Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

Discussion Papers

1136

Florian Ade • Ronny Freier

When Can We Trust Population Thresholds in Regression Discontinuity Designs?

Berlin, June 2011

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

When can we trust population thresholds in regression discontinuity designs?

Florian Ade
Humboldt University Berlin, DIW Berlin
Email: fade@diw.de

Ronny Freier
Stockholm School of Economics, DIW Berlin
Email: Ronny.Freier@hhs.se

Keywords: regression discontinuity design, population thresholds, local elections, government spending
JEL classification: C2, D7, H7

Abstract: A recent literature has used variation just around deterministic legislative population thresholds to identify the causal effects of institutional changes. This paper reviews the use of regression discontinuity designs using such population thresholds. Our concern involves three arguments: (1) simultaneous exogenous (co-)treatment, (2) simultaneous endogenous choices and (3) manipulation and precise control over population measures. Revisiting the study by Egger and Koethenbuerger (2010), who analyse the relationship between council size and government spending, we present new evidence that these three concerns do matter for causal analysis. Our results suggest that empirical designs using population thresholds are only to be used with utmost care and confidence in the precise institutional setting.

Acknowledgments: We would like to thank Charles B. Blankart, Peter Haan, Magnus Johannesson, Juanna Joensen, Henrik Jordahl, Erik Lindqvist, Christian Odendahl, Viktor Steiner and David Strömberg as well as colleagues at Stockholm School of Economics and DIW Berlin. Comments at the Applied Microeconomics Seminar at Stockholm School of Economics are also gratefully acknowledged. We would like to thank Peter Egger and Marko Koethenbuerger for making their data available online and Viktor Steiner and Peter Haan who invited us to work at their department for public economics at the DIW Berlin. Ronny Freier would like to thank the Jan Wallander and Tom Hedelius scholarship for continued funding. The usual disclaimer applies.

1 Introduction

Scholars in political economy devote much attention to the causal identification of the effects of fundamental rules and features of governmental organization (e.g. Persson and Tabellini (2002)). Estimating the causal effects of institutional designs and constitutional rules, however, is generally a difficult task for a number of reasons (see Acemoglu (2005)). The main challenges are that institutional rules are usually endogenous and seldom change, that different aspects of constitutional designs often correlate and change simultaneously and that data analysis is often limited by small samples. The new interest has therefore turned to subnational levels and statistical methods from the program evaluation literature in the attempt to use quasi-random variation in specific rules to estimate their impact.

One specific class of designs being used in a range of different applications are regression discontinuity designs that focus on population thresholds. Pettersson-Lidbom (2006) and Egger and Koethenbuerger (2010) investigate whether the municipality's council size, which changes at deterministic population thresholds, affects local government spending.[1] Regression discontinuity designs, based on population thresholds, are also used to study performance of politicians when salaries increase (Gagliarducci and Nannicini (2009) and Ferraz and Finan (2009)), the effect of fiscal transfers on local elections (Litschig and Morrison (2010)), corruption and the quality of politicians (Brollo, Nannicini, Perotti, and Tabellini (2009)), as well as the impact of representative democracy versus direct democracy on government spending (Hinnerich and Pettersson-Lidbom (2010)).

In this paper, we devote attention to the specific challenges of using population thresholds for reliable causal inference. Our concern is threefold: (1) The first challenge is that the population threshold used may not only define the treatment considered but also additional simultaneous exogenous co-treatments. (2) The identification and interpretation of the treatment effect is further complicated as additional endogenous choices on other institutions are often taken simultaneously. The timing of events is likely to coincide as it is often at specific times that changes are implemented (e.g. at the beginning of an election period). (3) Given that the official population count is usually observable at any moment, the specific concern is that the precise number can be manipulated. In the empirical analysis, we revisit the study by Egger and Koethenbuerger (2010) and present new evidence on the importance of these concerns.

[1] Both studies use rules (given by the federal or state law) that set the number of members of legislature based on population count of the locality. Those contributions are perceived to be of great importance as they focus on causal identification. For further studies on this topic compare e.g. Baqir (2002) for cities and counties in the US, Gilligan and Matsusaka (2001) for state and local governments in the US, Schaltegger and Feld (2009) for Swiss cantons.

Researchers using quasi-experimental designs recognize the importance of ensuring that a particular treatment effect stands in isolation to other confounding simultaneous treatments. In applications using difference-in-difference designs it is common practice to make an explicit argument that the treatment group is not simultaneously affected by additional treatments. Similarly, for the use of instrumental variables, special care is taken in arguing that the instrument is only of importance to the specific treatment under investigation (see Acemoglu (2005)).[2] We argue that this problem is critical when using regression discontinuity designs with population thresholds. Population count is an intuitive and easy way for higher level governments to impose differential rules on lower tier structures. Hence, it is very likely that the same thresholds might be used in different dimensions of rules and institutions.

According to Bavarian state law the number of town council members must increase at 10 different population thresholds. However, researching the details of Bavarian laws and bylaws we find that the same thresholds also determine large changes in both local institutions (e.g. referendum quotas, politicians wages) and communal finances (e.g. additional funds from the state government). In total, we find 14 additional legislative laws or bylaws that induce differential rules by population size at the same thresholds. Applying these rules, we show that a part of the spending effect at population thresholds is financed by increases in state grants and additional revenue from fees.

The second challenge in using population thresholds for identification is the timing of changes in the treatment and changes of other endogenous choices. Local institutions are often changed simultaneously at the beginning of new electoral cycles. In Bavaria, e.g., apart from the change in council size set by the exogenous rule, communities must make various (endogenous) decisions just before the next election cycle starts. Those decisions involve, for example, the nature of the mayor position (full-time or part-time) as well as what new tasks the community wants to be responsible for. For the identification of a specific treatment effect, this problem becomes twofold: First, because timing coincides it is harder for the researchers to disentangle the effect in question and, in small samples, pure statistical variation is more likely to introduce bias in the estimation. Second, the interpretation of the effect might crucially hinge on the timing and interdependency of the treatment changes and the choice making. We supplement the data of Egger and Koethenbuerger (2010) with information on one such endogenous decision (work status of the mayor) and no longer find a significant spending effect at the council size thresholds.

[2] Acemoglu (2005) discusses the fundamental problem for the standard instrumental variable approach: that the instrument used for e.g. an institutional setting may not only be an instrument for the institution under investigation but also for a different institution. In this case it is not clear what effect is estimated.

Our final concern is that the population count may be precisely manipulated. As population thresholds are critical for the remuneration of government personnel, for allocation of finances from other government levels, as well as for the council size, there are large incentives for sorting around the respective population thresholds. The concern of sorting on the precise threshold is well-recognized in the literature on regression discontinuity design (RDD) (see Lee and Lemieux (2009), McCrary (2008)). For population thresholds this is of particular concern as in Germany the number of inhabitants is by no means a surprise, but rather a regularly updated publicly known number. Hence, whenever a community comes close to a threshold the administration can purposefully manipulate the number and precisely sort on the desired population count. For Bavarian municipalities during the period of observation we find evidence that there is sorting around the population thresholds. Thus, the spending effects found by Egger and Koethenbuerger (2010) could be potentially biased by selection effects.

Given the results of our empirical analysis, we conclude that researchers must be cautious when using population thresholds for identification in regression discontinuity designs. To be covered against the threats of simultaneous exogenous co-treatment and simultaneous endogenous decisions, researches must acquire deep institutional knowledge while checking legal norms and customs thoroughly. Furthermore, testing of the key identifying assumptions of RDD is required.[3]

This study proceeds as follows. Section 2 introduces the empirical model and the underlying identification assumptions. Section 3 presents the institutional setting and the specific use of population thresholds in Bavarian law, before Section 4 highlights our results and draws comparisons to the findings reported by Egger and Koethenbuerger (2010). Section 5 concludes the analysis.

2 Empirical model and methodology

In this section, we consider the empirical methodology that is involved in the use of deterministic population thresholds in regression discontinuity designs (RDD). After discussing the basic empirical setup, we clarify the general and the specific identifying assumptions needed to allow for causal inference of the treatment effect in question.

[3]Caughey and Sekhon (2010) revisit the empirical evidence presented in Lee (2008), who uses a RDD on close election outcomes. The former rigorously investigate the RDD assumptions and find that, for the US House of Representatives, observations in close elections still exhibit crucially unbalanced predetermined variables that are likely to invalidate the research design.

2.1 Basic model

Assume, for simplicity, that we consider a case in which a treatment is determined at a single population threshold. Define $v_{i,t}$ as the distance of the number of inhabitants in location i at time t from the threshold. Assume further that the treatment (e.g. number of council members) is discontinuously determined at the threshold (sharp discontinuity design). The relationship between treatment $d_{1_{i,t}}$ and $v_{i,t}$ is as follows:

$$d_{1_{i,t}} = 1\left[v_{i,t} > 0\right] \tag{1}$$

The estimation equation specifying the RDD then reads as follows:

$$y_{i,t} = \delta_0 + \delta_1 d_{1_{i,t}} + h(v_{i,t}, \theta) + \epsilon_{i,t} \tag{2}$$

where $y_{i,t}$ is the outcome in question (e.g. local government spending), δ_1 is the parameter of interest and $h(\cdot)$ is a flexible function that represents the underlying general relationship between the distance to the threshold (hence population size) and the outcome variable. This simple framework can also be easily adjusted to accommodate the fact that treatment is changing at multiple thresholds (see Egger and Koethenbuerger (2010)).

The (see Lee and Lemieux (2009)):

2.2 Identifying assumptions

The parameter δ_1 is an unbiased estimate of the treatment effect under the critical assumption of continuity(see Lee and Lemieux (2009)). It is only if all observable and unobservable covariates, except treatment, are distributed continuously around the threshold that one can assume to have valid counterfactual observations on either side. If observations just right from the required population count are systematically different from the ones just to the left, then identification fails.

The first implication of the continuity assumption is that we must ensure that there exist no further co-treatments. It is obvious that causal inference of an individual effect cannot be upheld when other treatments are simultaneously determined at the same threshold. Formally, the existence of co-treatment implies the following. Let a second treatment, $d_{2_{i,t}}$, be determined at the same threshold v_0:

$$d_{2_{i,t}} = 1\left[v_{i,t} > 0\right] \tag{3}$$

The individual effects of $d_{1_{i,t}}$ and $d_{2_{i,t}}$ cannot be identified. The second treatment is omitted in eq. (2) and the flexible function in v cannot control for it as the threshold coincides.

Even if the second treatment is observable, one cannot include it in the regression due to multi-collinearity. The researcher can only identify a joint effect. If one has several thresholds and differences in when certain treatments apply, one may hope to disentangle the isolated treatments. However, the major concern is when and if we fail to recognize the existence of additional (co-)treatments. Although an outcome was in fact induced by several changing factors, we instead falsely attribute the effect to only one treatment.

It is important to note that co-treatment is of particular concern because standard RDD tests are not very likely to detect it. If a second treatment is implemented at the same threshold, distribution tests of predetermined covariates cannot be expected to detect such differences. Further, there is no reason to believe that a direct test of the distribution of the score variable (McCrary (2008)) will be of help.

For population thresholds the problem of co-treatments has to be critically reviewed on a case by case basis. Population thresholds are tools for legislatures to induce differences in laws and bylaws by the population size of local entities. Thus, the same thresholds are likely used in several legislative rules.

Apart from exogenous co-treatments, we consider the precise sorting around thresholds to be of particular concern when using population count in RDD. Lee (2008) shows that it is sufficient to show that there is a random component in the scoring variable to uphold the continuity assumption. If agents cannot precisely control the variable that determines the treatment, then in some neighborhood of the threshold, assignment of treatment is effectively random. This random assignment close to the threshold directly implies continuity.

Thus, any RDD application relying on this argument must investigate whether agents can precisely sort around the threshold. For population count this issue is indeed crucial. The official population of a municipality in Germany is known at any given time. Day to day changes in the population number are very small and even well ahead of the defining deadline, the precise population number can be well anticipated by the authorities of the municipality.[4]

If the agent can indeed ensure that she ends up just to the right of a certain threshold then the continuity assumption is likely to fail. Instead of valid counterfactuals, observations on either side are likely to differ. Variables that drive the selection will be systematically different on both sides and confound the treatment effect estimate.

[4]Caughey and Sekhon (2010) make the point that it is important to assess the magnitude of the random component in the score variable as compared to the precision with which agents can actively manipulate this variable.

One advantage of the RDD is that the sorting argument can be tested in a number of different ways. McCrary (2008) suggests a direct test of the distribution of the score variable. If it is profitable for agents to sort on a threshold and the possibility of doing so exists, then one should observe a higher frequency of observations on one side rather than on the other. Also, we can look at the distribution of predetermined variables which - given sufficient randomness - should not systematically differ around the thresholds.

Our last concern is a particularity of the use of population thresholds. Typically, a legal rule that applies a population threshold defines a definite point in time when the population count is taken. This point in time is often distinct from the actual implementation of the new rule. For example, for Bavarian municipalities council size is determined by a population count about a year before the council election.office.[5] This year, however, is also the time during which important institutional decisions are made. The municipal council, for example, decides whether the next mayor will be part- or full-time and / or whether the local authorities will be responsible for specific services.

These endogenous decisions can be problematic in two different ways. First, interpretation of the effect can be complex. Say, a municipality knows that it will increase the council size in one year. It must decide whether the future council will become responsible for certain tasks (e.g. water management). Their decision to do so might depend on the anticipated council size. A larger council might, for example, be able to support additional committees that can oversee these responsibilities. Those kinds of effects, however, are not included in the definition of the treatment effect, as treatment, *per se*, has not yet started. The researcher might be willing to redefine treatment, however, specific care has to be given to exactly what the object of interest is.

Apart from problems in interpretation, these endogenous decisions might also pose a threat to the validity of the identification, particularly in small samples. Assume that the simultaneous endogenous decision is in fact independent from the treatment. As with any postdetermined variable the researcher can only hope to not pick a strangely unbalanced sample in which the observations just right and just left are different by random selection. However, as opposed to other variables the timing of those endogenous institutional choices coincides exactly with the determination of treatment. This makes it much more likely that an unfavorable sample is picked. In Bavaria, for example, there is a trend to have more full-time mayors and to locate additional responsibilities at the local level. Naturally, these

[5] In preparation for a new election, the new council size must be known well ahead of time as certain preparations directly depend on it. For example, each party will name a number of candidates that is (typically) exactly as many as there are council seats.

decisions are made at exactly the same time as council size increases: just before the new election cycle.

3 Institutional setting

The German federal system comprises of four tiers. Apart from the federal, state and county level, some key decisions are taken by local municipalities. This local authority decides, among other issues, on local roads, theaters, cultural events, local business development, as well as school buildings and administers social welfare programs, kindergarten spending, etc. Furthermore, they often own and control key parts of the local economy such as waste disposal, public transport, as well as the energy and water supply. As the municipalities are the lowest tier of the federal system, most laws regulating their decision mechanisms, freedoms, and duties are exogenous to them: in Bavaria alone there are hundreds of state laws and by-laws that refer to municipalities.[6] As the state laws apply to all communities, the state legislature often uses population thresholds to adjust the rules to different requirements of smaller and larger communities.

We worked through the most relevant parts of the applicable state legislation to detect rules that use population thresholds. We found 14 (excluding the changes in council size) rules that cover nine of the ten thresholds determining the council size (see table 1 for an overview of the thresholds relevant to the council size and corresponding further rules).[7]

We found rules that define both local institutions and affect budget size. With regards to budgeting, we found five rules (see table 6 in the appendix for a detailed description).[8] For example, communities with more than 5,000 inhabitants receive 7.6 percent of the vehicle tax collected in their territory, while smaller communities do not receive anything - at the same threshold the council increases from 16 to 20 members. Another rule states that a city with more than 50,000 inhabitants may apply to become a county free city. If choosing to do so, the city takes over all the duties previously provided by the county and receives additional transfers accordingly. We conclude that larger cities receive more transfers and

[6] All laws and by-laws are available online at a website provided by the government of Bavaria: http://www.gesetze-bayern.de/jportal/portal/page/bsbayprod.psml (retrieved in March 2011)

[7] Given the amount of laws and bylaws that regulate municipal decision making, it is still possible that there are further undetected thresholds.

[8] Following the state development program (*Landesentwicklungsprogramm Bayern*, December 22, 2009 version) communities may also be grouped into one of five levels that indicate regional relevance for public services. While there are population thresholds among the criteria, these thresholds do not refer to the community itself but rather to the total population of the respective community plus the population of the communities it serves.

Table 1: Changes at council size population thresholds

	Population thresholds at # of inhabitants (in tsd)									
	1	2	3	5	10	20	30	50	100	200
	Panel 1: *Thresholds used in Egger and Koethenbuerger (2010)*									
Council size	x	x	x	x	x	x	x	x	x	x
Wage of elected civil servants					x		x	x	x	
	Panel 2: *Further thresholds defining local institutions*									
Wage of full-time mayors		x	x	x	x		x	x	x	
Wage of part-time mayors	x		x	x						
Full-time council members					x					
Petition for referendum					x	x	x	x	x	
Referendum quota								x	x	
City districts									x	
Open council					x				x	
Accounting committee				x						
Mayor status				x	x					
	Panel 3: *Further thresholds defining budgeting rules*									
County free city								x		
Status of larger city							x			
Vehicle Tax				x						
Fiscal equalization				x	x			x	x	

Notes: We have included only those rules and thresholds that correspond to the thresholds relevant for the council size. For a more detailed description the changes at each of those population thresholds see tables 6 and 7. *Source*: Own research.

provide more services than smaller cities and that the thresholds typically coincide with the community council size thresholds.

Regarding local institutions, there are even more rules that depend on population thresholds (we found nine, see table 7 in the appendix for details). These prescriptions range from relatively minor directives (e.g. that the council in cities with more that 5,000 inhabitants must include an accounting committee) to setting key rules of the local game. The latter include stipulations for the remuneration of the mayor[9] and a detailed regulation concerning local referendums. The requirements to bring a proposed referendum to the ballot are

[9]This is likely of significant importance for their performance and may hence affect government spending. Compare e.g. Besley (2004), Gagliarducci and Nannicini (2009), Messner and Polborn (2004), Ferraz and Finan (2009)

much higher in small communities (where 10 percent have to sign the petition) than in very large cities (where only three percent need to sign) with multiple steps in the signature requirement in between. The same general logic applies to the participation quota in the referendum itself too.[10]

We conclude that while the number of seats in the council is an important feature of local institutions, there are several other important local institutions that also change at the same population thresholds. This leads to double or multiple simultaneous treatment at any given threshold. The same applies when population thresholds also change the financial endowment of the community (see paragraph above).

Apart from setting population thresholds with budgetary and institutional relevance, the state laws also provide the communities with many choices in those dimensions. For example, the communities may decide to take over tasks from the county, which brings along new responsibilities, funds and administrative work. Tasks that communities may take over from higher government levels include: maintaining certain types of roads (compare *Finanzausgleichsgesetz*, Article 13a, version of 2010/06/03), construction supervision (compare Article 53, *Bayerische Bauordnung*, version of 2010/02/25), and waste disposal (compare Articles 5, 7 *Bayerisches Abfallwirtschaftsgesetz*, version of 2010/03/24). While some of the tasks that communities acquire from higher levels of government only induce small changes in revenue (e.g. construction supervision), others trigger large increases in revenues and spending (e.g. waste disposal).

Moreover, the municipality must regularly make decisions about institutional questions such as: Should we have a full-time mayor?, How many full-time working council members do we want? As mayors are elected at the same time that the council is elected and because additional responsibilities are also more likely to be taken over at the beginning of the electoral cycle it is probable that such decisions also coincide with increases in council size. The econometric implications of the resulting possible simultaneous endogenous decision making are discussed in section 2.

4 Results

Our empirical analysis generally builds on the strategy and data presented by Eggers and Koethenbuerger (2010).[11] Similar to their work, we use data on 2,056 municipalities from the German state of Bavaria over the 1983 to 2004 period. To investigate our specific

[10] e.g. Romer, Rosenthal, and Munley (1992) and Holcombe and Kenny (2008) show that referendum requirements affect spending.

[11] We obtained the data and Stata-dofiles from the authors through the journal website.

points of interest, we complement their data base with information on the revenue side of the municipality budget and the status of the town mayor (part-time or full-time).[12] The result section is split into three parts: (1) the importance of simultaneously determined finances, (2) the need to control for endogenous (but simultaneous) decisions of the municipality, and (3) the potential manipulation by local authorities of population numbers around thresholds.

4.1 Importance of simultaneous exogenous co-treatment

In the above section, we illustrate that legislative population thresholds in Bavaria not only affect council size but also a number of other important institutional features. Among those features are legislative rules that affect the budgeting of communities. Naturally, those direct fiscal consequences for the community are crucial in the identification of the effect of legislative size and governmental spending.

To understand the structure of municipality budgets in more detail, we illustrate some of the basic figures in table 8 (see appendix). The average municipality budget in Bavaria (for all 2056 municipalities in the period 1983-2004) reaches 1909 Euro per capita in total expenditures.[13] On the expenditure side, we highlight the shares of three major budget components that are (partly) in the discretion of the municipality: expenditures on personnel, on materials, as well as investments.[14]

Below, we present important categories from the revenue side. Under the full discretion of the local authority are three tax rates[15]: (1) property tax A on agricultural land, (2) property tax B on all non-agricultural property and (3) a trade tax on local businesses. In column (2) we highlight the share that each category has in the overall budget. We find that local taxes account for 20.6 percent of the total revenue with the major part of that income generated by the trade tax. Local property taxes contribute only about 4.4 percent to local finances.

Equally important are grants from other tiers of government and fees for communal services. We document that about 12.7 percent of revenues come from grants and 9.4 percent are

[12] This additional data is publicly available and can be obtained free of charge from the Bavarian statistical service.

[13] The same figure has to appear on the revenue side, as all expenditures have to be refinanced either by tax income, grants or increases in new debt.

[14] Other important expenditures that are not included are expenditures for debt repayments and expenditures for mandated social services. These are usually not directly in the control of the municipality.

[15] Only the tax rates are at the discretion of the municipalities - the rules defining the tax base are set by higher government levels.

raised through fees on local services.[16] It is important to note that the communities have only very limited decision power when it comes to influencing grants. Of course, they can lobby to receive grants and they have some discretion in setting fees, however, for the most part grants underlie legislative rules and many fees are regulated by the state.[17]

Table 2: Replication of the results by Egger and Koethenbuerger (2010)

	Outcomes (15 percent window)				
	Log Total Expenditures	Log Debt	Log Prop Tax Rate A	Log Prop Tax Rate B	Log Trade Tax Rate
	(1)	(2)	(3)	(4)	(5)
Treatment	0.109**	0.137*	0.055***	0.058***	0.009**
	(0.052)	(0.074)	(0.009)	(0.008)	(0.004)
N	22631	22162	22631	22631	22631
R2	0.00	0.00	0.00	0.00	0.00

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We use the program provided by Egger and Koethenbuerger (2010) to replicate their main results (see their table 3 and 6). Standard errors in parentheses are robust. The dependent variable are indicated above. Treatment is defined by being right to the population thresholds. All regressions include a third order polynomial in population that is flexible on both sides of the threshold. The estimation is done within the 15 percent window of population around the thresholds. *Source*: Own calculations based on the program provided by Egger and Koethenbuerger (2010).

Table 2 replicates the results reported by Egger and Koethenbuerger (2010).[18] Egger and Koethenbuerger estimate a version of eq. 2 for the thresholds at which population size changes and document the effect on total expenditures to be an immense 10.9 percent increase (see column 1, table 2) or 4.2 percent for every additional seat.[19] Given the average total expenditure (per capita) of 1909 Euro, the effect is hence argued to increase expenditures by the order of about 200 Euro (per capita and year). The size of this effect is to be evaluated even larger considering that only part of the local expenditures is effectively under local control.

Looking at the revenue side of the effect, Egger and Koethenbuerger (2010) report the estimates on debt, both types of property tax and the local trade tax (see columns 2-5 in table 2). They find sizable and significant effects primarily for the two property tax rates A and B of 5.5 percent and 5.8 percent increases respectively. They argue that municipalities

[16] Those fees are levied on services such as water supply, sewage and waste management, kindergartens, etc.

[17] The remaining part of the revenue comes from higher levels authorities and are raised, in part, through income tax and VAT. The community gets a fixed share of these revenues and has no control over the tax rate. Moreover, revenues can also be generated through new debt.

[18] We use the same dofiles and data and get the exact some results from their table 3 and table 6.

[19] The average increase in the number of seats at the thresholds is 2.56.

mainly rely on those sources of revenue to finance the additional expenditures (see Egger and Koethenbuerger (2010, p. 211)). However, given that the property taxes only account for 4.4 percent of total revenue, an increase by about 5.8 percent in the those taxes is likely to increase revenue by only 5 Euro (per capita and year).[20]

Table 3: The results of council size on state grants and municipality fees

	Outcome (15 percent window)		
	Log General Grants	Log Invest Grants	Log Muni Fees
	(1)	(2)	(3)
Treatment	0.089*	0.273***	0.236***
	(0.047)	(0.085)	(0.070)
N	21513	21383	21511
R2	0.00	0.00	0.00

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We use the program provided by Egger and Koethenbuerger (2010) to estimate the regressions using their setup, however, with the outcome variables that we have added to the data. Standard errors in parentheses are robust. The dependent variable are indicated above. Treatment is defined by being right to the population thresholds. All regressions include a third order polynomial in population that is flexible on both sides of the threshold. *Source*: Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).

Given the exploration on the institutional changes at population thresholds, we suspect that the additional spending is largely driven by automated changes in the rules that affect grants as well as the services communities provide (and hence fees collected). In table 3 we apply the design to estimate the increase in grants and fees at the thresholds. We show that there are substantial, significant increases both in grants (27.3 percent in additional investment grants) and communal fees (23.6 percent additional revenue). In total, the increases in grants and fees account for about 81 Euro per capita. Hence, we can explain a substantial part of the expenditure increase (although not the entire effect) by automated changes in revenues provided by higher levels of government to larger communities.

We document that the identification of the council size effect may suffer from co-treatment effects that are simultaneously determined by the same population thresholds. Without

[20]The effect that they report on the local trade tax is very small with less than 1 percent increase, which implies that this source of revenue does not play an important part in financing the expenditure increase. The point estimate on the debt is in fact quite large and could explain a substantial part of the expenditure increase. However, it is insignificant at the 5 percent level.

the detailed institutional knowledge and more specific data on all aspects of the revenue sources[21] it is impossible to distinguish between the "true" effect of legislative size and other changes at the same thresholds.

4.2 Simultaneous endogenous decision making

Causal analysis is further complicated by the fact that municipalities must make important decisions on the structure of the local institutions and the services provided by the community. It is in the nature of political cycles that the timing of those endogenous choices coincides with adjustments in council size. As explained above those decisions can have substantial consequences for municipal fiscal situations. We investigate one important feature of local institutions, namely whether the mayor is working part-time or full-time.

In figure 2 in the appendix we highlight the share of full-time mayors over the population distribution. The vertical lines indicate the population thresholds used in the council size analysis. Significant differences in the share of full-time majors can be observed around the 2000 and 3000 inhabitants thresholds.[22] To further investigate the differences, we present the distribution of full-time majors just around those two marks in figure 3 in the appendix. As indicated by the local kernel regression fitted onto the data, we observe substantial differences right of the cut-off points.

Moreover, we show that the differences in mayor status also prevail in the entire sample. Using the same estimation setup, we use the analysis to predict the mayor status (see table 4). If mayor status is independent of council size, the estimates should be insignificant from zero. Our results indicate that there is a sizable and significant effect of the population thresholds on the probability of choosing a full-time mayor.[23]

The above findings indicate that mayor status differs significantly between the observations just to the left and just to the right of the thresholds. The important question is how that affects the estimates of the council size effect. In table 5 we repeat the estimation including a dummy for each observation that takes the value one when the municipality employs a full-time mayor during that year. For comparison, we highlight the results from Egger and

[21] To rule out all automated changes in revenues / expenditures one would have to get detailed data on very specific grants that are given from the state or federal level to the communities.

[22] As described above, there are two legal thresholds at 5,000 and 10,000 inhabitants that prescribe consequences for the major status. However, those rules are non-binding and only suggestions. We find that the probability of choosing a full-time mayor does not seem to be altered systematically at those thresholds. Rather, the important changes are at thresholds below 5,000.

[23] We present the results within the 4 window sizes used by Egger and Koethenbuerger (2010). A window size of 15 implies that only municipalities within 15 percent of the population threshold are used in the analysis.

Table 4: The results of council size on predetermined mayor status

	Probability of having a full-time mayor			
Window size	15 Percent	20 Percent	25 Percent	30 Percent
	(1)	(2)	(3)	(4)
Treatment	0.068***	0.058**	0.051**	0.064***
	(0.026)	(0.023)	(0.020)	(0.019)
N	22611	29783	37177	44295
R2	0.00	0.00	0.00	0.00

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In this regression, we estimate the effect of council size on the mayor status defined as a dummy variable which takes the value 1 if the mayor is full-time employed. Standard errors in parentheses are robust. Treatment is defined by being right to the population thresholds. All regressions include a third order polynomial in population that is flexible on both sides of the threshold. In general, the setup is similar to the regressions run by Egger and Koethenbuerger (2010). *Source*: Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).

Koethenbuerger (2010) in panel 1 of 5. In the panels below we show the results controlling for mayor status (panel 2) and additionally controlling for year fixed effects (panel 3).

Including a control for mayor status, we find that the point estimates reported in Egger and Koethenbuerger (2010) drop significantly and are no longer statistically significant in the preferred specifications. The specification using a 15 percent window is shown to exhibit no more effects of the council size on governmental spending.[24] If the mayor status is a predetermined variable, this result would immediately suggest that the RDD is invalid. However, as argued above, the timing is specific. The mayor status is determined after the new council size is fixed, yet before this new council takes office. One could argue that it is in anticipation of the new council members that the old council decides to employ a full-time mayor and that this in turn increased spending. This would mean to change the interpretation of the effect drastically. However, we believe that the choice of the mayor's status is in fact unrelated to the future council size. Under this assumption, results imply that the RDD fails to identify the causal effect of an increase in council size.

Interestingly, we observe a large estimate for the mayor status dummy, implying a 124

[24]In table 9 in the appendix, we show that this is also true for the effect of council size on the disaggregate spending categories (referring to table 5 in Egger and Koethenbuerger (2010)). However, when we include the mayor dummy in the analysis on revenue sources, we can not reject their estimates for the local tax rates. In table (10) in the appendix, we show that the effects on property taxes and trade tax remain stable when we include the mayor dummy. As these categories are true choice variables of the community, they might reflect a different dynamic than the categories steered by choices of responsibilities and grants from other tiers.

Table 5: Main results - council size effect with controlling for mayor status

	Log Total Expenditures			
Window size	15 Percent	20 Percent	25 Percent	30 Percent
	(1)	(2)	(3)	(4)
	Panel 1 : Replication of Egger and Koethenbuerger (2010)			
Treatment	0.109**	0.121***	0.149***	0.202***
	(0.052)	(0.045)	(0.041)	(0.037)
N	22631	29803	37197	44324
R2	0.00	0.00	0.00	0.00
	Panel 2 : Controlling for mayor status			
Treatment	0.024	0.049	0.087***	0.124***
	(0.040)	(0.034)	(0.031)	(0.028)
Mayor status	1.240***	1.233***	1.227***	1.217***
	(0.010)	(0.009)	(0.008)	(0.007)
N	22611	29783	37177	44295
R2	0.42	0.42	0.41	0.41
	Panel 3 : Controlling for mayor status and year effects			
Treatment	-0.006	0.019	0.044	0.063***
	(0.034)	(0.030)	(0.027)	(0.024)
Mayor status	1.157***	1.146***	1.129***	1.118***
	(0.009)	(0.008)	(0.007)	(0.006)
N	22404	29483	36766	43809
R2	0.47	0.47	0.47	0.46

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are robust. The dependent variable is the log of total expenditures. Panel 1 replicates the results in Egger and Koethenbueger (2010) (see their table 3). In Panel 2, we control for the mayor status in the estimation by including a dummy variable that takes the value 1 if the mayor is full-time employed. Panel 3, additionally also controls for year fixed effects. In those last estimations we also excluded bigger communities that had the status of a county free city. Treatment is defined by being right to the population thresholds. All regressions include a third order polynomial in population that is flexible on both sides of the threshold. *Source*: Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).

percent increase in total spending when the mayor is full-time. We do not intend to interpret this to be a causal effect. Rather, we believe that is it very likely that this estimate reflects the importance of the complete choice set of the community at the time of elections. It is precisely when the council decides to take over more responsibilities from other tiers (like child care, water supply, etc.) that they also choose to have a full-time mayor that administers the local administration.

4.3 Manipulation of the population numbers around the thresholds

We show above that simultaneous co-treatment as well as simultaneous endogenous choices are posing a threat to the validity of the causal inference. As argued in the section on the empirical methodology, the important identifying assumption is the argument that municipalities can not sort around the population thresholds (continuity assumption). This is of specific concern here as the population figure is generally observable at all times and action can be taken to manipulate the precise figures before the deadlines in question. Municipalities could act to manipulate the statistical numbers directly within the administration or (more likely) they could start programs designed to attract new residents.[25] Given that the thresholds involve a multitude of consequences both for the political institutions and the fiscal budgeting of the municipality, the concern of manipulation needs to be taken seriously.

Egger and Koethenbuerger (2010) indicate the total number of observations just left (10914 obs) and just right (11690 obs) within the window of the limited sample.[26] However, it is standard praxis in applications of RDD to show a histogram of the frequencies just around the thresholds. In figure 1, we present such a histogram for the data used throughout this analysis (within the 15 percent window).

The graph indicates that the frequencies of observations just right and left of the thresholds differ systematically. There is a definite jump in the number of observations if one compares the two groups just below the threshold (about 700 obs each) compared to the one group just above (about 900 obs). The difference is further indicated by the gap in the local kernel regression fitted onto the data.[27] In figure 4 in the appendix, we investigate the frequency histogram for each of the thresholds individually. With the exception of the 2,000 threshold, the frequency tend to be always higher just to the right of the threshold. Of considerable difference are the jumps in the graphs for the 1,000, 3,000, 5,000 and 50,000 thresholds.[28]

This observation has direct consequences to the validity of the estimation design. If communities have the capability to manipulate their population figure at the margin, the continuity

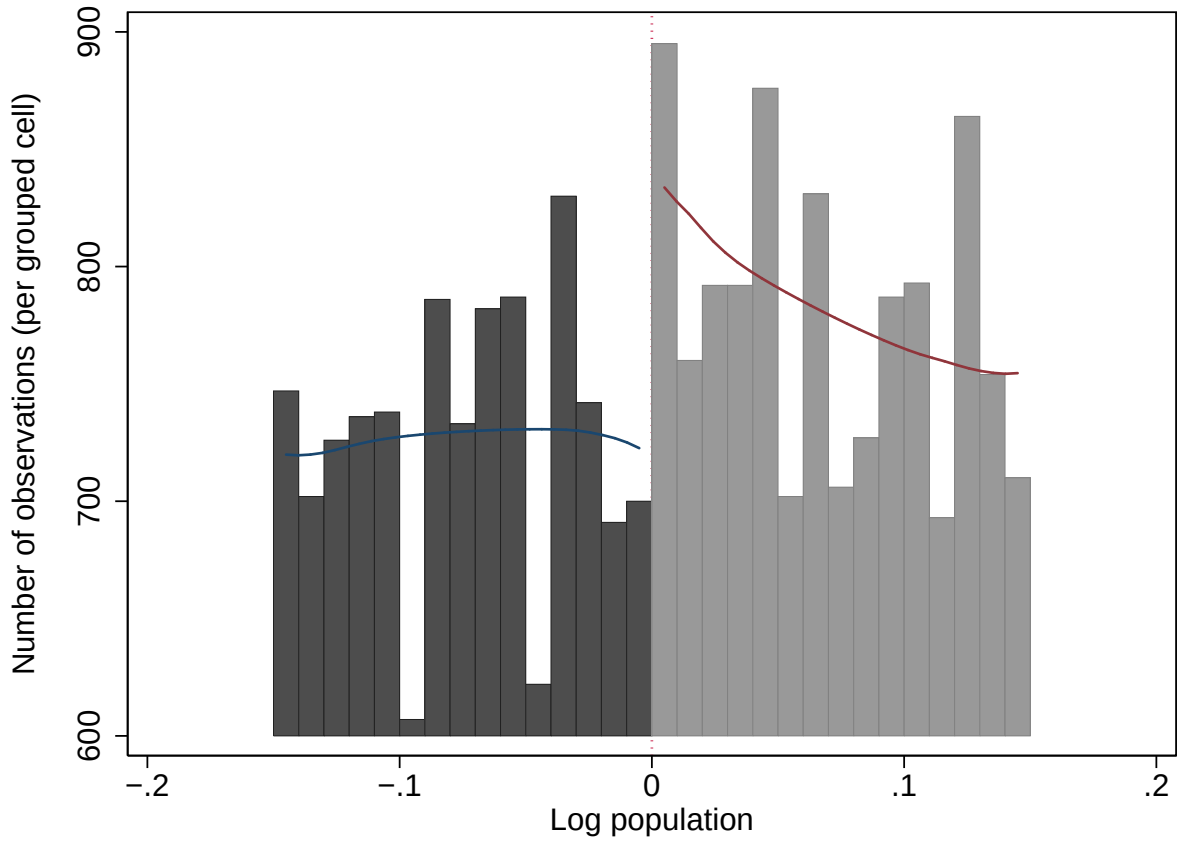
[25] Municipalities could, for example, open new community areas for housing projects, guarantee kindergarten spots to newcomers or give direct financial incentives to move to the town.

[26] Their results are presented in their table 3.

[27] We also ran flexible polynomial regressions on the binned data to obtain standard errors on the observed difference. As shown in the histogram, the data are quite variable between the bins. Nevertheless, the jump is significant at the 10 percent significance level when we run linear and quadratic specifications and at the 5 percent level using a third order polynomial that is flexible on either side.

[28] We also tested differences of important predetermined variables such as total expenditures the year before the election and mayor status during the last election period. However, we found no significant differences there, which implies that sorting is not along those dimensions.

Figure 1: RDD validity - frequency check



Notes: The figure shows the frequencies of observations in the analysis in grouped bins within the 15 % window (using 30 bins with a size of one percent each). The thresholds in the analysis have been normalized to zero. Again, as in Egger and Koethenbuerger (2010), the line fitted onto the data is based on a local kernel regression using Epanechnikov weights. A regression analysis using those bin averages and a flexible polynomial specification in the log of population shows that the difference at the threshold is both sizable and significant at least at the 10 percent level (at 5 percent level when one uses a third order polynomial specification). *Source:* Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).

assumption for the RDD does not necessarily hold and the estimates are potentially biased due to selection effects. It might be the fast growing communities that seek to take on more responsibilities which manage to end up just to the right of the thresholds.

5 Conclusions

In this paper, we identify and discuss three main challenges requiring careful attention when using population thresholds in regression discontinuity designs. First, the population threshold may define changes in multiple rules and not only the treatment considered. Second, when changes in population trigger the observed treatment it is likely that endogenous

institutional choices occur simultaneously. Third, political entities may seek to manipulate the official number of inhabitants knowing that it affects institutions at certain thresholds.

Revisiting Egger and Koethenbuerger (2010), we find evidence that our three concerns are of practical relevance. (1) The population thresholds used by the authors not only trigger changes in council size but also affect many other budgetary and institutional rules. Bigger communities have more responsibilities, receive more transfers and differ in institutions. We show that a large share of the spending increase at population thresholds relevant for the seat count stems from increases in grants form higher government levels. (2) We observe that when city councils grow in seats, the communities often decide endogenously to have their mayor work full instead of part time. When we include the information on the status of the mayor the spending effects of additional seats found by Egger and Koethenbuerger (2010) become insignificant. (3) We find evidence for manipulation of the population count, again indicating that the reported estimates are likely to be invalid.

From our results we find that it is crucially important to thoroughly verify the identifying assumptions of the RDD when using population thresholds. Taking into account the complexity of institutions and exogenous rules, notably at the local level, we recommend that researchers carefully document the legislative setting throughout the period of observation. In particular, they should discuss in detail whether population thresholds are used for other institutions (including the applicable laws and thresholds) and what other endogenous decisions might be taken simultaneously. The responsibility lies entirely on the researcher as the precise institutional setting is hard to assess for any outsider and near to impossible to judge from the pure data analysis. Further, the underlying continuity assumption needs to be discussed with great detail and all available tests have to be carried out such as to illustrate that the assumption is indeed supported by the data.

References

Acemoglu, D. (2005): "Constitutions, Politics, and Economics: A Review Essay on Persson and Tabellini's The Economic Effects of Constitutions," *Journal of Economic Literature*, 43, 1025–1048.

Baqir, R. (2002): "Districting and Government Overspending," *The Journal of Political Economy*, 110(6), 1318–1354.

Besley, T. (2004): "Joseph Schumpeter Lecture: Paying Politicians: Theory and Evidence," *Journal of the European Economic Association*, 2(2-3), 193–215.

Brollo, F., T. Nannicini, R. Perotti, and G. Tabellini (2009): "Federal Transfers, Corruption, and Political Selection: Evidence from Brazil," *University of Bocconi Working Paper*.

Caughey, D. M., and J. S. Sekhon (2010): "Regression-Discontinuity Designs and Popular Elections: Implications of Pro-Incumbent Bias in Close US House Races," Travers Department of Political Science, UC Berkeley.

Egger, P., and M. Koethenbuerger (2010): "Government Spending and Legislative Organization: Quasi-experimental Evidence from Germany," *American Economic Journal: Applied Economics*, 2(4), 200–212.

Ferraz, C., and F. Finan (2009): "Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance," *NBER Working Paper*, w14906.

Gagliarducci, S., and T. Nannicini (2009): "Do Better Paid Politicians Perform Better? Disentangling Incentives from Selection," *IZA Discussion Paper*, 4400.

Gilligan, T. W., and J. G. Matsusaka (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.

Hinnerich, B. T., and P. Pettersson-Lidbom (2010): "Democracy, Redistribution, and Political Participation: Evidence from Sweden 1919-1950," *Stockholm University Working Paper*.

Holcombe, R., and L. Kenny (2008): "Does restricting choice in referenda enable governments to spend more?," *Public Choice*, 136, 87–101.

Lee, D. S. (2008): "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, 142, 675 – 697.

Lee, D. S., and T. Lemieux (2009): "Regression Discontinuity Designs in Economics," Working Paper 14723, National Bureau of Economic Research.

Litschig, S., and K. Morrison (2010): "Government Spending and Re-election: Quasi-Experimental Evidence from Brazilian Municipalities," *University of Pompeu Fabra Working Paper*.

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142(2), 698 – 714.

Messner, M., and M. K. Polborn (2004): "Paying politicians," *Journal of Public Economics*, 88(12), 2423 – 2445.

Persson, T., and G. Tabellini (2002): *The economic effects of constitutions*. MIT Press.

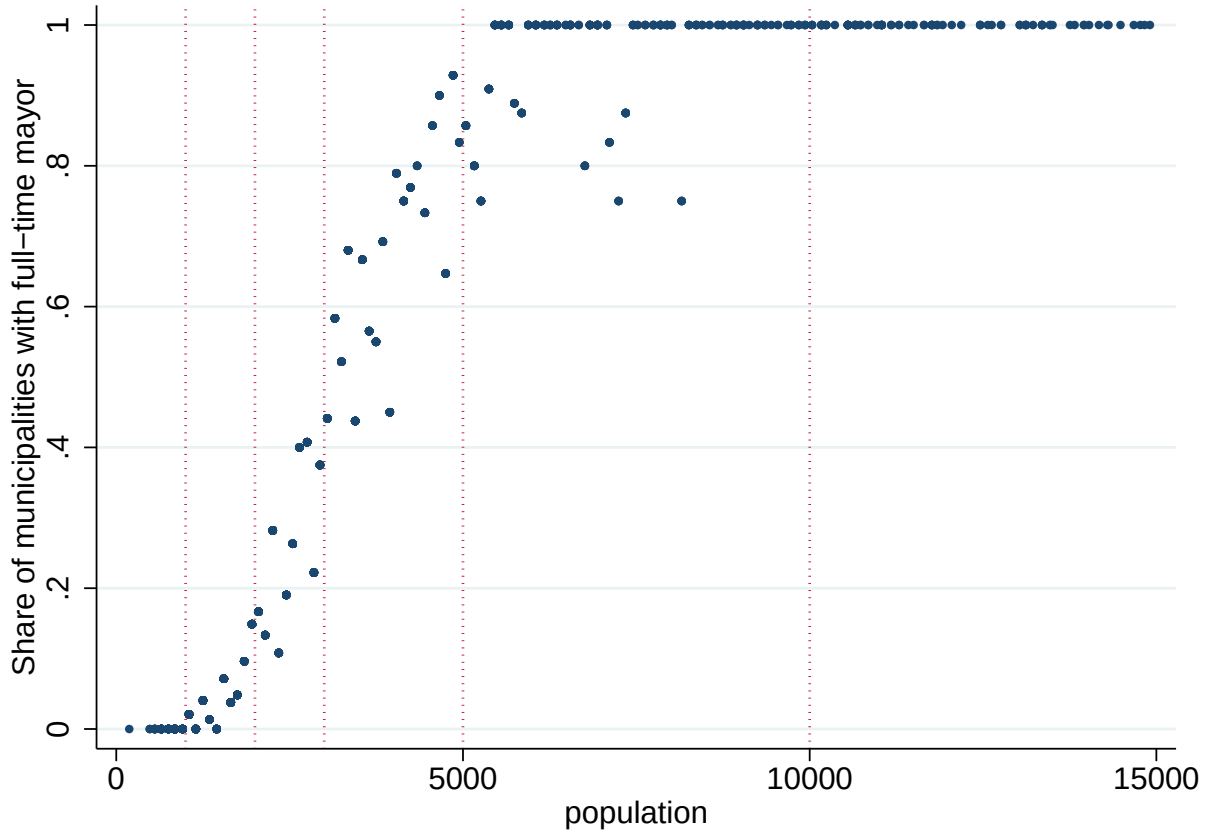
Pettersson-Lidbom, P. (2006): "Does the Size of the Legislature Affect the Size of Government? Evidence from Two Natural Experiments," *Stockholm University Working Paper*.

Romer, T., H. Rosenthal, and V. G. Munley (1992): "Economic incentives and political institutions: Spending and voting in school budget referenda," *Journal of Public Economics*, 49(1), 1 – 33.

Schaltegger, C. A., and L. P. Feld (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.

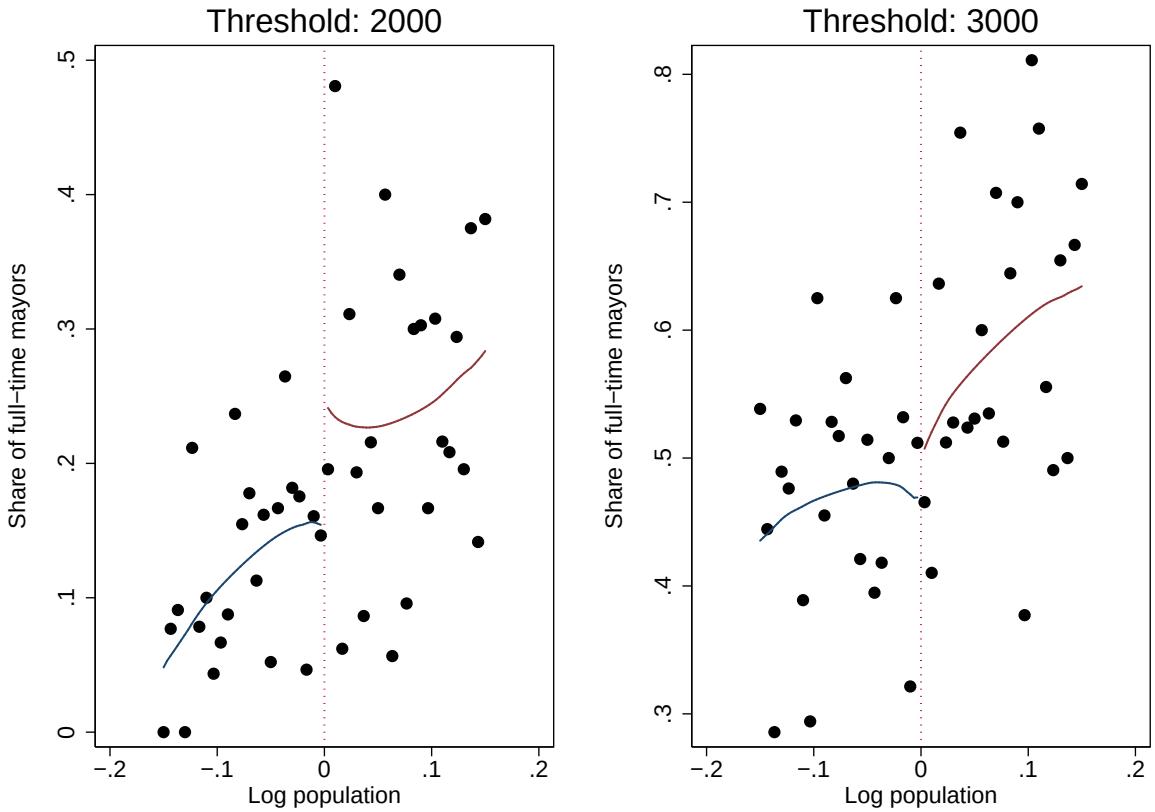
Appendix

Figure 2: Share of full-time mayors over the population distribution (in 1984)



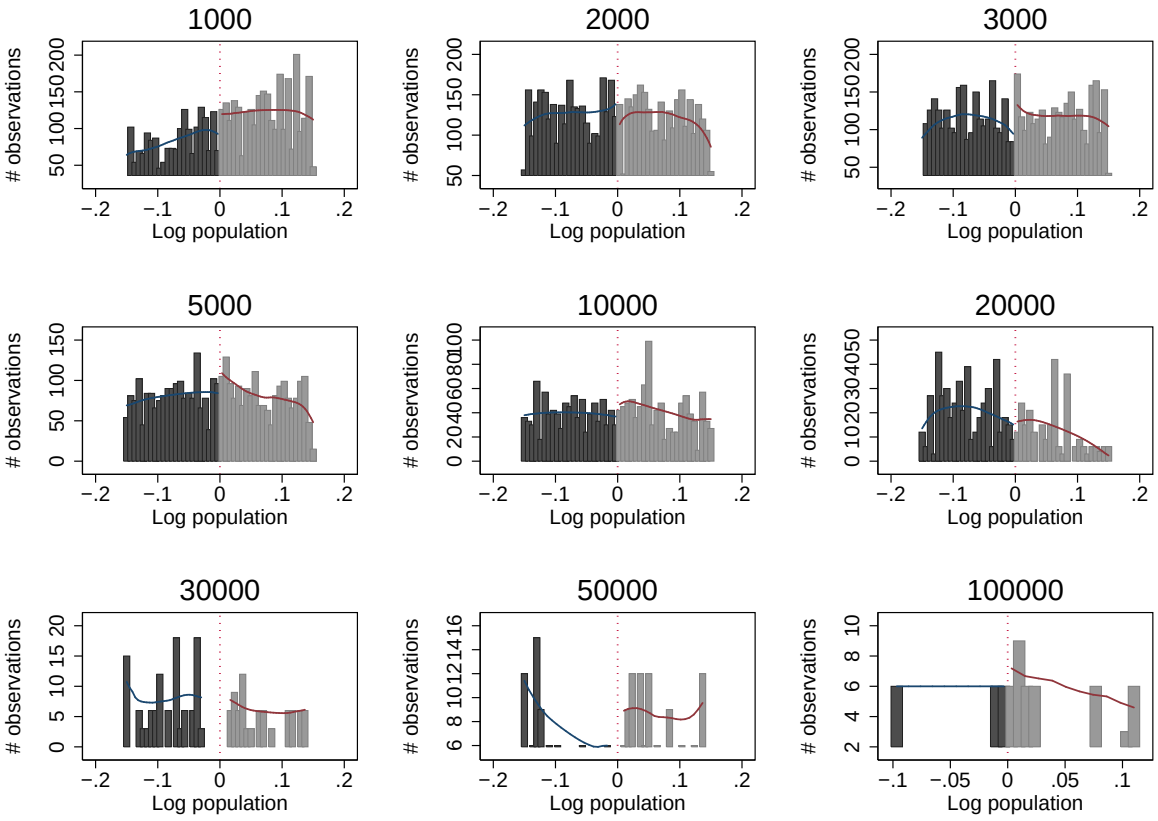
Notes: This figure illustrates the share of full-time mayors over the distribution of the population below 15000 inhabitants (there are no changes above). Each point represents the share within a bin (bandwidth equals 100 inhabitants) including all municipalities of that size. The vertical lines illustrate where the thresholds at with the council size changes are. *Source:* Own calculations.

Figure 3: Distribution of full-time mayors around thresholds



Notes: The figure investigates the share of full-time mayors just around the thresholds of 2000 and 3000 inhabitants. Similar to Egger and Koethenbuerger (2010), we present the results within the 15 % window (using 46 bins representing about 2/3 of one percent). The lines fitted onto the data is based on a local kernel regression using Epanechnikov weights. *Source:* Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).

Figure 4: RDD validity - frequency check



Notes: The figure shows the frequencies of observations at all thresholds below 200000 in grouped bins within the 15 % window (using 30 bins with a size of one percent each). The line fitted onto the data is based on a local kernel regression using Epanechnikov weights. *Source:* Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).

Table 6: Defining the budget: population thresholds in Bavarian law

Rule	Population thresholds	Legal source	Description
County free city	50,000	e.g. FAG Art. 2, 9, 12	Cities with more than 50,000 inhabitants can apply to become a county free city ('Kreisfreie Stadt'). That means that the city legally leaves the county it belonged to before and takes over all tasks that were executed by the county until then. These tasks include such important and costly assignments as social welfare, vocational schools, and car registration. The transfer of tasks also includes a transfer of the budget associated with these tasks. The additional funds that the council free city enjoys stem both from additional state transfers and a county levy that the city now no longer needs to pay.
Larger city	30,000	GO, Art. 5a	Communities with more than 30,000 inhabitants can apply for the larger city status ('Große Kreisstadt'). Once awarded the larger city status the community takes over some of the duties that are usually carried out by the county to which the community belongs. The additional tasks include construction supervision and road traffic authority. Some of these tasks also create additional revenue for the cities (e.g. fees of the construction supervision authority).
Vehicle Tax	5,000	FAG, Art. 13a	Communities with more than 5,000 inhabitants receive 7.6 percent of the vehicle tax collected on their territory, the rest of the tax revenue remains with the state government. Smaller communities do not receive a share of the vehicle tax. Whenever communities - regardless of their size - take over the responsibilities for certain types of roads they receive an even larger share of the vehicle tax.
Water management	90,000; 143,750; 300,000; 600,000	FAG, Art. 9	County free cities are in charge of managing all issues related to water in their city. To help them cover the associated cost county free cities receive a per capita allowance of 80 cents per inhabitant but not more than 115,000 EUR which is equivalent to to a population threshold of 143,750. Beyond they also receive a supplemental lump sum allowance that depends on their size. This lump-sum is 25,000 EUR for cities below 90,000 inhabitants, 35,000 EUR for cities with more than 90,000 and up to 300,000 inhabitants, 50,000 EUR for cities with more than 300,000 and up to 600,000 inhabitants and 100,000 EUR for all cities with more than 600,000 inhabitants.
Fiscal equalization	5,000; 10,000; 25,000; 50,000; 100,000; 250,000; 500,000	FAG, Art. 4	Next to own revenue the communities in Bavaria rely on transfers from the state government to finance their expenditures. One of the most important transfers is the so called 'key allocation' (Schluesselzuweisung). The larger the community is the more it receives in key transfers per capita. Communities with more than 5,000 inhabitants receive a per capita bonus for every additional inhabitant. The increase is 8 percent per inhabitants once the community reaches 5,000 inhabitants and is then a linear function on the population with several points a which the slope changes. For example at exactly 50,000 inhabitants the bonus is 35 percent compared to a community with less than 5,000 inhabitants. The population thresholds mentioned in this tables do not change the level of the key transfers discontinuously, however, the slope of the key transfer schedule (as a function of population size) changes. This change in the slope might in turn induce discontinuous changes in the optimizing strategy of the economic agent.

Legal sources: FAG - Finanzausgleichsgesetz fuer Bayern (version: June 2, 2010), GO - Gemeindeordnung für den Freistaat Bayern (version: August 17, 2009). *Source*: Own research.

Table 7: Defining local institutions: population thresholds in Bavarian law

Rule	Population thresholds	Legal source	Description
Full time council members	10,000	GO Art. 40	In communities with more than 10,000 inhabitants the council can elect additional full time council members that take over management functions in the local administration.
Wage elected civil servants [a]	2,000; 3,000; 5,000; 10,000; 15,000; 30,000; 50,000; 100,000.	BKBV, Art. 1; KWBG, supp. 1	The full-time mayor, the full-time deputy mayor(s), and the full time council members have the status of elected civil servants. Their remuneration depends on the size of the community. At each of the threshold indicated their salaries increase by one or two steps in the civil servants salary scale. For example in ordinary communities with more than 15,000 and up to 30,000 inhabitants the mayor is at the level B2 or B3 while in communities with more than 30,000 inhabitants she is at the level B3 or B4. An increase in one step in this system (from B3 to B4) implies about six percent wage increase (or EUR 400 per month) in Bavaria in 2010. Furthermore, elected civil servants receive a monthly expense allowance that depends on the population if the city is either a community free city or has the large city status. Then the allowance thresholds are 50,000, and 100,000 inhabitants. For example the mayor receives between 305.68 and 833.76 EUR in such communities with less than 50,000 inhabitants and between 437.72 and 965.79 EUR in communities with more than 50,000 but less than 100,000 inhabitants.
Cost allowance part time mayors	1,000; 3,000; 5,000	GO, supp. 1	Part time mayors receive a cost allowance that depends on the population size. The law sets overlapping allowance ranges for the different community size classes. For examples mayors in communities with up to 1,000 inhabitants receive a monthly allowance between 430.48 and 1,907.88 EUR and while this amount increases to between 1,830.10 and 3,307.51 EUR in the size class above.
Petition for referendum	10,000; 20,000; 30,000; 50,000; 100,000; 500,000	GO, Art. 18a	The inhabitants in a community can request a referendum. To be considered, the organizers of the request need to gather signatures from the inhabitants. What share of the inhabitants they need to sign depends on the size of the community. Ten percent have to sign if the community has up to 10,000 inhabitants, 9 percent up to 20,000, 8 percent up to 30,000, 7 percent up to 50,000, 5 percent up to 100,000 and 3 percent have to sign if the community has more than 500,000 inhabitants.
Referendum quota	50,000; 100,000	GO, Art. 18a	Once a petition for referendum has reached the necessary number of signatures the voters are called to the ballot. Then the referendum needs to fulfill two conditions to pass. First, it has to get more then fifty percent of the votes. Second, these votes must represent at least 20 percent of the eligible electorate in communities with up to 50,000 inhabitants. In communities with up to 100,000 inhabitants the supporting votes need to represent at least 15 percent of the electorate and in communities with more than 100,000 inhabitants this share has to be at least 10 percent.
City districts	100,000	GO, Art. 60	Communities with more than 100,000 inhabitants are subdivided into several city districts. These city districts take over part of the administration of the city. The city council can decide to hand over decision powers to the districts. In this case they also have a district councils whose members are elected.
Open council	10,000; 100,000	GO, Art. 18a	In an open council ('Buergerversammlung') the constituents meet to discuss topics of local relevance. The meetings are chaired by the local administration. The open council can pass recommendations that then have to be discussed in the community council. In communities with less than 10,000 inhabitants five percent of the voters have to request the open council meeting, in communities with more than 10,000 inhabitants only two and a half percent need to request the open council. In cities with more than 100,000 inhabitants open council can also be held within a city district.
Accounting committee	5,000	GO, Art. 103	In communities with more than 5,000 inhabitants the council has to establish an accounting committee composed of council members to conduct an examination of accounts.
Mayor status	5,000; 10,000	GO, Art. 34	Communities with more than 10,000 inhabitants must have full-time mayors. Communities with more than 5,000 but not more than 10,000 inhabitants have a full time mayor by default. However, they can choose to have a part-time mayor instead if the council decides this not later than 67 days before the mayor election. Communities with 5,000 inhabitants or less are governed by part-time mayors by default. However, they can also choose to have a full-time mayor if the council decides so not later than 67 days before the mayor election.

Legal sources: BKBV - Bayerische Kommunalbesoldungsverordnung (version: May 29, 2009), FAG - Finanzausgleichsgesetz fuer Bayern (version: June 2, 2010), GO - Gemeindeordnung für den Freistaat Bayern (version: August 17, 2009), KWBG - Gesetz ueber kommunale Wahlbeamte (version: August 17, 2009) *Source*: Own research.

a: Egger and Koethenbuerger (2010) state that the remuneration of the full time council members depends on the population thresholds. However, this also applies to mayors and deputy mayors. As full time council members can only be elected in communities with more than 10,000 inhabitants the authors conduct a robustness that only includes communities with less than 10,000 inhabitants to make sure that remuneration steps for full time council members do not play a role for identification.

Table 8: Communal budgeting - shares of expenditure and revenues

	Mean	Share in %
Expenditures (Euro per capita)		
Total	1909.5	100.0
on Personnel	369.6	19.4
on Materials	251.2	13.2
in Investment	368.3	19.3
Revenues (Euro per capita)		
Total	1909.5	100.0
from Property Tax A	6.3	0.3
from Property Tax B	77.9	4.1
from Trade Tax	310.1	16.2
from General Grants	146.4	7.7
from Investment Grants	96.3	5.0
from Fees	178.6	9.4
Average debt (Euro per capita)	788.4	

Notes: This table illustrates basic figures of an average municipality budget. Column 2 presents the population weighted average number of Euro per capita in each category over all 2056 municipalities during the period of observation (1984-2004). Column 2 highlights the share that the individual item has on total expenditures or revenues respectively. The categories named are not exclusive, rather, we refer to all categories used in the analysis. *Source*: Own calculations.

Table 9: The results on expenditure categories controlling for mayor status

	Expenditure Categories (Window size 15 percent)		
	Invest Expend	Material Expend	Personnel Expend
	(1)	(2)	(3)
	Panel 1 : Replication of Egger and Koethenbuerger (2010)		
Treatment	0.103*	0.143**	0.169***
	(0.061)	(0.059)	(0.059)
N	22623	22626	22631
R2	0.001	0.001	0.001
	Panel 2 : Controlling for mayor status and year effects		
Treatment	-0.047	0.026	0.052
	(0.048)	(0.039)	(0.038)
Mayor status	1.157***	1.373***	1.367***
	(0.012)	(0.010)	(0.010)
N	22396	22399	22404
R2	0.310	0.488	0.505

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are robust. The dependent variables are indicates above. Panel 1 replicates the results in Egger and Koethenbueger (2010) (see their table 5). In Panel 2, we control for the mayor status in the estimation by including a dummy variable that takes the value 1 if the mayor is full-time employed. Additionally, we also controls for year fixed effects and we excluded bigger communities that had the status of a county free city. Treatment is defined by being right to the population thresholds. All regressions include a third order polynomial in population that is flexible on both sides of the threshold. *Source*: Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).

Table 10: The results on revenues controlling for mayor status

	Revenues in Logs (Window size 15 percent)					
	Debt	Prop Tax A	Prop Tax B	Trade Tax	Gen. Grants	Invest Grants
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.031	0.066***	0.066***	0.009**	0.136*	0.106**
	(0.060)	(0.009)	(0.008)	(0.004)	(0.077)	(0.049)
N	21935	22404	22404	22404	21186	21314
R2	0.256	0.050	0.052	0.019	0.181	0.454

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variables are indicates above. In this regression, we estimate the effect of council size on the mayor status defined as a dummy variable which takes the value 1 if the mayor is full-time employed. Standard errors in parentheses are robust. Treatment is defined by being right to the population thresholds. All regressions include a third order polynomial in population that is flexible on both sides of the threshold. In general, the setup is similar to the regressions run by Egger and Koethenbuerger (2010). *Source*: Own calculations partly based on the program provided by Egger and Koethenbuerger (2010).