases
3,001-5,000	16	✓	✗	Slight changes of electoral laws (Art. 25, 31, 34 GLKrWG, § 45, 75 GLKrWO, some of those only s. 2006)
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases
5,001-10,000	20	✗	✗	Full-time mayor
				Honorary mayor in other municipalities if council desires
				Accounting agency
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases

		Table 9 – continued from previous page		
Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
10,001-20,000	24	✗	✗	Quota requirement for citizen assembly increases
				Quota requirement for citizen petition changes
				below, council may opt for honorary mayor
				Full-time council members as consultants possible
				Number of signatories for candidates increases
				Slight changes to electoral processes (§ 45, 88 GLKrWO, s. 2006)
				Wage of (deputy) mayor increases
20,001-30,000	30	✓	✗	Quota requirement for citizen petition changes
				Number of signatories for candidates increases
30,001-50,000	40	✗	✗	Municipalities can be "*große Kreisstadt*"
				Quota requirement for citizen petition changes
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases
50,001-100,000	44	✗	✗	Municipalities can be large independent city
				Quota requirement for citizen petition changes
				Quota requirement for referenda changes
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases

continues on next page

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
		Table 9 – continued from previous page		
100,001-200,000	50	✓	✗	Quota requirement for citizen petition changes
				Quota requirement for referenda changes
				City districts necessary
				Number of signatories for candidates increases
				Wage of (deputy) mayor increases
200,001-500,000	60	✓	✓	
≥500,001	14		✗	Quota requirement for citizen petition changes
Nuremberg	70			
Munich	80			

Notes: See notes below table 10.

---

[13]Threshold at 2,000 (at least 06/2004-04/2012).
[14]There is no council size given for municipalities whose population exceeds 500,000.

Table 10: Coucil Size Thresholds and Confounding Policies, 2002-2014 – Part 2

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
Saarland				
0-10,000	27			
10,001-20,000	33	✗	✗	Number of deputy mayors increases
				Wage of mayor increases
20,001-30,000	39	✗	✗	Number of signatories necessary for citizen petition increases
				One of top employees has do be allowed to hold judicial office
				Number of deputy mayors increases
				Full-time deputy mayors
				Full-time equal opportunities officer
				Accounting agency
				Wage of (deputy) mayor increases
30,001-40,000	45	✗	✗	Status of a larger city
				Changes to the (deputy) mayor's title
				Wage of (deputy) mayor increases
40,001-60,000	51	✗	✗	Number of signatories necessary for citizen petition increases
				Number of deputy mayors increases
				Wage of (deputy) mayor increases
60,001-100,000	57	✗	✗	Number of signatories necessary for citizen petition increases
				Wage of (deputy) mayor increases
≥100,001	63	✗	✗	Number of deputy mayors increases
				Changes of officials' titles
Mecklenburg-Vorpommern				
0-500	7			
501-1,000	9		✗	Number of councilors in "*Amtsausschuss*" increases (2005-2011)
				Municipalities should not have less than 500 inhabitants
1,001-1,500	11		✗	Municipalities may be "independently administered" under certain conditions
				Number of councilors in "*Amtsausschuss*" increases

Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
		Table 10 – continued from previous page		
1,501-3,000	13		●	Wage of (deputy) mayor increases (until 2011)
3,001-4,500	15		✗	Municipalities may have an "*Amt*"
				Number of councilors in "Amtsausschuss" increases
4,501-6,000	17		✓	
6,001-7,500	19		●	Number of councilors in "Amtsausschuss" increases (s. 2011)
7,501-10,000	21		✓	
10,001-20,000	25		✗	Full-time equal opportunities officer
				Wage of (deputy) mayor increases
20,001-30,000	29		✗	One of municipality's mayor or at least one of their employees must have qualification to hold judicial office
				Wage of (deputy) mayor increases
30,001-50,000	37		✗	Wage of (deputy) mayor increases)
50,001-75,000	43		●	Quota requirement for citizen petition changes (until 2011)
75,001-100,000	45		✓	
100,001-150,000	47		✗	In independent large cities increase in deputy mayors possible
				Wage of (deputy) mayor increases
≥150,001	53		✗	Wage of (deputy) mayor increases (s. 2011)
				Wage of (deputy) mayor increases
Saxony				
0-500	8			
501-1,000	10	✓	✗	Wage of mayor increases
1,001-2,000	12	✓	✗	Wage of mayor increases
2,001-3,000	14	✓	✗	Number of signatories for candidates increases
				Wage of mayor increases
3,001-5,000	16	✓	✗	Mayoral candidates with less than five votes not mentioned by name
				Wage of mayor increases

		Table 10 – continued from previous page		
Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
5,001-10,000	18	✗	✗	Mayor must work full-time
				Number of signatories for candidates increases
				Wage of mayor increases
10,001-20,000	22	✗	✗	Municipalities may opt for deputy mayors
				Number of signatories for candidates increases
				Wage of mayor increases
20,001-30,000	26	✗	✗	Full-time equal opportunities officer
				Accounting agency
				Number of signatories for candidates increases (s. 2013)
				Wage of (deputy) mayor increases
30,001-40,000	30	✗	✗	Groups in parliament must receive funds (s. 2014)
				Maximum number of deputy mayors increased
				Wage of (deputy) mayor increases
40,001-50,000	34	✓	✗	Wage of (deputy) mayor increases
50,001-60,000	38	✗	✗	Number of signatories for candidates increases
60,001-80,000	42	✗	✗	Maximum number of deputy mayors increase
				Wage of (deputy) mayor increases
80,001-150,000	48	✓	✓	
150,001-400,000	54	✓	✓	
≥400,001	60	✗	✗	Maximum number of deputy mayors increased
Saxony-Anhalt				
0-100[15]	4			
101-500[16]	8	✓	✓	
501-1,000	10	✓	✓	
1,001-2,000	12	✓	✓	
2,001-3,000	14	✓	✓	

		Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	
Population	Council size			Other policy changes
3,001-5,000	16	✓	✗	Municipalities may have multiple constituencies
				Wage of mayor increases
5,001-10,000	20	✗	✗	Wage of (deputy) mayor increases
10,001-20,000	28	✗	✗	Maximum number of signatories needed for citizen request increases (s. 2014)
				Wage of (deputy) mayor increases
20,001-30,000	36	✗	✗	Maximum number of signatories needed for citizen request increases (s. 2014)
				Maximum number of signatories needed for citizen petition increases
				Full-time equal opportunities officer (only in 2005)
				Wage of (deputy) mayor increases
30,001-50,000	40	✗	✗	Maximum number of signatories needed for citizen request increases (s. 2014)
				Wage of (deputy) mayor increases
50,001-150,000	50	✗	✗	Maximum number of signatories needed for citizen request increases
				Maximum number of signatories needed for citizen petition increases (until 06/2014)
				Formation of parliamentary group requires more council members (s. 2009)
				Wage of (deputy) mayor increases
150,001-300,000	56	✓	✗	Wage of mayor increases
≥300,001	60	✓	✓	

Thuringia

0-500	6			
501-1000	8	✗	✓	
1,001-2,000	12	✗	✗	Municipalities must form "*Hauptausschuss*"
2,001-3,000	14	✗	✓	

		Table 10 – continued from previous page		
Population	Council size	Noncon-founded Threshold According to Höhmann (2017)	Actual Non-Con-founded Threshold	Other policy changes
3,001-5,000	16	✗	✗	Municipalities can form "Landgemeinde" (s. 2008)
				Municipality must be part of municipal administrative association if below
				Full-time mayor
5,001-10,000	20	✗	✗	Number of votes a voter has changes
				Wage of mayor increases
10,001-20,000	24	✗	✗	Status of mayor changes
				Equal opportunities officer (at least 01/2003-03/2013)
				Accounting rules change (s. 2013)
				Wage of mayor increases
20,001-30,000	30	✗	✗	Equal opportunities officer (s. 2013)
				Wage of (deputy) mayor increases
30,001-50,000	36	✗	✗	Wage of (deputy) mayor increases
50,001-100,000	42	✗	✗	Number of deputy mayors increases
100,001-200,000	46	✗	✗	Number of deputy mayors increases
				Wage of (deputy) mayor increases
≥200,001	50	✗	✗	Number of deputy mayors increases
				Wage of (deputy) mayor increases

Notes: Tables 9 and 10 are similar to Höhmann's (2017, p. 360-364, table 5). Höhmann's legal analysis is limited to the 2008-2010 time frame whereas my analysis aims to cover the 2002-2014 period. A "✓" indicates a nonconfounded threshold, whereas the "✗" marks a threshold as confounded. "●" stands for a change in confounding policies over the course of the data set so that this threshold is not considered confounded for the entire time period. The item refers to the lower cutoff of the interval. "s." is short for "since". Legal research for Lower Saxony and Brandenburg available upon request. Tables 9 and 10 have been separated for technical reasons.

---

[15]Until 2014, after that 10 councilors.
[16]Until 2014, after that 10 councilors.

State Regulations

Table 11: State Regulations for Local Elections, 2002-2014

State	Sharp or Fuzzy?	Reason for More/Less Seats Than Standard	Relevant Population Data Cutoff Date	Population Cutoff Dates	Population Cutoff Dates Confirmed by Authorities?
SH	fuzzy	Overhang & compensatory seats; standard council size only decreasing after ministry's approval	December 31, three years prior to the election (§7 KWG)	E. 2013: 12-31-2010, E. 2008: 12-31-2005, E. 2003: 12-31-2000, E. 1998: 12-31-1995	✗
NW	fuzzy	Overhang & compensatory seats; possibility for municipalities to change council size	Biannually updated population figure published 18 months prior to the end of the election period (§78 KWahlO)	E: 2014: 06-30-2012, E. 2009: 06-30-2007, E. 2004: 06-30-2002, E. 1999: 06-30-1998	✗
HE	fuzzy	Possibility for municipalities to decrease council size	Population figure published by the last date prior to the determination of the election day (§148 HGO)	E. 2011: 09-30-2009, E. 2006: 12-31-2004, E. 2001: 12-31-1999	✓
RP	sharp		June 30, year prior to the election (§ 130 GemO RP)	E. 2014: 06-30-2013, E. 2009: 06-30-2008, E. 2004: 06-30-2003, E. 1999: 06-30-1998	✓

		Table 11 – continued from previous page			
State	Sharp or Fuzzy?	Reason for More/Less Seats Than Standard	Relevant Population Data Cutoff Date	Population Cutoff Dates	Population Cutoff Dates Confirmed by Authorities?
BW	fuzzy	In certain municipalities overhang & compensatory seats; possibility for municipalities to change council size	September 30, two years prior to the election (§ 57 KomWG)	E. 2014: 09-30-2012, E. 2009: 09-30-2007, E. 2004: 09-30-2002 E. 1999: 09-30-1997	✓
BY	sharp		Last population figure published earlier than six month prior to the election (Art. 55 GLKrWG); if population falls below threshold, council size shall not be reduced to the number prescribed by law until the next but one election	E. 2014: 03-31-2013, E. 2008: 12-31-2006, E. 2002: 12-31-2000, E: 1996: 12-31-1994	✓
SL	sharp		Population figure published 60 days prior to the election (§ 219 KSVG)	E. 2014: 09-30-2013, E. 2009: 09-30-2008, E. 2004: 09-30-2003, E. 1999: 09-30-1998	✗

continues on next page

			Table 11 – continued from previous page		
State	Sharp or Fuzzy?	Reason for More/Less Seats Than Standard	Relevant Population Data Cutoff Date	Population Cutoff Dates	Population Cutoff Dates Confirmed by Authorities?
MV	fuzzy	Decrease of council size in communities administered on a honorary basis	After 2015: Last available data as of December 31 (§60 LKWG MV); until 2014: Population cutoff date determined by Ministry of the Interior (§60 LKWG MV, similarly §4 KWG MV, old version, before 2011)	E. 2014: 12-31-2012, E. 2009: 06-30-2008, E. 2004: 06-30-2003, E. 1999: 06-30-1998	✓
SN	fuzzy	Possibility for municipalities to change council size	December 31, two years prior to the election (§65 KomWG)	E. 2014: 12-31-2012, E. 2009: 12-31-2007, E. 2004: 12-31-2002, E. 1999: 12-31-1997	✓
ST	sharp		December 31, two years prior to the election (since 2014: §158 KVG LSA, until 2014: §149 GO LSA)	E. 2014: 12-31-2012, E. 2009: 12-31-2007, E. 2004: 12-31-2002, E. 1999: 12-31-1997	✓
TH	sharp	One-time increase after mergers	Last population data published three months prior to the election (§ 37 ThürKWG)	E. 2014: 06-30-2013, E. 2009: 06-30-2008, E. 2004: 06-30-2003, E. 1999: 12-31-1997	✓

Notes: "E." is the abbreviation for "election". The column "Population Cutoff Dates Confirmed by Authorities" indicates whether a ministry or statistical agency confirmed ("✓") the population-cutoff dates used to determine council size. If not ("✗"), population-cutoff dates are my best guess based on legal research as well as communication with ministries, municipalities, and statistical agencies. Information on Lower Saxony and Brandenburg available upon request.

## Balance Tests for Nonsignificant Covariates

Table 12: Local-Polynomial RDD – Balance Tests: Non-Significant Covariates

Dependent variable	Unemployment rate	Total area in ha	City (dummy)	Lagged council size	Population (as of 12-31)
RD effect	-0.0015	-51.4934	0.0048	-0.0846	15.7506
	(0.0011)	(53.9151)	(0.0064)	(0.1809)	(22.6144)
	$p = 0.19$	$p = 0.30$	$p = 0.35$	$p = 0.53$	$p = 0.77$
Obs.	103,618	103,618	103,618	103,618	103,618
Bandwidth	0.0428	0.0760	0.0743	0.0420	0.0355
Obs. below	6,176	11,350	11,130	5,225	5,219
Obs. above	6,797	12,028	11,833	5,798	5,725

Notes: Results from local-polynomial RD balance tests on significant covariates. For the statistically significant covariates, see table 2. Model includes threshold, year and state FE. Robust standard errors clustered at the threshold level in parenthesis. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ' $p < 0.10$.

The Main RDD Result Visually

Figure 15: Local Linear Regression from Model (1) of Table 6
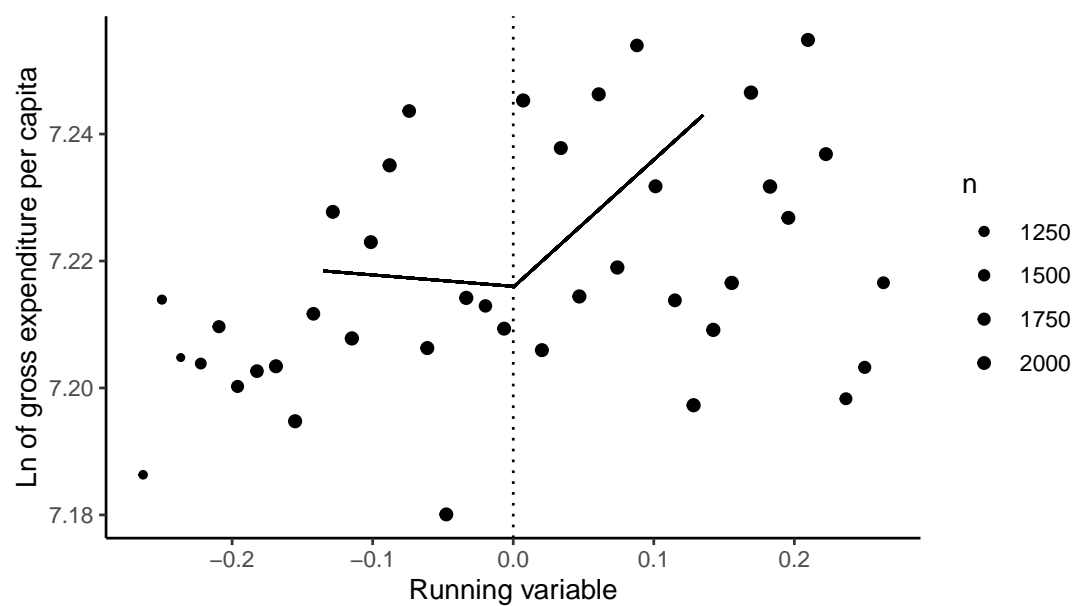
Notes: Plot shows the local linear regression run in model (1) of table 6 as well as evenly spaced binned averages for a span of the running variables that is double the size of the RD bandwidth ($n = 77,635$ municipalities displayed; 40 total bins with $min = 1,040$, $mean = 1,941$, $max = 2,240$ municipalities per bin).

Differences in the treatment effects between States – Sharp RDD

Figure 16: Local-Polynomial Sharp RDDs for Each State Separately
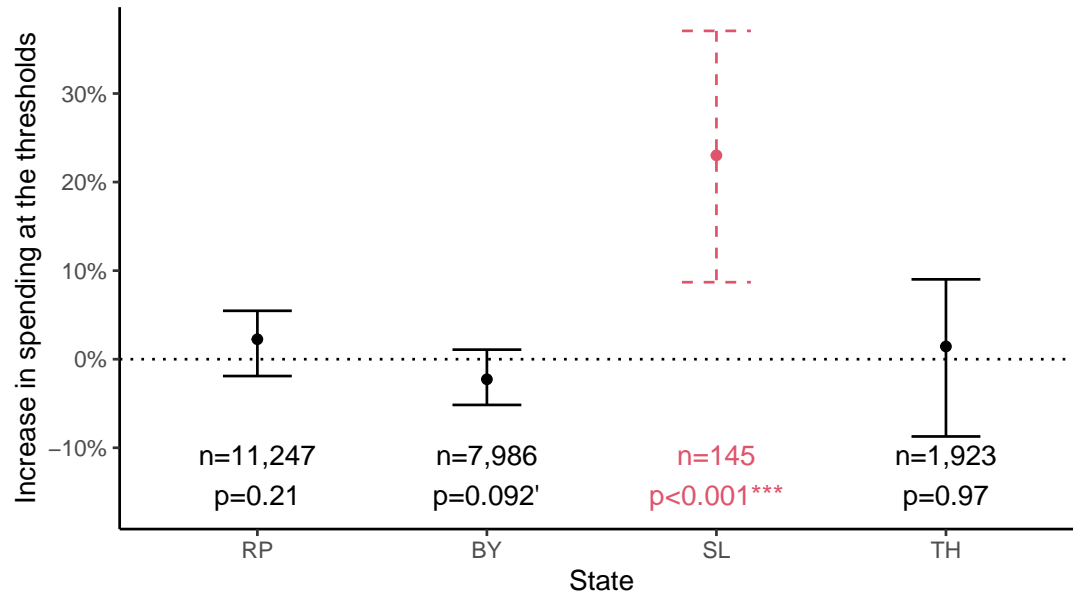
Notes: Plot shows standard coefficients, robust bias-corrected confidence intervals, robust bias-corrected *p* values, and the number of observations within the respective bandwidths from separate (split-sample) sharp local-polynomial RD estimations for each state that has sharp population thresholds: Rhineland-Palatinate, Bavaria, Saarland, & Thuringia. Red dashed confidence intervals indicate an effect statistically significantly different from zero at $\alpha = 0.01$.

Documentation

The code used in the analysis is available on GitHub: https://github.com/flo-fox/Master_Thesis.

Tables are exported using the 23rd commit to the GitHub repository, except table 7 which is based on the 24th commit to the repository. All tables have been manually modified in addition, as documented in the 28th commit.

Figures are based on the 28th commit as well.

Data as well as Stata and R scripts used to compile and merge the data are available upon request

The project study can be found in the GitHub repository as well.