

Sieger, Philip

Conference Paper

Incumbency Effects in Germany: Federal and Mayoral Elections

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2012: Neue Wege und Herausforderungen für den Arbeitsmarkt des 21. Jahrhunderts - Session: Microeconometrics and Statistics, No. D09-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Sieger, Philip (2012) : Incumbency Effects in Germany: Federal and Mayoral Elections, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2012: Neue Wege und Herausforderungen für den Arbeitsmarkt des 21. Jahrhunderts - Session: Microeconometrics and Statistics, No. D09-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/62084>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Incumbency Effects in Germany: Federal and Mayoral Elections

Philip Sieger†

September 2011

Abstract:

In this paper incumbency effects in Federal Elections and Mayoral Elections in Germany are estimated using a quasi-experimental research design which allows for causal inference under a set of rather mild assumptions. Relying on a nonparametric estimation procedure and exploiting a recently developed bandwidth selection criterion, incumbency effects for the two main parties in Germany cause an increase in vote share of 1.4-1.7% in Federal Elections. Analyzing Mayoral Elections, the causal effect of incumbency is about ten times larger with an increase in vote share of 14-17%. Both results are robust with respect to the inclusion of further covariates. These huge differences might possibly be explained by the differences in visibility and popularity of mayors compared to candidates in Federal Elections.

Key Words: Regression Discontinuity; Incumbency; Average treatment effect

JEL classification: C14, C21, C52, D72

† Goethe University Frankfurt am Main, Faculty of Economics and Business Administration, Department of Applied Econometrics, Chair of Econometrics, Grueneburgplatz 1, RuW/PF 48, 60323 D-Frankfurt am Main, Germany, Tel: +49 69 798 34721, Email: sieger@wiwi.uni-frankfurt.de

1. Introduction

The interaction of politicians and voters can be understood as a principle-agent problem in which political incumbents try to increase their chance of re-election by the way they act while in office. They are able to influence, for example, tax and expenditure policy, use the office to sell political favors to campaign donors or generally vote on legislation in a way that pleases their constituencies (Lee, 2001). And it is in fact the case that in 90 percent of the time when Representatives in the U.S House tried to be re-elected they were successful in doing so (Jacobson, 1997). There is, however, the possibility that incumbents enjoy re-election success for different reasons apart from their incumbency status (Lee, 2001). It may simply reflect the systematic differences in electoral districts. In a left-leaning district a Democrat is simply more likely to win in any given election than a Republican, and the same also holds true for the subsequent election. This means that persistent heterogeneity across districts could generate the observed 90 percent incumbent re-election rate even if there is no structural advantage to incumbency. To disentangle the causal effect from incumbency from these potentially confounding effects it has become common practice to exploit the special nature of elections in a Regression Discontinuity Design (RDD from here on). Current contributions using this approach are Lee (2001, 2008) and Lee, Moretti and Butler (2004). These studies primarily focused on the incumbency advantage inherit in elections to the United States House of Representatives. Much less attention has been dedicated to Germany with Hainmüller and Kern (2008) being a noticeable exception. In addition, most studies focus on incumbency effects in Federal Election while elections at the local level have rather been neglected. This paper tries to fill this gap by taking a deeper look at elections in Germany, contrasting the results of Federal Elections to the one obtained looking at Mayoral Election. By exploiting the quasi-experimental design of elections I estimate incumbency effects for the two main parties in Germany for Federal and Mayoral Elections. The estimation is done via local linear regression using a recently developed bandwidth selection criterion. The results show that incumbency increases vote shares by about 1.4-1.7% in Federal Elections. Although this coefficient is rather small in magnitude, especially compared to corresponding numbers found for the United States, it is nonetheless economically important since many districts are won by a very small margin. This is why incumbency raises the probability of winning the next election by about 20% in Federal Elections. Contrast these results to the one obtained by looking at Mayoral Elections: in this case incumbency causes an increase in vote share which is about ten times larger in magnitude than the effect in Federal Elections. The estimated incumbency effect is between 14-17%, which translates into an increased probability of

winning the next election of about 60%. The robustness of the results is reassessed by including further covariates as far as they are available and by estimating the incumbency effect based on a parametric approach. The reason for these large differences in incumbency effects is not clear without ambiguity. One explanation might be the difference in visibility and popularity of mayors compared to candidates in Federal Elections. While in Federal Elections any single candidate's result is not as important as the result in the aggregate (people might not care as much whether the Senate seats of Iowa are won by a Democrat or Republican as they care about which party won the majority of Senate seats), this logic does not apply to Mayoral Elections. In mayoral elections it is really the single candidate's result that matters to the electorate. In addition, candidates in Federal Elections (e.g. candidates for the House of Representatives) are usually not as well known to the public as are mayors, implying a larger potential for incumbency effects in Mayoral Elections compared to Federal Elections.

The structure of the paper is as follows: section 2 discusses the econometric methodology and the data, in section 3 results and robustness checks are presented. Section 4 concludes.

2. Econometric Methodology and Data

In this section I outline the econometric methodology used in the empirical analysis and describe the data.

2.1. Econometric Methodology

Since its first appearance in the scientific literature in 1960 (Thistlethwaite and Campell, 1960), RDD as a method of identifying causal effects has gone a long way (for a historical review of RDD see for example Cook, 2008). Since its use was thought to be of limited applicability, it has not attracted much attention in economics until the mid 90s. In a recent paper Lee and Lemieux (2010) provide a comprehensive list of studies using the RD Design; the applications vary from Education (Angrist and Lavy, 1999, Hoxby, 2000 and Van der Klaauw, 2002) over Health (Ludwig and Miller 2007) and Crime (Lee and McCrary 2005) to Political Economy (Lee, 2001 and 2008, Lee, Moretti and Butler 2004, Hainmueller and Kern, 2008). The rising popularity of this approach can be explained by the fact that it allows identification of causal effects of interventions or treatments under a set of rather weak assumptions (Hahn, Todd and Van de Klaauw 2001).

2.1.1. Regression Discontinuity Design

The RD Design is best explained within the Rubin Causal Model with potential outcomes (Rubin, 1974). In order to evaluate a certain program one would like to compare the outcome of interest Y_i of certain individuals (labeled from $i = 1, \dots, N$), namely the outcomes of those who participated in the program to the outcomes of those who did not. Denoting the potential outcome of person i if she participated in the program by $Y_i(1)$ and the potential outcome if she did not participate by $Y_i(0)$, one can express the observed outcome of person i , Y_i , as:

$$(1) \quad Y_i = (1 - D_i) * Y_i(0) + D_i * Y_i(1) = \begin{cases} Y_i(0) & \text{if } D_i = 0 \\ Y_i(1) & \text{if } D_i = 1 \end{cases}$$

with D_i being an indicator for participation in the program. The indicator for participation can only take two distinct values (since there is either participation, $D_i = 1$, or non-participation, $D_i = 0$), but it can itself depend on what is called an assignment (or sometimes forcing or running) variable X_i .

In the RDD participation of person i is switched on as X_i exceeds a certain known threshold or cutoff c so that one can write $D_i = \mathbb{I}\{X_i \geq c\}$, where \mathbb{I} stands for the indicator function¹. Individuals with a value of the assignment variable greater than or equal to the cutoff value do receive treatment while the others do not. It is, however, important to stress that the existence of a treatment being a discontinuous function of the assignment variable is not sufficient to justify the validity of the RD Design (Lee and Lemieux, 2010). In the RDD it is also necessary that individuals only have (at most) *imprecise* control over the assignment variable². To rephrase the identification strategy, close to the cutoff participation is “as good as randomized” (Lee, 2008), therefore unconfounded with observables and unobservables. The randomization follows directly from the inability to *precisely* control the assignment variable close to the cutoff.

To formulate necessary assumptions and resulting theorems, Lee (2008) introduces the following data generating process:

- 1) Randomly draw an individual from a population of individuals.

¹ The focus of this paper is exclusively on the so called Sharp RD Design, where participation is a deterministic function of the assignment variable. In the so called Fuzzy RD Design the probability of participation does not jump from zero to one at the cutoff but from a to b with $a < b$ and a and b strictly between zero and one.

² It is easy to imagine a treatment assignment mechanism in which individuals do not have any control over the assignment variable. This desirable situation is, however, not given in most applications and it is not a prerequisite for the RDD to be applicable.

- 2) Assign a score X , drawn from a non-degenerate, sufficiently “smooth” individual-specific probability distribution.
- 3) Assign treatment status based on the rule $D = \mathbb{I}\{X \geq 0\}$ (where Lee assumes without loss of generality that the cutoff is at 0).
- 4) Measure all variables, including the outcome of interest Y and any predetermined variables Z .

Based on this data generating process, the following two assumptions are needed:

Assumption A1: Let (W, X) be a pair of random variables (with W being the unobserved “type” and X being the observed assignment variable) and let $Y \equiv y(W, X)$ and $Z \equiv z(W)$, where for each w , $y(\cdot, \cdot)$ is continuous in the second argument except at $X = 0$, where the function is only continuous from the left. Define the functions $y^-(w) = \lim_{\varepsilon \nearrow 0} y(w, \varepsilon)$ and $y^+(w) = y(w, 0)$.

Assumption A2: $F(x|w)$, the cdf of X conditional on W , is such that $0 < F(0|w) < 1$, and is continuously differentiable in x at $x = 0$, for each w in the support of W . Let $f(\cdot)$ and $f(\cdot | \cdot)$ be the marginal density of X and the density of X conditional on W , respectively, with $f(0) > 0$.

According to these two assumptions, individuals can take action to influence their probability of receiving the treatment, but there is sufficient random variation to ensure that for each type the probability is strictly between zero and one ($0 < F(0|w) < 1$). In addition, the probabilities of getting a draw just above or just below the cutoff are the same ($F(x|w)$ is continuously differentiable in x at $x = 0$). As Lee points out, assumption A2 allows for arbitrary correlation in the population between X and any of the functions Y_1 , Y_0 or Z .

The theorem which builds on these assumptions reads as follows:

Theorem 1: If assumption A1 and A2 are fulfilled, then it holds true that

- a) $\mathbb{P}(W \leq w | X = x)$ is continuous in x at $x = 0$, for all w .
- b) $\mathbb{E}(Y | X = 0) - \lim_{\Delta \rightarrow 0^-} \mathbb{E}(Y | X = \Delta) = \mathbb{E}(Y_1 - Y_0 | X = 0) = \int_{-\infty}^{\infty} (y^+(w) - y^-(w)) \frac{f(0|w)}{f(0)} dG(w) = ATE^{**}$
- c) $\mathbb{P}(Z \leq z_0 | X = x)$ is continuous in x at $x = 0$, for all z_0 .

To rephrase the statements of Theorem 1, a) establishes that receiving treatment at the cutoff is “as good as randomized”, b) establishes that what is identified by the discontinuity gap in

the conditional expectation function is a weighted average treatment effect, where the weights are the probability that an individual draws an X “close” to 0, and c) establishes that all observed pre-determined covariates should be balanced at the cutoff.

One important question is which modeling assumption one needs to impose in order to express that individuals only have imprecise control over the assignment variable. One reasonable way to do so is to assume that the observed outcome of the assignment variable is decomposable into two components: $X_i = \eta_i + \varepsilon_i$, where η_i can be influenced by the individual, whereas ε_i is a random variable with continuous density. The question whether this is a reasonable description of the individual assignment processes has to be tackled by logical reasoning. For the case of incumbency effects, politicians surely can influence their vote share a great deal by the amount of money they spend on their election campaign or the number of public appearances. But still they can never be sure to win the election, as there remain factors which influence the candidate’s vote share while not being under his or her control, as weather on Election Day or personal scandals being exposed. Therefore it is reasonable to assume that the margin of victory, i.e. the difference in own vote share and runner-up’s vote share if the candidate wins or the difference in own vote share and the winner’s vote share if the candidate loses respectively, can only be imprecisely manipulated.

In order to measure the effect of incumbency to a party’s electoral success one would ideally change the incumbent party in one district exogenously from party A to party B, while holding all other factors fixed. The increase in party’s Y vote share associated with this change would then represent the electoral benefit due to being the incumbent party. Of course it is impossible to conduct such kind of analysis. Instead one has to rely on econometric methods in order to estimate the effect of incumbency.

In order to do so consider the following reduced form specification of electoral success incorporating the potential electoral gains from incumbency (Lee, 2008):

$$(2) VS_{i,t+1} = \alpha W_{i,t} + \beta MV_{i,t} + \tau D_{i,t} + \varepsilon_{i,t}$$

$$(3) D_{i,t} = \mathbb{I}\{MV_{i,t} \geq 0\}$$

$$(4) f_{i,t}(MV|W) \text{ is continuous in } MV$$

$$(5) \mathbb{E}(\varepsilon_{i,t} | W_{i,t}, MV_{i,t}) = 0$$

In this model, $VS_{i,t+1}$ is the vote share for party A’s candidate in district or city i in election held at time $t+1$. $D_{i,t}$ is a dummy variable reflecting whether party A is the incumbent party in

the electoral race between time t and time $t+1$. $W_{i,t}$ can be thought of as capturing characteristics determined as of election day in year t , for example party resources or the quality of the candidate. Note that the assignment variable $MV_{i,t}$, the margin of victory in the t election, is also allowed itself to have an impact on $VS_{i,t+1}$. To allow for such an effect seems reasonable since a large margin of victory in the previous election might have an influence on the vote share a candidate receives this election, for example by attracting more campaign donors and thereby boosting the vote share in the current election (Lee, 2008).

In addition to considering the vote share a party receives in the election in $t+1$ it is also possible to consider the probability of winning the next election as a measure of electoral success. This is a binary variable which equals 1 if a party wins the election in $t+1$ and 0 otherwise. The estimation techniques used in this context are identical to the ones used when considering vote share in $t+1$ as a measure of electoral success. So what is estimated in this case is in fact a linear probability model.

The main issue with the model (2)-(4) is that $W_{i,t}$ might be either unobserved or unavailable to the econometrician. It is well known that an omitted variable would lead to inconsistent estimates of the causal effect, given $W_{i,t}$ and $MV_{i,t}$ are correlated (because $D_{i,t}$ and $W_{i,t}$ would then be correlated as well). This will likely be the case, since general party affiliation or the quality of the candidate (which would be typical elements of $W_{i,t}$) will have an impact on the party's vote share and therefore on the margin of victory.

There is, however, hope to still estimate the causal effect of incumbency consistently, since assignment to treatment (becoming the incumbent or not) depends in a deterministic way on the observed assignment variable $MV_{i,t}$. In addition it is reasonable to assume that candidates only have imprecise control over their vote share and hence over their margin of victory conditional on unobservable characteristics. These two points together render it possible to apply Theorem 1 and to use the estimation techniques and selection procedures outlined in section 2.1.2 (for further details, see again Lee, 2008).

2.1.2. Estimation and bandwidth choice

Two frequently applied estimation techniques used in RDD are local linear regression and polynomial modeling. These two estimation techniques will be described in the next subsection, followed by an analysis of respective selection procedures.

In the case of local linear regression the causal effect is estimated the following way (Imbens and Lemieux, 2008): define $m(x) = \mathbb{E}(Y_i|X_i = x)$, then the causal effect τ is the difference $\mu^+ - \mu^-$, where $\mu^+ = \lim_{x \searrow c} m(x)$ and $\mu^- = \lim_{x \nearrow c} m(x)$. Estimation of $m(x)$ is accomplished via

$$(6) \quad \hat{m}_h(x) = \begin{cases} \hat{\alpha}^-(x) & \text{if } x < c \\ \hat{\alpha}^+(x) & \text{if } x \geq c \end{cases}$$

where $\hat{\alpha}^-(x)$ and $\hat{\alpha}^+(x)$ solve the following equations:

$$(7) \quad \left(\hat{\alpha}^-(x), \hat{\beta}^-(x) \right) = \underset{\alpha, \beta}{\operatorname{argmin}} \sum_{i=1}^{N_-} \mathbb{I}_{\{X_i < x\}} (Y_i - \alpha - \beta(X_i - x))^2 K\left(\frac{X_i - x}{h}\right)$$

$$(8) \quad \left(\hat{\alpha}^+(x), \hat{\beta}^+(x) \right) = \underset{\alpha, \beta}{\operatorname{argmin}} \sum_{i=1}^{N_+} \mathbb{I}_{\{X_i \geq x\}} (Y_i - \alpha - \beta(X_i - x))^2 K\left(\frac{X_i - x}{h}\right)$$

K is the kernel function that weighs observations relative to their importance. The kernel function used in the estimation is the edge kernel. The use of the edge kernel stems from the fact that one is primarily interested in estimation at the boundary and the edge kernel is particularly suited for this purpose (Cheng, Fan and Marron, 1997). The estimated causal effect then equals $\hat{\tau} = \hat{\mu}^+ - \hat{\mu}^-$, the difference in intercepts of the two regression functions, where $\hat{\mu}^+ = \lim_{x \searrow c} \hat{m}_h(x) = \hat{\alpha}^+(c)$ and $\hat{\mu}^- = \lim_{x \nearrow c} \hat{m}_h(x) = \hat{\alpha}^-(c)$.

In the case of polynomial modeling a flexible p -th order polynomial is estimated at both sides of the cutoff over the whole range:

$$(9) \quad Y_i = \alpha_l + f_l(X_i - c) + \varepsilon$$

$$(10) \quad Y_i = \alpha_r + f_r(X_i - c) + \varepsilon$$

for observations left to the cutoff (7) or right to the cutoff (8) respectively. The functions $f_r(X_i - c)$ and $f_l(X_i - c)$ represent any polynomial in X_i . The causal effect is simply estimated as the difference in intercepts, $\hat{\tau} = \hat{\alpha}_r - \hat{\alpha}_l$.

This subsection focuses on how to choose the bandwidth using local linear regression in a RDD context. Choosing a bandwidth in a non-parametric setting involves a tradeoff between bias and variance in the mean squared error expression of the respective estimate. While the

bias is an increasing function of the bandwidth, the variance is a decreasing function of the bandwidth. The bandwidth choice therefore tries to balance these two conflictive tendencies (Härdle, 1990).

While cross validation methods are usually designed to produce a good fit over a broad range of values, this is not directly relevant for the case of RDD, where the interest is exclusively in estimating the regression function at one point, namely the cutoff. Imbens and Kalyanaraman (2010) therefore propose the following mean square error criterion to be minimized by the choice of the bandwidth:

$$(11) \quad MSE(h) = \mathbb{E}[(\hat{\tau} - \tau)^2] = \mathbb{E}\left[\left((\hat{\mu}^+ - \mu^+) - (\hat{\mu}^- - \mu^-)\right)^2\right]$$

This is not a feasible bandwidth since it entails true model parameters. So instead of focusing on this criterion, Imbens and Kalyanaraman (2010) use a modified criterion which is based on the asymptotic mean squared error (AMSE). This is defined (under assumptions stated in therein) as:

$$(12) \quad AMSE(h) = C_1 h^4 \left(m_+^{(2)}(c) - m_-^{(2)}(c)\right)^2 + \frac{C_2}{N * h} \left(\frac{\sigma_+^2(c)}{f_+(c)} - \frac{\sigma_-^2(c)}{f_-(c)}\right)$$

where C_1 and C_2 are kernel specific constants, $m_+^{(2)}$ and $m_-^{(2)}$ are the second derivative of the conditional expectation function from the right and the left respectively, σ_+^2 and σ_-^2 are the right and left limit of the conditional variance function and f_+ and f_- are the marginal distributions of the assignment variable. The first term, the squared bias, is increasing in the bandwidth while the second term, the variance, is decreasing in the bandwidth.

The following theorem taken from Imbens and Kalyanaraman (2010) justifies the use of this criterion.

Theorem 2: Assume the assumptions 3.1-3.5 formulated in IK hold. Then

- i. $MSE(h) = AMSE(h) + o(h^4 + (N * h)^{-1})$
- ii. $h_{opt} = \underset{h}{argmin} AMSE(h) = C_K * \left(\frac{\frac{\sigma_+^2(c)}{f_+(c)} + \frac{\sigma_-^2(c)}{f_-(c)}}{(m_+^{(2)}(c) + m_-^{(2)}(c))^2} \right) * N^{-1/5}, C_K = \left(C_2 / (4C_1) \right)^{1/5}$

Imbens and Kalyanaraman (2010) modify this bandwidth in several ways, leading to the feasible optimal bandwidth they propose:

$$(13) \quad \hat{h}_{opt} = C_K * \left(\frac{2 * \frac{\hat{\sigma}^2(c)}{\hat{f}(c)}}{(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c))^2 + (\hat{r}_+ + \hat{r}_-)} \right)^{1/5} * N^{-1/5}$$

For details on how to operationalize this bandwidth the reader is referred to Imbens and Kalyanaraman (2010). They develop a fully data-dependent algorithm for which they establish certain optimality properties. The routine which calculates this bandwidth has been implemented in MATLAB by Imbens and Kalyanaraman (2010) and is publicly available on their website. This routine will later on be used in the analysis without any changes to the code.

2.2 Data

As opposed to the political system in the United States, Germany has adopted a mixed electoral system after the Second World War. This system combines single member districts (SMD) in one tier with proportional representation (PR) in a second tier (Hainmüller and Kern, 2008). In each electoral district the candidate receiving the highest vote share is elected, independent of any given threshold (in principle a candidate with less than 30% of the votes can win, if all other candidates receive a lower vote share³). This candidate is elected with the so called *Erststimme* (first vote). With the *Zweitstimme* (second vote) the voter does not vote for a single person, but for a whole party instead. For a detailed description of the German electoral system see Bawn (1993,1999). The focus of this paper will be on SMD votes, or the candidates who are elected with the *Erststimme* respectively. With the data at hand it is, however, not possible to distinguish between an incumbent legislator running for re-election and a new candidate from the same party replacing the incumbent. So although the focus will be on SMD votes, what is estimated is party incumbency as defined by Lee (2008)⁴. In principle it is also possible to track the candidates in each district and focus on so called legislator incumbency as defined by Gelman and King (1990).

The data used in order to estimate this incumbency effect are based on German Federal Elections from 1953 to 1998⁵ and on mayoral elections in the state of Hesse from 1992 to 2011. The former have been downloaded from the official webpage www.bundeswahlleiter.de

³ This was indeed the case in 8 districts in the 2009 election, with the lowest vote share necessary to win equal to 26%.

⁴ There is, however, the possibility of spillover effects from SMD votes to PR votes, as analyzed by Hainmüller and Kern (2008).

⁵ . Between the election in 1998 and 2002 the number of districts was reduced from 328 to 299. During the following legislative period there was again a minor redistricting. One district switched from the state of Thuringia to Bavaria, which also changed the boundary of several districts in Thuringia and Upper Bavaria.

whereas the latter have been provided by the *Hessisches Statistisches Landesamt*. Concerning the Federal Elections it was not possible to link each election outcome to the outcome of the previous election because of redistricting⁶. After taking the redistricting into account four waves are available for estimation: the elections from 1953-1961, from 1965-1972, from 1980-1983 and from 1987-1998. All together this gives a total of 2070 observations. The calculation of incumbency effects will be restricted to two parties in Germany, the SPD and the “Union”. These are the two major parties in Germany during the period under consideration. Concerning the political spectrum they cover, they can roughly be compared to the Democrats (SPD) and Republicans (Union). SPD is an abbreviation for *Sozialdemokratische Partei Deutschland* (Socialdemocratic Party of Germany). The “Union” actually comprises two parties, the *Christlich Demokratische Union*, CDU, (Christian Democratic Union) and the *Christlich Soziale Union*, CSU, (Christian Social Union). While the CSU only runs candidates for Federal Elections in Bavaria, the CDU runs candidates in all other states except Bavaria. On the federal level they form a parliamentary group in the *Bundestag* (the German parliament), called the CDU/CSU fraction. In the following analysis it will not be distinguished between CDU and CSU, because this distinction is not relevant for the question, and both parties are referred to as “the Union”.

There are some districts in which one of these parties did not run a candidate, especially in the early elections. In the 1953 election for example, the Union did not run a candidate in districts in Hamburg (district numbers 17, 18, 19, 21, 22), Lower Saxony (district numbers 23, 30, 32, 33, 34, 36, 37, 46, 56), North Rhine-Westphalia (75, 87, 109), Hesse (126) and Bavaria (220, 221, 222). This “non-running” was partly due to an agreement with the *Deutsche Partei* (the coalition partner of the Union at the federal level at that time) on who runs a candidate in certain districts in order not to split the conservative vote. The SPD did not run candidates in the 1957 election in some districts in Bavaria (district numbers 196, 204, 206 and 211) for the same reason, cooperating with the *Föderalistische Union*. This practice, known as *Huckepackverfahren* (pick aback practice), only played a role in the 1953 and 1957 elections, however. Excluding these particular districts seems to be the methodologically cleanest way to handle the problem of a potential bias. In the reduced sample, both SPD and Union place candidates in all districts in all elections.

⁶ Remember that an “observation” in this setting includes, according to (2) the margin of victory in the previous election and the vote share in the current election. So for each district information for two *consecutive* elections is needed. If there had been redistricting between the previous and the current election, it is not possible to construct an observation as described above.

Concerning mayoral elections, mayors in the state of Hesse were directly elected for the first time by the electorate in 1992. Their term in office lasts for 6 years. The vote share necessary to win an election is 50%. If none of the candidates running for office receives a share above this threshold in the first round there is a second round in which only the first two candidates from the first round enter the election. In this second round, the candidate receiving the highest vote share is then elected. Thus, by construction, the winner of a mayoral election always receives at least 50% of the vote. The margin of victory for the Mayoral elections therefore is computed as the difference between a candidates own vote share and the 50% threshold (as in the case of Federal Elections, a negative margin of victory means that a certain candidate has lost the election).

Some elections had to be removed from the sample when estimating the effect of incumbency however, since SPD and Union do not run candidates in all cities in all Mayoral Elections. Doing so reduces the sample from 3354 observations to 1115 observations.

3. Results for Federal Elections and Mayoral Elections

In this section results of the local linear regression as well as the polynomial modeling approach will be presented for the SPD and the Union using two different measures of electoral success. These are the vote share in the next election and the probability of winning the next election. The results based on local polynomial modeling have been produced using STATA/SE 11.0; the estimation of the local linear model based on the optimal bandwidth has been implemented in MATLAB 7.9.0 by the author.

3.1. Results for Federal Election

The estimated incumbency effect in Federal Elections for the SPD and the Union are displayed in table 1. Using local linear regression with the optimal bandwidth developed by Imbens and Kalyanarman (2009) as estimation strategy, I estimate electoral gains to incumbency on vote share in the next election to be 1.33%. Using a flexible polynomial on both sides of the cutoff (fourth order), the effect is slightly higher, namely 1.78%. In terms of magnitude, estimated coefficients seem to be rather small in general, especially when compared to results found in US House Elections, where the estimated incumbency effect is quantified to be around 6% to 8% in terms of vote share. Although these effects are rather small they are nonetheless economically important. This can be seen by the fact that these small increases in vote share translate into an increased probability of winning the next

election of 21.88% or 28.6% respectively. That this small increase in vote share translates into such a large increase in the probability of winning the next election can be explained as a result of a lot of districts being won by a very small margin. Thus even small gains in vote share increase the probability of winning by a large amount. This relationship is also apparent from figure 1 and figure 2, where the estimated regression function is shown together with local averages of the dependent variable.

While the jump in the regression function is very small in figure 1, it is much larger in figure 2. Compared to the results of Hainmüller and Kern (2008), the results presented here are very similar in magnitude, although they use a slightly different sample⁷. They use a third order polynomial on both sides of the cutoff and report an incumbency effect in terms of vote share for the SPD of 1.51% and 1.88% for the Union compared to 1.33% and 1.78% for both parties pooled together. Their results are also significant at 5%. So the effect seems to be quite robust (given the specification of Hainmüller and Kern, 2008) with respect to minor changes in the data used (as it should be, of course).

Table 1: Effect of Incumbency in Federal Elections

Estimation Strategy	Effect of Incumbency on	Coefficient	Optimal bandwidth	Polynomial order
Local linear regression	Vote share	1.33%** (0.0055)	0.132	-
	Probability of winning	21.88%*** (0.0541)	0.09	-
Polynomial regression	Vote share	1.78%** (0.0059)	-	4
	Probability of winning	28.6%*** (0.0675)	-	6

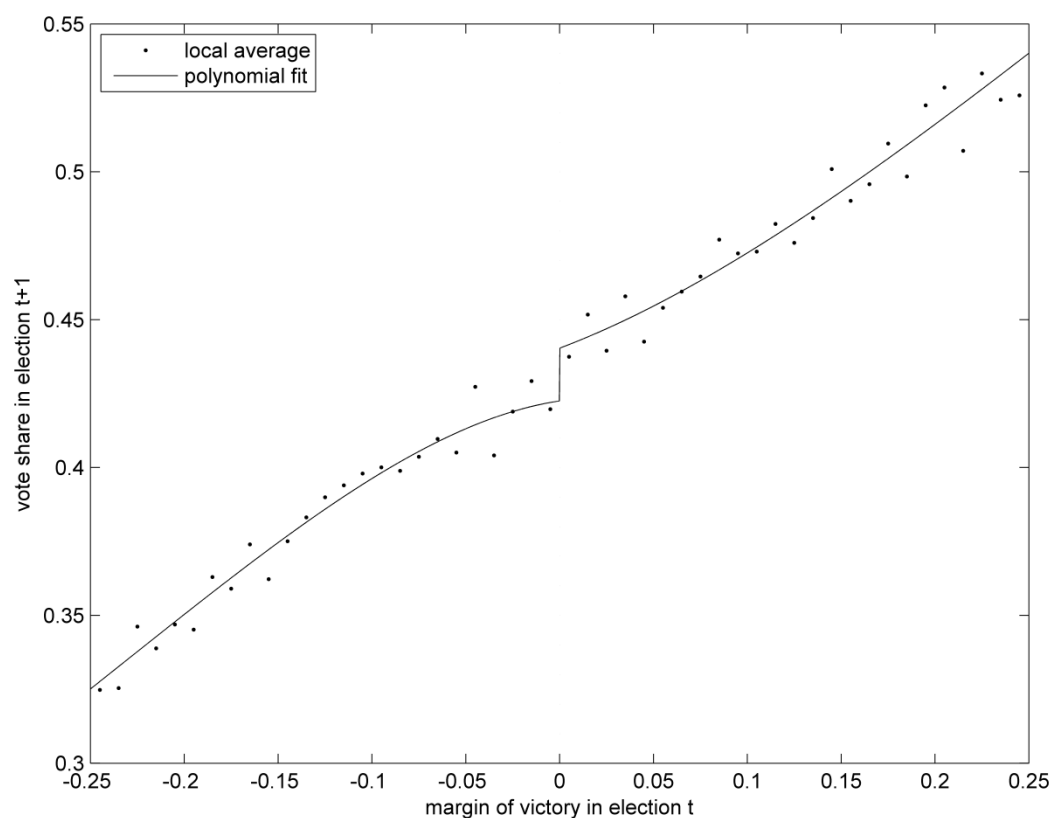
Note: Robust standard errors in parenthesis. Number of observations is 4140 in all estimations. Choice of polynomial order based on Akaike information criterion. *, ** and *** denote significance at the 10%, 5% and 1% level respectively.

Using local linear regression as opposed to polynomial regression basically yields a very similar point estimated with respect to the effect of incumbency status on vote share. The difference is somewhat larger when considering the effect on the probability of winning the next election, where the difference is almost 7% points. In general the results based on the local linear specification should be trusted more, since the bandwidth used has certain optimality properties and the edge kernel is particularly suited for the purpose of estimating a regression function at a certain point (namely the cutoff). In addition, the order of the

⁷ For details on their sample see Hainmüller and Kern (2008).

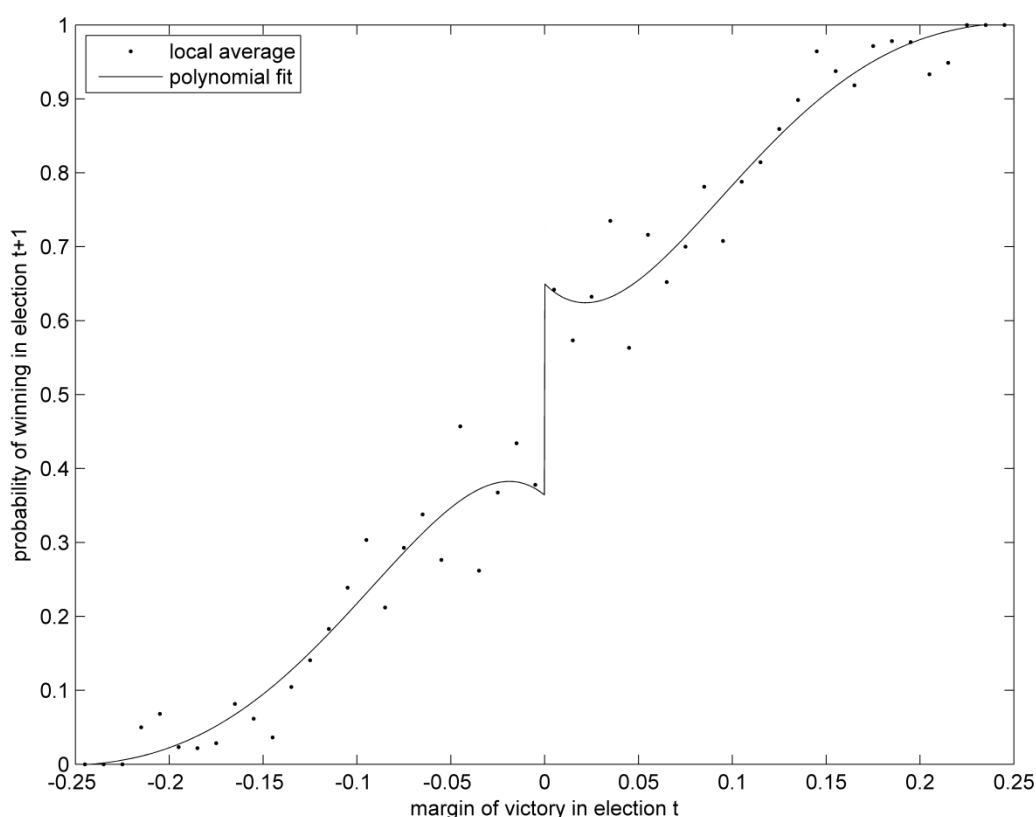
polynomial when analyzing the effect on the probability of winning (sixth order polynomial) seems rather larger, although it had been chosen based on a formal criterion.

Figure 1: Estimated Regression Function for Federal Elections, impact on vote share



Note: This figure illustrates the discontinuity in the estimated regression function. For visual clarity, the data have been grouped in bins, representing an interval of 1 percent in the margin of victory.

Figure 2: Estimated Regression Function for Federal Elections, impact on the probability of winning



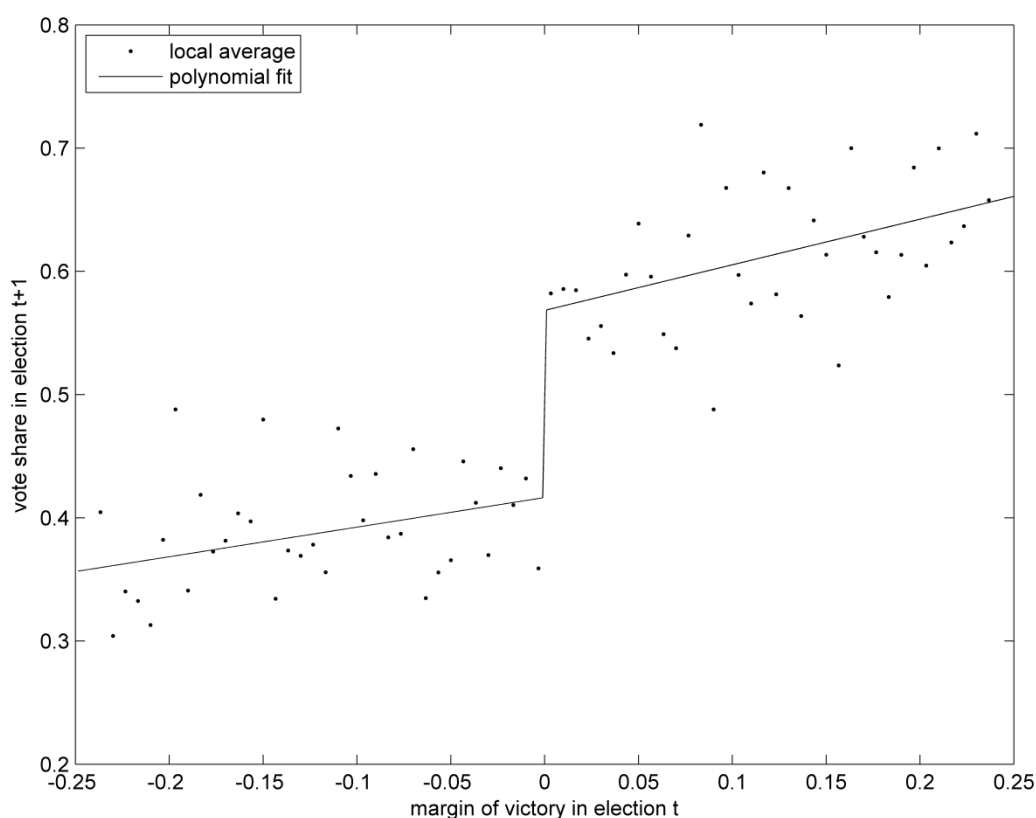
Note: This figure illustrates the discontinuity in the estimated regression function. For visual clarity, the data have been grouped in bins, representing an interval of 1 percent in the margin of victory.

3.2. Results for Mayoral Elections

The results for Mayoral Elections for the SPD and the Union are displayed in table 2. The estimated effect of incumbency is much larger than in Federal Elections. While the impact of incumbency on vote share is merely between 1.3% and 1.8% depending on the estimation strategy, it is about ten times larger when looking at Mayoral Elections. In the case of local linear regression (again using the optimal bandwidth and the triangular kernel) the estimated coefficient of incumbency status on vote share is 16.47% and highly significant. This translates into an increase in the probability of winning the next election of 57%. In the case of polynomial regression the specification chosen by the Akaike information criterion is interestingly a polynomial of order one so that the only differences in the estimation is the range of observations included. The estimated effect of incumbency on vote share in this specification is 13.99% and the effect on the probability of winning the next election is 57% as in the case of local linear regression. These figures are much higher than the 32% increase

in the probability of winning that Ferreira and Gyourko (2009) find in their analysis of incumbency effects in Mayoral Elections in the United States and the 45% increase estimated by Lee (2008) in his analysis of United State House Elections. These effects are visualized in the same way as for Federal Elections in figure 3 and figure 4.

Figure 3: Estimated Regression Function for Mayoral Elections, impact on vote share



Note: This figure illustrates the discontinuity in the estimated regression function. For visual clarity, the data have been grouped in bins, representing an interval of 2/3 percent in the margin of victory.

Comparing the results of both estimation strategies shows that the estimated coefficient is quite robust with respect to the strategy chosen. While the local linear regression finds an effect of incumbency on vote share of about 16.5%, the polynomial regression estimates an effect of 14%. That these coefficients are similar is not very surprising, since both estimates are based on a first order polynomial. The only difference is that the polynomial regression uses the full sample without applying a weighting scheme; the local linear regression only uses those observations which are within an 18.99% bandwidth to the cutoff and weights observations relative to its distance to the cutoff. The effects concerning the probability of

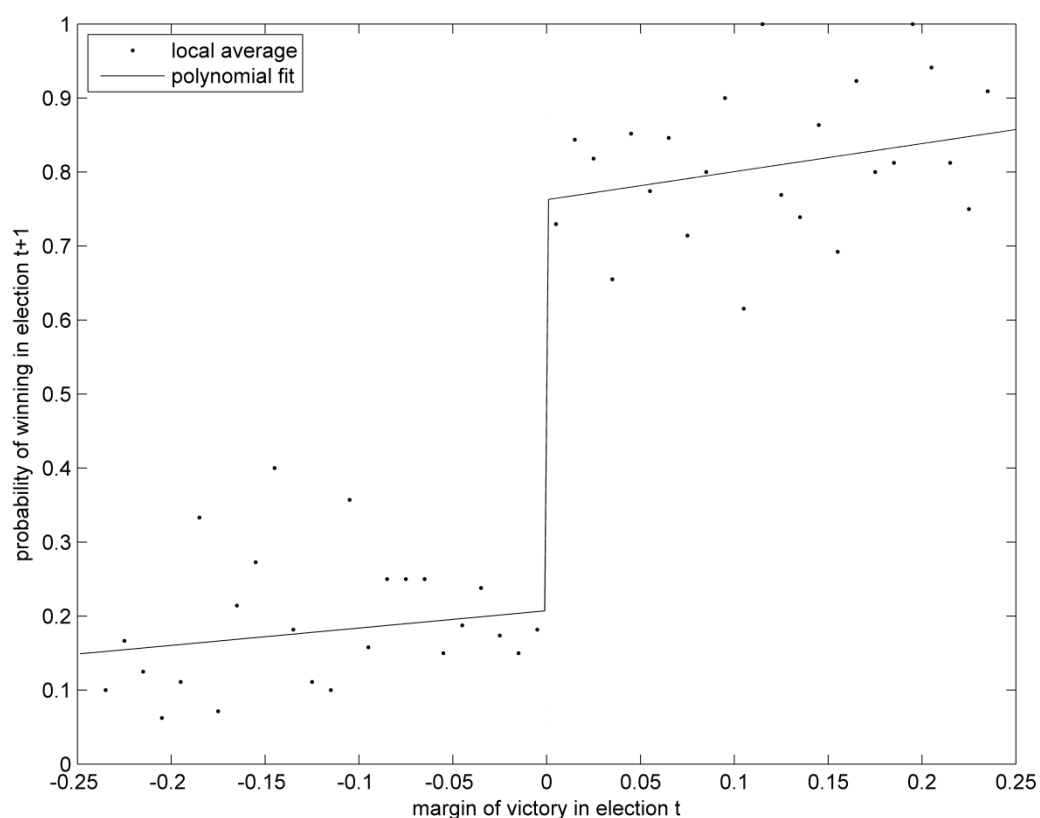
winning the next election are even more similar, since the optimal bandwidth chosen basically covers the whole range of observations.

Table 2: Effect of Incumbency in Mayoral Elections

Estimation Strategy	Effect of Incumbency on	Coefficient	Optimal bandwidth	Polynomial order
Local linear regression	Vote share	16.47%*** (0.0183)	0.1899	-
	Probability of winning	57.04%*** (0.0465)	0.3736	-
Polynomial regression	Vote share	13.99%*** (0.0138)	-	1
	Probability of winning	57.08%*** (0.0418)	-	1

Note: Robust standard errors in parenthesis. Number of observations is 1114 in all estimations. Choice of polynomial order based on Akaike information criterion. *, ** and *** denote significance at the 10%, 5% and 1% level respectively.

Figure 4: Estimated Regression Function for Mayoral Elections, impact on probability of winning



Note: This figure illustrates the discontinuity in the estimated regression function. For visual clarity, the data have been grouped in bins, representing an interval of 2/3 percent in the margin of victory.

As already outlined briefly in the introduction, these huge difference between Federal and Mayoral Elections in the effect of incumbency on vote share and the probability of winning might be explained by the differences in visibility and popularity. While single member district candidates are in general not as well known to the electorate⁸, mayors certainly are.

4. Robustness checks

One way to check the robustness of the results presented is to include predetermined covariates in the estimation and see whether the estimated coefficient on the incumbency dummy changes. Inclusion of covariates in the regression should not affect the estimated incumbency effect (in analogy to a true randomized experiment). If anything, this inclusion should result in lower standard errors as some variability in the data is reduced. Unfortunately, the only covariates at hand are voter turnout and time and state effects for Federal Elections and voter turnout in the case of Mayoral Election. Another way to test whether voter turnout in the current election is balanced around the cutoff is to run a local linear regression⁹ using turnout as dependent variable. It turns out that the jump at the cutoff is insignificant for this variable for both Federal and Mayoral Elections. The respective results are available from the author upon request.

Table 3 and table 4 show the results of the local linear and polynomial regression when voter turnout and time and state fixed effects are included in the analysis (the latter two only in the case of Federal Elections). As can be seen from these tables, the estimated coefficient is very robust to the inclusion of further covariates.

Table 3: Robustness check Federal Elections

Estimation Strategy	Effect of Incumbency on	Coefficient	Optimal bandwidth	Polynomial order
Local linear regression	Vote share	1.37%*** (0.0045)	0.132	-
	Probability of winning	20.69%*** (0.0541)	0.09	-
Polynomial regression	Vote share	1.76%*** (0.0050)	-	4
	Probability of winning	28.61%*** (0.0677)	-	6

Note: Robust standard errors in parenthesis. Number of observations is 1114 in all estimations. Choice of polynomial order based on Akaike information criterion. *, ** and *** denote significance at the 10%, 5% and 1% level respectively. Additional covariate included in the regression: voter turnout, state fixed effects, time fixed effects.

⁸ In fact, no one I asked knew the name of the person who had won the respective district in the last Federal Election, and only few knew from to party he or she belonged to.

⁹ In this regression, the optimal bandwidth and the edge kernel were used as in the main estimation outlined in section 2 and applied in section 3.

Including voter turnout, time fixed effects and state fixed effects in the regression as additional covariates changes the coefficient only slightly from 1.33% to 1.37% (local linear regression) or from 1.78% to 1.76% (polynomial regression) respectively in terms of vote share. Concerning the probability of winning, the coefficient changes from 21.88% to 20.69% (local linear regression) or from 28.6% to 28.61% (polynomial regression) respectively. Standard errors slightly decrease.

Table 4: Robustness check Mayoral Elections

Estimation Strategy	Effect of Incumbency on	Coefficient	Optimal bandwidth	Polynomial order
Local linear regression	Vote share	16.52%*** (0.0180)	0.1899	-
	Probability of winning	58.02%*** (0.0455)	0.3736	-
Polynomial regression	Vote share	13.67% (0.0138)	-	1
	Probability of winning	56.98% (0.0418)	-	1

Note: Robust standard errors in parenthesis. Number of observations is 1114 in all estimations. Choice of polynomial order based on Akaike information criterion. *, ** and *** denote significance at the 10%, 5% and 1% level respectively. Additional covariate included in the regression: voter turnout.

The estimated effect of incumbency on vote share in the next Mayoral Election and the probability of winning the next Mayoral Election using local linear regression changes from 16.46% to 16.52% and from 57.04% to 58.02% while standard errors decrease from 0.0183 to 0.0180 and from 0.0465 to 0.0455.

These robustness checks show that including predetermined covariates in the estimation does not alter the estimated coefficient on the effect of incumbency; hence the effect of incumbency as estimated in this analysis is credible.

5. Conclusions

In this paper I show that incumbency effects are much larger for Mayoral elections in the state of Hesse compared to Federal Elections in Germany using a quasi-experimental research design. While the effect of incumbency on vote share in the next election is only about 1.5% in terms of vote share in Federal Elections, it is roughly ten times larger in Mayoral Elections in the state of Hesse with about 15%. These huge differences also translate into a much larger probability of winning the next election. The results obtained in this paper are robust in the following sense: on the one hand, using two different estimation strategies (local linear regression and polynomial modeling) virtually yield the same results for the effect of

incumbency on vote share and the probability of winning the next election; on the other hand, inclusion of further covariates (as far as they are available) does not change the estimated coefficient on incumbency. Both these statements hold true for Federal Elections and Mayoral Elections in the state of Hesse. In addition, the usage of formal criteria for choosing the bandwidth or the polynomial order, respectively adds further credibility to the results. A possible explanation for these large differences is the difference in the character of the election and the candidates. In Federal Elections, voters mostly care about the election outcome in the aggregate: it is not that much which candidate won in their district but which party won the majority of seats in the parliament. In Mayoral Elections the electorate of course cares directly about the success of their favored candidate. Furthermore, single member district candidates in Federal Elections are much less well known and much less visible than candidates running for mayor's office. In addition, it is much harder to make a single member district candidate responsible for what has happened on the Federal level in the past legislation period. Mayors and opposition leaders in the city council can much easier be called to account for what has happened on the local level in the past legislation period. These differences in visibility (and usually popularity) might serve as another explanation of the observed differences.

References:

- Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114(2): 217-221.
- Bawn, Kathleen. 1993. "The logic of institutional preferences: German electoral law as a social choice outcome." *American Journal of Political Science*, 37(4): 965-989.
- Bawn, Kathleen. 1999. "Voter response to electoral complexity: ticket splitting, rational voters and representation in the federal republic of Germany." *British Journal of Political Science*, 28(3): 487-505.
- Der Bundeswahlleiter. 4.4.2011. <www.bundeswahlleiter.de>
- Cheng, Ming-Yen, Jianqing Fan, and J.S. Marron. 1997. "On Automatic Boundary Corrections." *The Annals of Statistics*, 25(4): 1691-1708.
- Cook, Thomas 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.
- Ferreira, Fernando and Joseph Gyourko. 2009. "Do Political Parties Matter? Evidence From US Cities." *Journal of Economics*, 124(1): 399-422.
- Gelman, Andrew, and Gary King. 1990. "Estimating Incumbency Advantage without Bias." *American Journal of Political Science*, 34(4): 1142-1164.
- Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. „Identification and Estimation of Treatment Effects with a Regression Discontinuity Design." *Econometrica*, 69(1): 201-209.
- Hainmüller, Jens, and Holger L. Kern. 2008. „Incumbency as a Source of Spillover in mixed Electoral Systems: Evidence from a Regression Discontinuity Design." *Electoral Studies*, 27(2): 213-227.
- Härdle, Wolfgang. 1990. *Applied Nonparametric Regression*. New York: Cambridge University Press.
- Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics*, 115(4): 1239-1285.
- Imbens, Guido W., and Karthik Kalyanaraman. 2010. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." Center for Microdata Methods and Practice Working Paper CWP05/10.
- Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Design: A Guide to Practice." *Journal of Econometrics*, 142(2): 615-635.
- Jacobson, Gary C. 1997. *The Politics of Congressional Elections*. Menlo Park, California: Longman.
- Lee, David S. 2001. "The Electoral Advantage to Incumbency and Voters' Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Close Elections." University of California Berkeley Center for Labor Economics Working Paper 31.
- Lee, David S. 2008. "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics*, 142(2): 677-697.
- Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2): 281-355.
- Lee, David S., and Justin McCrary. 2005. "Crime, Punishment and Myopia." National Bureau of Economic Research Working Paper 11491.

- Lee, David S., Enrico Moretti, and Matthew J. Butler. 2004. "Do Voters Affect of Elect Policies? Evidence from the U.S. House." *Quarterly Journal of Economics*, 119(3): 807-859.
- Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chance? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics*, 122(1): 159-208.
- Rubin, Donald D. 1974. "Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5): 688-701.
- Thistlethwaite, Donald L., and Donald T. Campbell. 1960. "Regression Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology*, 51(6): 309-317.
- Van der Klaauw, Wilbert. 2002. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach." *International Economic Review*, 43(4): 1249-1287.