Role-models and female political participation

Evidence from a regression discontinuity approach to Canadian elections

Author:
Sjors Hoogeveen

Student ID:
412903

Supervisor:
Dr. Jurjen Kamphorst

Second assessor:
Dr. Sacha Kapoor

Master Thesis Policy Economics

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
The Netherlands
29 July 2021

Contents

Role-models and female political participation

Evidence from a regression discontinuity approach to Canadian elections

Sjors Hoogeveen

Abstract. Around the globe, women remain under-represented in parliaments. One frequently invoked explanation behind this gender gap in politics is the dearth of visible female politicians itself. Not only could female politicians act as role-models for other women, but examples of successful female politicians could also reduce the biases, barriers, or disincentives that currently hamper women's access to political office. By applying a regression discontinuity design to federal elections in Canada, I investigate whether the election of a woman causes an increase in the number of female candidates within an electoral district. The main results robustly show that the election of a woman indeed increases the number of female candidates in the next election. However, in a number of extensions, I show that this effect is most likely attributable to the tendency of incumbent parliamentarians to run for re-election. No evidence suggests that the election of a woman makes it more likely that *other* women become candidates.

1 Introduction

A hundred years ago, Agnes Macphail was the first woman elected to the Canadian House of Commons in 1921. Since the election of Macphail, the number of women in the Canadian parliament has grown steadily, and in the 2019 general election, a record number of 98 women became Members of Parliament (MPs). Although the level of female political participation has increased substantially over the last century, gender gaps in Canadian politics remain. Even with the record of 98 elected women, only 29.0% of all MPs is female. A quick glance at Figure 1 leaves little doubt that women have been, and continue to be (vastly) under-represented in the Canadian federal parliament.

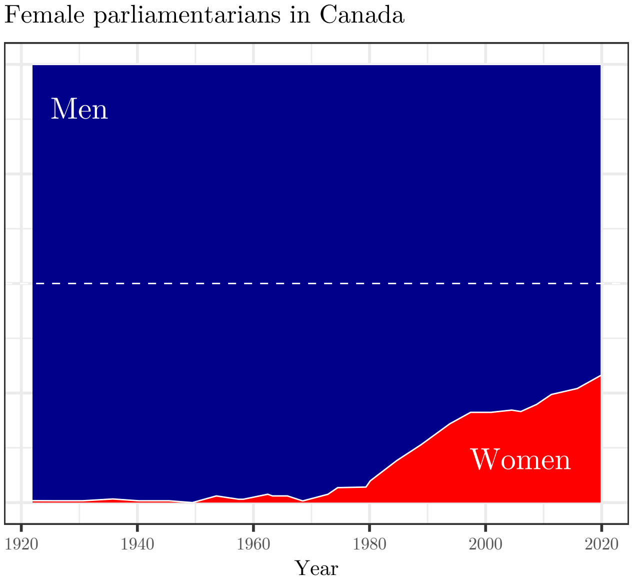

Figure 1: Gender composition of the Canadian parliament since the election of Macphail in 1921. The dotted horizontal white line denotes the point at which a gender-balanced parliament would be reached.

Despite Justin Trudeau's simple explanation for a gender-balanced cabinet: 'because it's 2015', full gender parity has not been reached in all aspects of Canadian politics (Hartviksen 2015). In 2017, some Canadian MPs have voiced their dissatisfaction with the disproportionate number of women in parliament more explicitly, stating that 'the current representation is a failure, and that prioritising women is still needed' (Galandy and Tavcer 2019, p. 16).

The under-representation of women in the Canadian parliament is far from an anomaly. Worldwide, only 25.6% of all MPs are women as of 1 January 2021,[1] and with the current rates of progress, the road the gender-parity remains long. UN Women calculated that gender-balance in national parliaments would take another 46 years (UN Women 2021), while the World Economic Forum estimated that a gender gap in political empowerment in general will continue to exist for another 146 years (World Economic Forum 2021).

Political scientists and philosophers have long argued that the under-representation of women—or any other (historically) marginalized groups—in politics can deteriorate the legitimacy of democratic institutions (Clayton et al. 2019; Dovi 2002; Mansbridge 1999) and, when the specific interests of such under-represented groups are insufficiently taken into account by public decision-making bodies, it can result in the creation or continuation of situations of injustice (Chattopadhyay and Duflo 2004; Dovi 2018). Unsurprisingly, closing the gender gap in politics is widely considered a legitimate and important policy objective. In the Beijing Declaration and Platform for Action, adopted in 1995 by 189 states, it is stated that 'Women's equal participation in decision-making is not only a demand for simple justice or democracy but can also be seen as a necessary condition for women's interests to be taken into account' (UN 2014, p. 119).

The under-representation of women in politics does not only pose normative questions, but also an important descriptive question: why do relatively few women hold elected public office? And, given the general con-

[1] Data from Inter-Parliamentary Union (2021).

sensus that gender parity in politics is an important objective, what tools and instruments could be used by policymakers to accelerate the progress towards gender equality?

Since the formal or legal constraints which once prevented women from running for political office have been abolished in almost all countries, naive observers could expect the proportion of female parliamentarians to be similar to the overall proportion of women in the total population. The fact that this is evidently not the case in most countries, indicates that somewhere along the process of winning an election, women face barriers, deterrents, or disincentives that are not faced by men.

Historically, much attention has been paid to inequalities in political resources—*e.g.* education, work experience, political knowledge and interest, or available free time—as explanatory factors for the under-representation of women in politics (for an overview, see Paxton et al. 2007). However, in the last decades many of these gender differences in political resources have shrunk considerably, and some have even reversed. In Canada for example, women are more likely to hold university certificates or degrees compared to men (Ferguson 2016), and in 2018, 46% of all Canadian lawyers—a profession generally seen as a potential 'springboard' into politics (Mariani 2008)—was female (Federation of Law Societies of Canada 2018). The persistence of the gender gap in politics in light of fast(er) closing gender gaps in many other areas has resulted in a shift in focus towards more contextual and institutional explanations for the under-representation of women in politics (Burns et al. 2001). In general, it has become widely acknowledged that resource-based supply-side explanations remain important, yet do not fully account for the gender gap in politics on their own (D. E. Campbell and Wolbrecht 2006).

One influential theory on how the political environment affects female political participation is that a lack of exposure to female politicians can itself impede women from becoming politically active (*e.g.* Karp and Banducci 2008; Wolbrecht and D. E. Campbell 2007). Sapiro (1981) observed 40 years ago that '[w]omen and men continue to think of politics as a male domain because the empirical truth at this moment is that politics *is* a male domain'. Proponents of the exposure-based explanation argue that a feedback loop between low exposure to female politicians and reduced political participation among women makes it unlikely that the gender gap in politics will disappear as long as politics remains (perceived as) a 'male domain'.

This theory is primarily based on a set of two related mechanisms. First, (visible) female politicians are believed to act as potential role-models and sources of motivation or inspiration for other women aspiring a career in politics (Beaman, Duflo, et al. 2012; D. E. Campbell and Wolbrecht 2006; Ladam et al. 2018). Especially given the widely documented gender gap in self-confidence regarding political knowledge and ambition (Fox and Lawless 2004; Fox and Lawless 2011; Gidengil et al. 2008; Thomas 2012), the presence of role-models might be crucial in showing that women can and do have success in the political arena (Latu et al. 2013). If an increased female presence in parliaments diminishes the association of politics being a 'men's job', this can be expected to stimulate the inflow of women into politics (Akerlof and Kranton 2000) and attenuate the gender-gap in self-confidence on political abilities (Bordalo et al. 2019).

Secondly, examples of successful female politicians are also believed to be capable of diminishing gender-specific biases among other actors and gatekeepers with influence over the political behaviour of women, such as the electorate, political elites, or the media, by showcasing the suitability of women to participate in the political arena (Atkeson and Krebs 2008; Beaman, Chattopadhyay, et al. 2009). For example, Sanbonmatsu (2006) notes that the under-representation of women could create a self-fulfilling prophecy when party leaders have doubts about the electability of women, and consequentially become reluctant to support or nominate female candidates, *because* so few women are elected (see also Shair-Rosenfield 2012). And, since (former) female MPs can likely be considered to be part of the local political elite themselves, they could potentially assert substantial influence over the emergence of female political candidates (C. Cheng and Tavits 2011).

Based on this exposure-based explanation for the under-representation of women in politics, I aim to investigate whether the the election of a female politician causes an increase in the number of female candidates in Canada within a federal electoral district. In Canada's single-member constituency voting system, MPs often have a strong connection with their electoral district and spend considerable amounts of time and resources on local constituency work (Franks 2007; Heitshusen et al. 2005). Furthermore, local party elites, most notably the Electoral District Associations (EDA), have substantial influence over the recruitment and nomination process of political candidates in Canada (Carty et al. 2003; Cross 2016). For example, both C. Cheng and Tavits (2011) and Cross and Pruysers (2019) find that women are more likely to participate in local nomination races, and actually be nominated, when women occupy positions of local party authority—in their case the EDA presidency.

Based on this close relationship between a constituency and its federal representative on the one hand, and a large influence of the local political elite on the process of becoming a candidate on the other hand, combined with the theorised potential of female politicians to act as role-models and to impact the behaviour or attitudes of the electorate and (local) political elites, the following hypothesis can be formulated:

> The election of a woman to the Canadian

> House of Commons increases the number of female candidates during the subsequent election in the same federal electoral district.

Finding empirical evidence for this hypothesis is however fraught with difficulty. It is almost certain that a simple empirical examination of the relationship between the exposure to female politicians and the number of female candidates running for election will suffer from the problem of endogeneity. Most likely, the chances of female electoral success and the number of women on the ballot are jointly correlated with a shared set of unobserved or unobservable factors which render naive estimates of the causal effect of female electoral success on the number of female candidates biased (Angrist and Pischke 2008, p. 61). Examples of factors which are correlated to both the likelihood of female electoral success and the number of female candidates could be the level of religious conservatism among the electorate, female labour participation, or income and wealth levels within the electoral district.

To overcome this problem of endogeneity, I will exploit the Canadian first-past-the-post electoral system with a regression discontinuity (RD) design. In electoral districts where the two strongest candidates are a combination of a woman and a man, treatment status—*i.e.* female electoral success—changes discontinuously around a certain cut-off point: as soon as the strongest female candidate gains a single vote more than any of her competitors, the election is won by a woman. By comparing electoral districts where the margin of victory is very close to this cut-off point, credible estimates of the local average treatment effect can be obtained.

The main results of the RD design indeed suggest that in Canada, the election of a woman to the House of Commons causes a substantial and statistically significant increase in the number of female candidates during the next election. After presenting evidence for the validity of the RD design, I will aim to provide a more in-depth exploration whether this causal effect can be attributed to a role-model effect. Based on four extensions of the main analysis, it must be concluded that *no* credible evidence can be found for expectation that the election of a woman increases the likelihood that *other* female candidates run for election. Instead the observed increase in female candidates is most likely attributable to the tendency of incumbent members of parliament to seek re-election.

The remainder of this thesis is structured as follows: in section 2, related research will be discussed, including research on the relationship between exposure to female politicians and female political participation and engagement, the existence of role-model effects in general, and the application of RD designs in electoral settings. The identification strategy is discussed in detail in section 3, followed by a description of the estimation and inference techniques in section 4. The data and RD sample are presented in section 5, and the main results of the RD analysis are given in section 6. Tests for the validity of the RD design are presented in section 7. In section 8, four complementary RD analyses are conducted to further investigate the potential causal mechanisms behind the treatment effect. Finally, section 9 concludes.

2 Related literature

This thesis is closely related to three bodies of research. First, this thesis builds upon a substantial body of literature on the relationship of female leadership in politics and female political participation. Secondly, this thesis is also related to studies on role-model effects in general. Thirdly, this thesis draws methodologically on the ever-popular application of a RD design in electoral settings to estimate the effects of the results of elections on political, economic, or social outcomes of interest. In this section, I will discuss these three related bodies of research.

2.1 The relationship between exposure and participation

Overall, the existing empirical evidence on the relationship between exposure to female politicians and political participation among women is mixed (Mariani et al. 2015). In one of the earliest empirical examinations of this relationship, Verba et al. (1997, p. 1069) concluded that the impact of women as part of the visible political elite on the political engagement of women—as proxied by measures on political knowledge and efficacy—was a 'definitive maybe'.

In a later 'initial attempt' to analyse existing survey data in the US to explore whether symbolic representation by female politicians affect the political behaviour and attitudes of women, Lawless (2004) argue that the 'presence of women in politics does not seem to affect women's political trust, efficacy, competence, and engagement'. In a more extensive study of American elections and in a large cross-national examination, Dolan (2006) and Karp and Banducci (2008) reach similar conclusions, finding no evidence in support of the idea that the presence of female politicians has a clear impact on the political attitudes and behaviour of women.

Multiple authors do however find (suggestive) evidence on a positive relationship between exposure to female politicians and political participation. In the work of Atkeson (2003), the earlier 'definitive maybe' from Verba et al. (1997, p. 1069) is substituted by a resolute yes. Based on regression analysis of data from the American National Election Studies, Atkeson (2003, p. 1053) concludes that 'the lack of visible female political players has helped the gender gap in political engagement to persist'. In an environment where competitive female politicians are present, she finds that women are more likely to discuss politics and that they are more confident in their level of political knowledge (see also Atkeson and Carrillo 2007). D. E. Campbell and Wolbrecht (2006) also note that the presence of female politicians is associated with a process of 'political socialisation' of adolescent girls in the US—who were more likely to expect to become involved in politics later in life when exposed to female politicians.

Overall, the idea that exposure to female politicians can have positive effects in the US seems to be broadly accepted be a large number of researchers. Reingold and Harrell (2010) conclude that an increased presence of women in American politics 'can have significant, mobilizing effects on women', and Fridkin and Kenney (2014) argue that 'women citizens are more active in politics when represented by women senators'. Furthermore, some researchers have also found evidence specifically in support of the hypothesis of this thesis. For example, Palmer and Simon (2005) analyse US House of Representatives election from 1956–2002 to argue that the presence of female incumbents 'increases the entry and participation of female candidates'.

The positive relationship between exposure to female politicians and political participation among women has also been documented outside the US. In the UK, Norris et al. (2004) observe that women tend to more politically active—measured by voter turnout, campaign interest, and voluntary campaign work—when female MPs are elected. In Switzerland—infamous for its late adoption of women's suffrage—Gilardi (2015) also finds evidence for a role-model effect, although he does note that the positive effect of female representation on the inflow of female candidates fades away over time, indicating that role-model effects become less important as the level of female representation in politics reaches a certain 'adequate level'. In light of this conclusion, it is no surprise that in the relatively young democracies in Sub-Saharan Africa, increases in the the percentage of women in parliaments are 'associated with increases in women's political engagement' (Barnes and Burchard 2012). Similarly, in Indonesia, Shair-Rosenfield (2012) found that the presence of female incumbents increased the likelihood of high-position nomination and election for female newcomers.

Both Wolbrecht and D. E. Campbell (2007), and Bühlmann and Schädel (2012) analyse large cross-national datasets to find evidence for the suggestion that increased female representation is associated with heightened political engagement among women and girls, 'inspiring other women to be politically motivated and active themselves'. Alexander (2012) also documents a positive association between the presence of women in parliaments and women's beliefs about women's ability to govern in 25 countries. Finally, Liu and Banaszak (2017) use data from 20 democracies to argue that not only the share of female legislators, but also the presence of female cabinet members is associated with increased levels of political participation among women.

Unfortunately, all these studies suffer from an important common drawback: the lack of credible identification strategies preclude drawing definitive conclusions on the potential causal effect of exposure to female politicians on the political behaviour and attitudes of women. Several studies with more credible research designs have been conducted, which again show mixed evidence for the effects of female representation on political engagement among women.

A number of researchers have exploited a policy experiment in India where in a group of randomly selected villages, one-third of the council seats was reserved for women. First, Chattopadhyay and Duflo (2004) found that political participation among women increased significantly in the West Bengali villages that witnessed an increase in female representation on the village council due to the policy experiment. Secondly, Beaman, Chattopadhyay, et al. (2009) observe that exposure to female leadership significantly reduces prejudice against female leadership. Thirdly, Beaman, Duflo, et al. (2012) also found that an increase in female representation in the village councils had a substantial positive impact on the educational and professional aspirations of girls, while it simultaneously contributed to closing the gender gap in educational outcomes and time-use patterns. Overall, there is substantial evidence that the quota-based increase in female representation in West Bengal created an social and political environment in which aspiring female politicians were more likely to prosper.

Most closely related to this thesis are empirical studies that apply RD designs to estimate the causal effect of the election of a woman on the female political behaviour during the subsequent election. The first author—to my knowledge—to utilise a RD design in such setting is Broockman (2014), who uses data from US State legislative elections. Although he finds that female electoral success has a large positive effect on the likelihood of a female candidate in the subsequent election in the same district, it seems that this effect is attributable to the tendency of politicians in the US to almost always run for re-election. None of Broockman's (2014) results indicate that the election of a woman induces *other* women to run for election. For example, the effect of female electoral success on the number of female candidates in *nearby* districts is found to be zero.

However, based on evidence on the effect of women holding more prominent offices (governors and senators) on the female proportion of candidates for US state legislatures, Ladam et al. (2018) challenge the overall conclusion that female role-models do not seems to cause female candidates to emerge in the US. Although their identification strategy—a variation of propensity score matching—is more prone to bias due to unobserved confounding variables, their conclusion can cast doubt on whether the results of Broockman (2014) also apply to situation where women get elected to more prominent offices—such as federal, instead of local legislative bodies.

In India, Bhalotra et al. (2018), use a similar identification strategy and draw similar conclusions as Broockman (2014). Bhalotra et al. (2018) apply a RD design to India's state elections to analyse whether female electoral success increases the share of women as candidates for subsequent elections. They find that the election of a woman results in a significantly higher share of female candidates in the following election. This higher share is however attributed to the propensity of females to run for re-election, and not to a larger entry of new female candidates. Analysing elections at different levels (national, regional, municipal) in Poland with a RD design, Jankowski et al. (2019) also note a large incumbency effect to be present, yet no empowerment effect: the election of women does not seem to cause an increase in the number of non-incumbent women running for election. Interestingly, Jankowski et al. (2019) do find limited evidence for a contagion effect, observing that political parties are more likely to nominate women at high positions on their lists when female politicians from opposing parties had enjoyed electoral success.

2.2 Role-model effect outside of politics

In areas outside politics where gender gaps are present, mitigating role-model effects have also been found in studies with credible identification strategies. The overall message of these studies is that the visibility of, and exposure to female role-models can cause changes in the behaviour and attitudes of other women in a wide variety of (unexpected) settings.

In corporate settings for example, a larger share of women in leadership positions is found to have a spillover effect on the likelihood that other women occupy positions in top management in a large-scale fixed-effects study conducted by Matsa and A. R. Miller (2011). In Norway, where gender quotas for corporate boards have recently been implemented, Kunze and A. R. Miller (2017) also find that a larger share of women in leadership positions is associated with a decrease in the gender gap in promotions. Although this effects could be largely attributable to the likelihood of more female-friendly cultures in firms where women occupy (more) leadership positions (Tate and Yang 2015), Bertrand et al. (2018) do observe that one of the effects of the mandatory Norwegian gender quota was that young women in business reported more positive career expectations after the quota went into effect. So, even though Bertrand et al. (2018) do not find credible evidence themselves for a direct spillover effect of the gender quotas, they do conclude that 'It is possible that the positive mindset the reforms induced among young women in business will ultimately encourage them to stay on the fast track for longer'.

In educational settings, role-model effects have also been found. In an experimental study where students were randomly assigned to their professors, Carrell et al. (2010) found that female students were more likely to pursue a major in male-dominated STEM disciplines when they had a female professor for the mandatory introductory courses in these disciplines. Similar results were found by Porter and Serra (2020) in a field experiment where female students who were exposed to successful and charismatic female economics graduates were more likely to major in economics themselves.

Several authors have also documented a role-model effect of exposure to less traditional gender roles in television. In rural India, Jensen and Oster (2009) find a positive effect of the introduction of cable television on the status of women based on a fixed effects model. In Brazil, researchers have exploited regional variation in the availability of television channels to conclude that that exposure to modern lifestyles and emancipated female roles in *telenovelas* was associated with increased divorce rates (Chong and Ferrara 2009), and has led to significantly lower fertility (La Ferrara et al. 2012).

Overall, there is considerable evidence that role-model effects do exist and that the exposure to role-models can influence behaviour and attitudes in a wide array of situations.

2.3 Regression discontinuities and elections

Since the seminal work of Lee (2008), who uses a RD design to estimate the electoral advantages of incumbency in US House elections, RD designs have been widely applied to first-past-the-post elections (*e.g.* Butler 2009; Eggers and Hainmueller 2009). The reason behind the popularity of RD designs in electoral settings is that RD designs do not only provide easily interpretable estimates of causal effects, but the characteristics of elections also provide strong theoretical and empirical evidence for the validity of the RD design (Eggers, Fowler, et al. 2015)—although Caughey and Sekhon (2011) have argued that the assumptions of the RD design were violated in several US House elections. The methodological aspects of RD designs, and validity tests will be discussed in section 3. In this section, I will first briefly discuss some notable examples of papers where RD designs have been applied to elections to study the effect of the electee's gender on a number of outcome variables.

Ferreira and Gyourko (2014) use a data set on mayoral elections in the US to study whether gender affects policy outcomes and whether the election of a female mayor has a political spillover effect on the success on other females competing in elections for public office. Although the authors find that female mayors are more likely to be reelected than male mayors, they do not find any significant effect on a range of other outcome variables, including the size of local government, the allocation of public resources, and crime rates. Furthermore, they conclude that having a woman as mayor does not increase female political participation, because it does not improve the electoral success of other females—although their definition of political participation as electoral success is somewhat problematic.

A similar study is conducted by Brollo and Troiano (2016), who analyse the effects of the election of a female mayor in Brazil on a range of outcome variables by applying a RD design to close elections. They find that female mayors are less likely to face corruption charges, but they are also less likely to be re-elected. Although Brollo and Troiano (2016) discuss the implications of their research on female political participation, they do not investigate whether the election of a female mayor itself raises participation or induces other women to run for office.

Another interesting line of research employs RD designs to investigate whether female politicians are outperformed by their male competitors. For example, Anastasopoulos (2016) applies a RD design to primary elections for the US House of Representatives in the period 1982–2012 to estimate whether a gender penalty exists in terms of campaign funding or general election voting shares. He concludes that this is not the case, even finding some evidence that women enjoyed a slight electoral advantage.

Baltrunaite et al. (2019) provide more explicit evidence for the existence of a preference for female politicians among the electorate in Italy. Their application of a RD design is however different from the examples discussed above, because they do not exploit a first-past-the-post electoral system. Instead, Baltrunaite et al. (2019) analyse a Italian law that introduced 'double preference voting' only in municipalities with over 5000 residents. The results of this study show that 'if voters are given the option of casting a preference vote for one candidate of each gender, they do select female candidates more often.' At least in Italian municipal elections, outright discrimination among the electorate does not seem to act as a barrier to female political success. On the contrary, female politicians might enjoy a (slight) competitive advantage compared to men, a conclusion which has also been drawn in the Canadian context (see for example Bashevkin 2011; J. H. Black and Erickson 2003; Thomas and Bodet 2013).

3 Identification strategy

The main challenge of applied econometric research is the credible identification of causal effects. Without a valid research design and an identification strategy with clearly articulated assumptions backed up by empirical evidence, standard regression analysis will generally not provide any persuasive causal estimates. Since the 'credibility revolution', design-based research has become the standard of empirical economics (Angrist and Pischke 2010). Applied economists have increasingly turned to carefully implemented quasi-experimental methods where causal inference is justified based on explicitly stated assumptions backed by empirical evidence.

In the following sections, I will first provide a brief explanation of the reason why causal inference is fundamentally problematic. Based on this discussion, I will show how a regression discontinuity (RD) design can be used to obtain a credible estimate of the causal effect of the election of a female politician to the Canadian House of Commons on the number of female candidates during the following election. I will pay explicit attention to the key identifying assumption of the RD design, and discuss three empirical tests that can provide evidence for the validity of this assumption.

3.1 The fundamental problem of causal inference

Before the fundamental problem of causal inference can be addressed, it is important to have a formal definition of causality. In economics, the most common definition of causality is based on the potential outcome framework popularised by Rubin (1974). In the Rubin Causal Model, each unit is considered to face multiple potential outcomes associated with that unit's exposure to the treatment (Angrist and Pischke 2008; Imbens and Rubin 2010; Rubin 2005; Splawa-Neyman et al. 1990). In the context of this thesis, this means that for each electoral district i during election t, there will be a potential number of female candidates Y_{1it} if a woman had won the previous election ($D_{it-1} = 1$), and a potential number of female candidates Y_{0it} if a woman had not won the previous election ($D_{it-1} = 0$). The potential number of female candidates in each electoral district can thus be defined as:

$$\text{Potential outcomes} \equiv \begin{cases} Y_{1it} & \text{if} \quad D_{it-1} = 1 \\ Y_{0it} & \text{if} \quad D_{it-1} = 0 \end{cases}$$

In the Rubin Causal Model, the causal effect of the treatment is then simply defined as the difference between the two potential outcomes: $Y_{1it} - Y_{0it}$.

Unfortunately, this causal effect can never be observed directly since the different treatment statuses are mutually exclusive. For each unit, only a single potential outcome associated with the treatment received by that unit will be observed. For example, in electoral districts where a woman won the previous election, only Y_{1it} will be observed, while the value of Y_{0it} will remain fundamentally unknown. This mutual exclusivity of treatment status gives rise to the 'The Fundamental Problem of Causal Inference', a notion first coined by Holland (1986). With this notion, Holland (1986) emphasized that the identification of a causal effect of treatment $Y_{1it} - Y_{0it}$ for each individual unit is fundamentally impossible, because only one of the potential outcomes will materialise and be observable.

Although the individual causal effect of a treatment at the level of the unit $Y_{1it} - Y_{0it}$ remains fundamentally unknown, it is still possible to obtain credible estimates of an *average* causal effect of the treatment for the population of interest: $\mathbb{E}[Y_{1it}] - \mathbb{E}[Y_{0it}]$. Since these averages can both be observed, in principle, the average treatment effect (ATE) can be estimated. Regrettably, a simple comparison of average outcomes will only provide an unbiased estimate of the ATE under very specific—and often unrealistic—conditions.

3.1.1 Selection bias

To see under which conditions a comparison of average outcomes can be used for the identification of the average treatment effect, note that the observed outcome Y_{it} can be expressed as:

$$Y_{it} = Y_{0it} + D_{it-1}(Y_{1it} - Y_{0it})$$

Using this formula, the observed average outcomes for the treated and untreated groups of units $\mathbb{E}[Y_{it}|D_{it-1}]$ can be expressed as:

$$\mathbb{E}[Y_{0it}|D_{it-1} = 0]$$

if $D_{it-1} = 0$, and

$$\mathbb{E}[Y_{0it}|D_{it-1} = 1] + \mathbb{E}[Y_{1it} - Y_{0it}|D_{it-1} = 1]$$

if $D_{it-1} = 1$.

In the latter equation, $\mathbb{E}[Y_{1it} - Y_{0it}|D_{it-1} = 1]$ denotes the average effect of the treatment on the treated (ATT). Differences in the observed average outcomes will however not only depend on this treatment effect, but also on any potential dissimilarities in the 'baseline' expected outcome $\mathbb{E}[Y_{0it}]$ between the treated and untreated groups of units:

$$\begin{aligned}
&\underbrace{\mathbb{E}[Y_{it}|D_{it-1} = 1] - \mathbb{E}[Y_{it}|D_{it-1} = 0]}_{\text{Comparison of observed outcomes}} = \\
&\underbrace{\mathbb{E}[Y_{1it} - Y_{0it}|D_{it-1} = 1]}_{\text{ATT}} + \\
&\underbrace{\mathbb{E}[Y_{0it}|D_{it-1} = 1] - \mathbb{E}[Y_{0it}|D_{it-1} = 0]}_{\text{Selection bias}} \qquad (1)
\end{aligned}$$

It is the last part of this equation, known as the *selection bias*, which makes the identification of the average treatment effect challenging. In most situations, there are good reasons to believe that selection bias will render a naive comparison of observed outcomes unsuitable for the identification of a causal effect. The reason for this is that on average, the units receiving treatment are likely to differ in various ways from the untreated units. If Y_{0it} is correlated with any factors or characteristics in which the treated and untreated groups are dissimilar, the observed difference in outcomes will not only be attributable to the treatment effect, but also to a selection bias (Angrist and Pischke 2008, p. 61).

In the context of this thesis, there can be several reasons to expect some form of selection bias. On average, the districts where women won the last election could very well differ from other district in terms of the share of female voters, the percentage of voters holding more traditional beliefs about gender roles, or the level of female participation in the labour market or local politics. If such factors also affect the number of female candidates in the electoral district, differences in the average observed number of female candidates are not only attributable to the treatment status, but to the average differences in these confounding factors.

3.1.2 The gold standard: randomized experiments

Ideally, the problem of selection bias can be solved through randomised assignment of the treatment, which would make the potential outcomes Y_{1it}, Y_{0it} independent of the treatment status. This independence eliminates the selection bias because it results in: $\mathbb{E}[Y_{0it}|D_{it-1}] = \mathbb{E}[Y_{0it}]$. With randomly assigned treatment, a comparison of the observed average outcomes will provide credible estimates of the average treatment effect. For this reason, the randomised experiment is often regarded as the gold standard of causal inference—and as a benchmark for research based on other identification strategies.

Random assignment of the treatment status is however not always possible or feasible. In the context of this study, it is somewhat obvious that it is not possible to randomly determine whether elections are won by female or male candidates. Unless the electorate votes at random, only a constitutional overhaul would be capable of creating a situation where seats in the House of Commons are assigned randomly to male and female politicians.

An alternative to randomised assignment of the treatment variable in real-life situations, is a simulation of the mechanism of interest in a controlled laboratory experiment. However, given the the long-term nature, the importance, and the complexity of real-life parliamentary elections it is almost certain that the results obtained in an artificial laboratory setting will lack external validity: what happens in a laboratory will likely be incomparable to the dynamics of real parliamentary elections.

3.1.3 Dealing with dependence: match & control

In principle, the absence of random assignment of the treatment can be addressed by the addition of control variables to regression analysis or by means of statistical techniques such as propensity score matching. These identification strategies all rely on the *conditional independence assumption* (Angrist and Pischke 2008, p. 69). The implication of this assumption is that causal inference based on these strategies is only credible and unbiased if the potential outcome Y_{0it} is independent of the treatment *conditional* on the vector of covariates $\boldsymbol{X}_{it}$ added to the analysis/used to perform the matching:

$$\mathbb{E}[Y_{0it}|D_{it}, \boldsymbol{X}_{it}] = \mathbb{E}[Y_{0it}|\boldsymbol{X}_{it}]$$

This conditional independence assumption is a very strong assumption of which the validity cannot be tested by empirical evidence alone. In almost any situation, there is a distinct risk that unknown or unobservable confounding factors cause a violation of this assumption. Given the complexity of many of the topics studied in applied econometrics, selection bias will remain a distinct possibility when causal inference is based on the conditional independence assumption, because any confounding factor that is not included in $\boldsymbol{X}_{it}$ can result in omitted variable bias (Angrist and Pischke 2008, p. 59–64).

3.2 Regression discontinuity designs

Up until now, credible identification of the causal effect of female electoral success on the number of female candidates in the next election seems to remain nearly impossible: the potential outcomes are unlikely to be independent of the treatment status and control- or matching strategies will almost certainly suffer from bias due to unobserved or unobservable confounding factors. Fortunately, the first-past-the-post electoral system used in the Canadian parliamentary elections can be exploited for credible causal inference through the application of a regression discontinuity (RD) design.

In their essence, RD designs are based on the idea that in situations where the treatment status changes discontinuously at a certain cut-off point, the observed outcomes of the units that are located just below this cut-off point—and thus did not receive treatment—can be used to construct a counterfactual for the units that are located slightly above the cut-off point—and did receive treatment. In the following section, I will provide a more detailed and formal account of the ideas behind RD designs, the key identifying assumption on the continuity of the potential outcome at the cut-off point, and a number of validity tests that can provide empirical evidence in support of this identifying assumption.

3.2.1 The good and bad of deterministic treatment

Any RD design is based on a discontinuity in treatment status as a function of some observable forcing variable. In the case of this thesis, treatment status can be seen as a discontinuous *and* deterministic function of the voting share obtained by the strongest female candidate.[2] If and only if the strongest female candidate in an electoral district obtained more votes than any of her competitors during the previous election, than the district is treated—*i.e.* a woman has won the previous election. More specifically, treatment D_{it-1} is determined in each district as:

$$D_{it-1} = \begin{cases} 1 & \text{if} \quad \tilde{v}_{it-1} > 0 \\ 0 & \text{if} \quad \tilde{v}_{it-1} < 0 \end{cases}$$

where $\tilde{v}_{it-1}$ is the normalised forcing variable defined as the difference between the vote share of the strongest female candidate v_i during election $t-1$ and the vote share of her strongest opponent c_i—who can be either male or female.

An important consequence of the deterministic relationship between the treatment status and the forcing variable is that the conditional independence assumption mentioned earlier will be trivially satisfied once controls for the forcing variable are added to the analysis. Conditional on the forcing variable, there is no variation in the treatment status left, and potential outcomes will thus be independent of the treatment status (Imbens and Lemieux 2008).

The flip-side of the coin is however that the deterministic manner in which treatment depends on the forcing variable also results in the violation of the *common support* assumption (Heckman et al. 1998; Rosenbaum and Rubin 1983). There exists no value of $\tilde{v}_{it-1}$ for which both treated and untreated units can be observed. In each of the electoral districts with $\tilde{v}_{it-1} > 0$, female candidates won the previous election, whereas in the districts with $\tilde{v}_{it-1} < 0$, elections were exclusively won by men. For any specific value of $\tilde{v}_{it-1}$, it is therefore impossible to compare the observed outcomes of treated and untreated districts, since for each value of $\tilde{v}_{it-1}$ either treated or untreated districts will exclusively be observed.

In RD design-based research, the common support assumption is substituted by a different, but relatively weak, identifying assumption. Instead of comparing the outcomes of treated and untreated units with common values of the forcing variable $\tilde{v}_{it-1}$, the outcomes of untreated and treated units *close* to the cut-off point $\tilde{v}_{it-1} = 0$ are used to obtain an credible estimate of the treatment effect.

3.2.2 A close cousin of randomized experiments

The first authors to introduce a RD design were Thistlethwaite and D. T. Campbell (1960), who proposed a somewhat alternate way of looking at situations where treatment is a deterministic and discontinuous function of a

[2] Since treatment is determined in a deterministic manner, this case represents a *sharp* RD. If treatment status would instead show variation at both sides of the cut-off (while still showing a discontinuity at the cut-off), *fuzzy* RD design-based causal inference would still be possible, although in a slightly different manner.

single and observable variable. In their study on the effect of receiving a Certificate of Merit on students' academic performance, they argued that a RD analysis can be regarded as a substitute for a 'true' experiment by focusing on the observations immediately around the cut-off point where treatment status is discontinuous (see also D. T. Campbell and Stanley 1963).

Lee and Lemieux (2010) follow this line of reasoning and regard RD design as a 'close cousin' of randomised experiments, providing a more formal discussion of the similarity of RD designs and experiments. Conceptually, in both an experimental setting, as well as in a RD design, treatment can be expressed as a deterministic and discontinuous function of a forcing variable. The main difference is however that in randomised experiments, treatment status can be thought of as a function of a forcing variable which is *itself* randomly distributed over the studied population—and thus independent of the potential outcomes.

In contrast, in most RD designs the forcing variable will not be randomly distributed, and is therefore likely to be correlated with the potential outcomes. For example, in the Thistlethwaite and D. T. Campbell (1960) study, students received a Certificate of Merit if they had a test score above a certain threshold. The students' test score will almost certainly be correlated with their potential future academic performance. Comparisons of the observed outcomes for the groups who do and do not receive treatment will not only capture the causal effect of the treatment, but also a type of selection bias.

To prevent this correlation between the forcing variable and the potential outcomes to bias the estimate of the causal effect, a RD design focuses on a comparison of the observations where the value of the forcing variable is just below, or just above the cut-off point. More specifically, in a RD design the causal treatment effect of interest can be identified comparing the observed outcomes as the forcing variable approaches the cut-off point:

$$\begin{aligned}&\underbrace{\lim_{\tilde{v}_{it-1}\downarrow 0}\mathbb{E}[Y_{it}|\tilde{v}_{it-1}] - \lim_{\tilde{v}_{it-1}\uparrow 0}\mathbb{E}[Y_{it}|\tilde{v}_{it-1}]}_{\text{Comparison of outcomes at the cut-off point}} = \\ &\underbrace{\mathbb{E}[Y_{1it} - Y_{0it}|\tilde{v}_{it-1} = 0]}_{\text{LATE}} + \\ &\underbrace{\lim_{\tilde{v}_{it-1}\downarrow 0}\mathbb{E}[Y_{0it}|\tilde{v}_{it-1}] - \lim_{\tilde{v}_{it-1}\uparrow 0}\mathbb{E}[Y_{0it}|\tilde{v}_{it-1}]}_{\text{Confounding discontinuity bias}} \qquad (2)\end{aligned}$$

Analogous to Equation 1, this comparison of the observed outcomes as the forcing variable approaches the cut-off point is composed of two distinct elements. First, there is the local average treatment effect (LATE) which measures the effect of female electoral success for the districts close to the cut-off point $\tilde{v}_{it-1} = 0$. The second component of the comparison is a potential bias term, which I will call the *confounding discontinuity bias*.

There is an important difference between the selection bias of Equation 1 and the confounding discontinuity bias appearing in Equation 2 which makes the condition for unbiased RD estimates considerably 'weaker' than the conditional independence assumption. Whereas selection bias disappears only when the expected (conditional) potential outcome $\mathbb{E}[Y_{0i}|D_{it-1}]$ is the same for the treated and untreated groups, the confounding discontinuity bias already disappears when the expected potential outcome $\mathbb{E}[Y_{0i}|\tilde{v}_{it-1}]$ behaves continuously around the cut-off point $\tilde{v}_{it-1} = 0$.

The identifying assumption of a RD design does thus not require the potential outcome $\mathbb{E}[Y_{0i}]$ to be (conditionally) independent of the treatment status. Instead, as Hahn et al. (2001) show, the identifying assumption of causal inference in a RD design is:[3]

$\mathbb{E}[Y_{0it}|\tilde{v}_{it-1}]$ is continuous at $\tilde{v}_{it-1} = 0$.

Under this assumption, there will be no confounding discontinuity bias, because $\lim_{\tilde{v}_{it-1}\downarrow 0}\mathbb{E}[Y_{0it}|\tilde{v}_{it-1}] = \lim_{\tilde{v}_{it-1}\uparrow 0}\mathbb{E}[Y_{0it}|\tilde{v}_{it-1}]$. Consequentially, the local average treatment effect (LATE) on the treated around the cut-off $\tilde{v}_{it-1} = 0$ can be identified as:

$$\begin{aligned}&\mathbb{E}[Y_{1it} - Y_{0it}|\tilde{v}_{it-1} = 0] = \\ &\lim_{\tilde{v}_{it-1}\downarrow 0}\mathbb{E}[Y_{it}|\tilde{v}_{it-1}] - \lim_{\tilde{v}_{it-1}\uparrow 0}\mathbb{E}[Y_{it}|\tilde{v}_{it-1}]\end{aligned}$$

In the following section, I will discuss statistical tests that can provide empirical evidence for the validity of the identifying assumption of a RD design. In section 4, I will then examine the techniques that can be used to estimate the local average treatment effect in more detail.

3.3 Validity tests

At first sight, the identifying assumption of a RD design seems equally untestable as the conditional independence assumption mentioned earlier, because $\mathbb{E}[Y_{0it}|\tilde{v}_{it-1} > 0]$ will remain fundamentally unobservable. Lee and Lemieux (2010) show however that as long as units have no *precise* control over the forcing variable $\tilde{v}_{it-1}$, treatment D_{it-1} can be regarded 'as good as' randomly distributed closely around the cut-off. The result of such local randomisation of treatment is that the distribution of the potential outcomes $\mathbb{E}[Y_{1it}|\tilde{v}_{it-1}]$ and $\mathbb{E}[Y_{0it}|\tilde{v}_{it-1}]$ will be continuous in any value of $\tilde{v}_{it-1}$.

[3] As both Hahn et al. (2001) and Lee and Lemieux (2010) point out, this is the bare minimum assumption of an RD. However, in both papers, the stronger assumption is made that $\mathbb{E}[Y_{0it}|\tilde{v}_{it-1}]$ and $\mathbb{E}[Y_{1it}|\tilde{v}_{it-1}]$ are continuous for any value of $\tilde{v}_{it-1}$. The intuition here is that assuming that only one of the conditional expectation functions is continuous at a particular point seems not very natural and reasonable—especially if this particular point is a normalised forcing variable as is the case in this thesis. Strictly speaking, this stronger assumption is however not necessary for the identification of a causal effect.

Intuitively, the local randomisation of treatment can be explained as follows: as long as the forcing variable is at least partly determined by a stochastic process, units with exactly identical (observed and unobserved) characteristics can end up with different values of $\tilde{v}_{it-1}$ by chance. This means that if it would be possible to hold a close mixed-gender election multiple times under the exact same circumstances—*i.e.* with the same candidates, in the same district, and at the same time—the presence of a stochastic effect on the candidates' voting share implies that in some of these elections the female candidate would win, and in others she would lose.

More formally, following Lee and Lemieux (2010), in the presence of a stochastic component η_i affecting the forcing variable, the RD design can expressed as the set of equations:

$$Y_{it} = \tau D_{it-1} + \gamma_1 \boldsymbol{X}_i + \epsilon_i$$

$$\tilde{v}_{it-1} = \gamma_2 \boldsymbol{X}_i + \eta_i$$

where D_{it-1} is the treatment variable and $\boldsymbol{X}_i$ is a vector of the systematic covariates affecting the outcome or the forcing variable (or both). Lee and Lemieux (2010) then define imprecise control as the situation where conditional on $\boldsymbol{X}_i$ and ϵ_i the density of η_i—and consequentially $\tilde{v}_{it-1}$—is continuous. In the context of this thesis, such continuity would mean that for each district, the probabilities of having a $\tilde{v}_{it-1}$ just below or above 0 are the same (Lee 2008, p. 679).

Using Bayes' theorem, Lee and Lemieux (2010) formally show that the continuity of the conditional density of $\tilde{v}_{it-1}$ implicates that the potential outcome Y_{0it}—as a function of $\boldsymbol{X}_i$ and ϵ_i—will also be continuous in $\tilde{v}_{it-1}$. In the absence of precise control over the forcing variable around the cut-off point, there is therefore a (refutable) prediction that the identifying assumption of the RD design holds, because treatment is locally randomised (Lee and Lemieux 2010).

This relationship between the lack of precise control over the forcing variable and the continuity of potential outcomes also sheds light on a number of ways in which the validity of a RD design can be tested. First, an inspection of how the forcing variable is determined for each district can provide some initial arguments in favour of the validity of RD designs. In the case of free democratic elections there is an strong expectation that the share of votes obtained by each candidate cannot be perfectly controlled—at least in the absence of voting fraud. Secondly, there are three widely used empirical tests that can provide evidence for the validity of RD designs (Eggers, Fowler, et al. 2015; Lee and Lemieux 2010). In the next sections, I will discuss these tests.

3.3.1 Forcing variable density test

Although it seems reasonable to expect that voting shares cannot be precisely controlled in a well-functioning democracy, direct proof for the lack of precise control will remain unattainable—because the forcing variable for each district is observed only once every election. Indirect evidence for the lack of precise control over the forcing variable can however be found based on an examination of the observed density of the forcing variable around the cut-off. McCrary (2008) was the first to introduce a formal test based on the idea that in the absence of precise control, the density of the forcing variable is expected to be continuous at the cut-off point. A discontinuity suggests that there is complete manipulation of the forcing variable. Although this would not necessarily mean that the RD design would be invalid—*i.e.* the expected potential outcomes could still be continuous conditional on the forcing variable at the cut-off—it would nonetheless call the credibility of the RD design into question.

Cattaneo, Jansson, et al. (2020) have more recently proposed a non-parametric density estimator based on local polynomial techniques for which they have provided easy-to-use software packages in `Stata` and `R` (see also Cunningham 2021, sec. 6.3.1).[4] In contrast to the original McCrary (2008) density test, this newer non-parametric manipulation test does not require pre-binning. As a result, the statistical power of Cattaneo, Jansson, et al.'s (2020) test is likely superior. To investigate whether there is precise control over of the forcing variable, I will present both density tests in section 7.1.

3.3.2 Covariate balance test

Related to such tests on potential discontinuities of the density of the forcing variable are tests that examine whether the treated and untreated units near the cut-off are similar in terms of baseline covariates. As discussed above, the local randomisation of treatment close around the cut-off should not only result in the continuity of the potential outcomes at the cut-off, but also in a continuous distribution of baseline covariates (or placebo outcomes known to be unaffected by the treatment). Lee and Lemieux (2010) note that discontinuities in any pretreatment covariates would cast doubt on the identifying assumption of RD designs. Intuitively, discontinuities in covariates which are known not to be affected by the treatment would not only raise questions on the local randomisation of treatment around the cut-off, but it would also make the presence of a confounding discontinuity bias around the cut-off plausible.

Implementing this test is relatively straightforward: potential discontinuities in the observed baseline covariates at the cut-off point can (and should) be analysed by application of the same RD methods and techniques used in the main analysis (Cattaneo, Idrobo, et al. 2019, p. 90). Each covariate is thus analysed as if it is the outcome of interest. Any non-zero effect raises doubt on the validity

[4] See https://rdpackages.github.io/.

of the RD design—although it is important to note that with a confidence interval of 95%, there is a 5% chance of a false identification of a non-zero effect. The results of this test are presented in section 7.2.

3.3.3 Placebo discontinuities test

The third validity test is a kind of placebo test suggested by Imbens and Lemieux (2008), that analyses whether the outcome variable is discontinuous at values of the forcing variable where there is no variation in the treatment status. If such discontinuities are found to exist, this would suggest that the expected potential outcomes conditional on the forcing variable around the cut-off could also be discontinuous. In the context of this thesis, this falsification test could for example be implemented by estimating the treatment effect around the placebo cut-off point of $\tilde{v}_{it-1} = 0.1$—where both the districts to the left and right had a female winner. If the identifying assumption of the continuity of the potential outcomes holds for all values of $\tilde{v}_{it-1}$, this should mean that no treatment effect should be found at the artificial cut-off points. After all, around this point, there is no actual change in treatment, and the outcome variable should thus be continuous. The results of this validity test are presented in section 7.4.

4 Estimation and inference

Broadly speaking, two main approaches are possible for the estimation of treatment effects in a RD design. First, treatment effects can be estimated with a global or parametric regression models where assumptions have to be made on the functional form of the relationship between the forcing variable and the outcome variables—*i.e.* whether the relationship between the forcing variable and the potential outcomes are linear, quadratic, or logarithmic. Since the true functional form of this relationship is almost always unknown, there will be a potential for misspecification bias when treatment effects are estimated using parametric models. In RD designs, misspecification bias can be especially problematic, since the treatment effect is estimated as the difference between two *local* values of a function at its boundary—*i.e.* at the cut-off point. Although global polynomial estimates will generally provide an accurate global approximation of a function, this is not always the case for boundary points due the Runge's phenomenon (Cattaneo, Idrobo, et al. 2019, p. 40).

To reduce the likelihood of biased estimates due to misspecified functional forms, local non-parametric estimation techniques have became increasingly popular in RD designs. Since Hahn et al.'s (2001) work on RD designs, the use of local linear (or low-order polynomial) regression models has been widely recommended. These non-parametric methods are based on the idea that over a sufficiently small interval, any continuous function can be accurately approximated by a linear function. Thus, by running linear regressions on observations in close proximity to the cut-off point, misspecification bias can be minimised. Furthermore, discarding observation far removed from the cut-off point reduces the sensitivity to observations with extreme values of $\tilde{v}_{it-1}$—which are unlikely to be informative for the estimation of $\lim_{\tilde{v}_{it-1}\downarrow 0} \mathbb{E}[Y_{it}|\tilde{v}_{it-1}]$ and $\lim_{\tilde{v}_{it-1}\uparrow 0} \mathbb{E}[Y_{it}|\tilde{v}_{it-1}]$, or could even result in biased estimates (Gelman and Imbens 2019). Unfortunately, local non-parametric approaches do have an obvious drawback: discarding observations further removed from the cut-off point reduces the sample size, which can consequentially deteriorate the precision of the estimates.

Estimation of treatment effects in RD designs thus involves a bias-variance trade-off: global polynomial approaches are more likely to produce biased estimates, while local non-parametric approaches suffer from a reduction in the sample size and a loss in precision. In this thesis, I will estimate the treatment effect using both approaches. Furthermore, I will pay explicit attention to the sensitivity of the estimates to changes in the bandwidth and functional specification. A high sensitivity to alternative bandwidth selections or functional forms would call for a more cautious interpretation of the estimates.

4.1 The global parametric approach

The easiest way to obtain an estimate of the treatment effect

$$\tau = \lim_{\tilde{v}_{it-1}\downarrow 0} \mathbb{E}[Y_{it}|\tilde{v}_{it-1}] - \lim_{\tilde{v}_{it-1}\uparrow 0} \mathbb{E}[Y_{it}|\tilde{v}_{it-1}]$$

is to estimate two separate regression models on both sides of the cut-off point:

$$\begin{aligned} Y_{it} &= \alpha_l + f_l(\tilde{v}_{it-1}) + \epsilon_{it} \quad \text{for} \quad \tilde{v}_{it-1} < 0 \\ Y_{it} &= \alpha_r + f_r(\tilde{v}_{it-1}) + \epsilon_{it} \quad \text{for} \quad \tilde{v}_{it-1} > 0 \end{aligned}$$

where α_l and α_r are the intercepts of the regressions and $f(\cdot)$ is the functional form of the relationship between the forcing variable and the outcome. This can be done using ordinary least squares (OLS) estimation methods. Since the cut-off point in this case is $\tilde{v}_{it} = 0$, the treatment effect can be calculated as $\hat{\tau} = \alpha_r - \alpha_l$.

A drawback of this simple method is that it does not directly produce standard errors. To avoid having to calculate the standard errors manually, the two separate regression models can be combined in a single pooled regression model:

$$Y_{it} = \alpha + \tau D_{it-1} + f(\tilde{v}_{it-1}) + D_{it-1} f(\tilde{v}_{it-1}) + \epsilon_{it}$$

where τ is the causal effect of interest, and the interaction term $D_{it-1} f(\cdot)$ allows for differences in the regression function at both sides of the cut-off point. As mentioned before, a potentially problematic aspect of the global parametric approach is finding the correct approximation of the underlying functional form of $f(\cdot)$. In the next sections, I will discuss two strategies that can be used to try to determine this underlying functional form.

4.1.1 Graphical inspection

As a starting point, graphical inspection of the data can provide an overall idea of the functional form of $f(\cdot)$. A simple scatter plot can be used first to inspect the raw data, but far more common in RD settings is to present a plot with (i) local sample means for binned data, and (ii) global polynomial fits estimated separately at both sides of the cut-off point (Cattaneo, Idrobo, et al. 2019; Lee and Lemieux 2010). As with virtually any aspect of applied econometrics, the canonical RD plot with binned data involves choices which can (drastically) affect the resulting graph: (i) the location of the bins, and (ii) the number of bins (Cattaneo, Idrobo, et al. 2019).

On the location of the bins, there are two main options: the bins can be constructed to be of equal length in terms of the forcing variable, or the bins can be constructed to include a similar number of observations. An important advantage of the latter option is that the similar number of observations in each bin results in local means that are estimated with similar precision, while it also provides a first glimpse at the distribution of the observations (Cattaneo, Idrobo, et al. 2019, p. 25).

With respect to the number of bins, there is a whole range of possibilities. At its core, the number of bins involves a trade-off in bias and variance (Lee and Lemieux 2010, p. 308). As the number of bins is inversely related to the number of observations in each bin, the estimated local averages lose precision when more bins are used. Reducing the number of bins will however increase the risk of bias because the local sample means can fail to provide a good approximation for the underlying functional form. Calonico, Cattaneo, and Titiunik (2015) propose two data-driven selectors that can be used to choose the number of bins: the integrated mean squared error (IMSE) method and the the mimicking variance method. The IMSE method is designed to 'trace out' the underlying regression function while the mimicking variance method is constructed to replicate the variability of the underlying data. For the main results, I will present RD plots created with both methods.

4.1.2 Sensitivity analysis & cross-validation

In addition to graphical representation of the RD design, it is common practice to provide estimates for various functional specifications to showcase the sensitivity of the results. As usual, estimates that are robust to alternative orders of polynomials suggest no substantial misspecification bias, whereas a high sensitivity calls for closer inspection and a more cautious interpretation of the results.

More formal criteria for model selection are also available. A commonly used method to determine the quality of a model relative to models with different specifications is to calculate the Akaike (1974) Information Criterion (AIC), defined as $2K - 2\ln(\hat{L})$—where k is the number of parameters of the model and $\hat{L}$ is the likelihood of the function (*e.g.* D. A. Black et al. 2007; van Kippersluis et al. 2011). This AIC is based on the trade-off between the goodness-of-fit of a specific model and that model's simplicity. The model with the lowest AIC is considered to be of the highest quality. It must however be stressed that indicators such as the AIC measure *global* goodness-of-fit, whereas the purpose of a RD design is to obtain *local* estimates of the function at the cut-off point (Gelman and Imbens 2019). The value of the AIC is therefore limited when determining the quality of a model at the boundaries points.

4.2 The local non-parametric approach

Although global parametric approaches were extensively used in early applications of RD designs, far more common in contemporary studies are local non-parametric estimation methods. The main intuition behind the use of local non-parametric approaches is that discarding observations far removed from the cut-off point and estimating low-order polynomial approximations of the regression function over a small interval reduces the risks of misspecification bias (Cattaneo, Idrobo, et al. 2019; Gelman and Imbens 2019; Hahn et al. 2001; Lee and Lemieux 2010). In this thesis, I will primarily focus on local *linear* regression models—which have become the standard of RD designs.

In a local non-parametric setting, only observations within a certain bandwidth h around the cut-off are used to estimate $\lim_{\tilde{v}_{it-1}\downarrow 0} \mathbb{E}[Y_{it}|\tilde{v}_{it-1}]$ and $\lim_{\tilde{v}_{it-1}\uparrow 0} \mathbb{E}[Y_{it}|\tilde{v}_{it-1}]$. In principle, these estimates can be obtained by means of similar OLS regression techniques as discussed in the section on the global parametric approach—only dropping the observations outside the bandwidth. The main challenge of the local non-parametric approach is selecting the appropriate bandwidth h. Here, the bias-variance trade-off is again important.

To reduce the risk of biased estimates of the regression function at the boundaries, it is important to use a sample that is as 'local' as possible. However, reducing the sample size also reduces the statistical power. Choosing a bandwidth that balances bias and variance is thus important. Furthermore, as Cattaneo, Idrobo, et al. (2019, p. 45) note, the value of h 'directly affects the properties of local polynomial estimation and inference procedures, and empirical findings are often sensitive to its particular value'. In the next section, I will discuss a number of bandwidth selection techniques that I will use in this thesis. After this discussion, I will discuss a slightly more sophisticated extension of local polynomial regression techniques where more weight is given the closer observations are to the cut-off (Angrist and Pischke 2008, p. 256).

4.2.1 Bandwidth selection

A relatively simple method to select the bandwidth for the local polynomial regression models is the Ludwig and D. L. Miller (2005, LM from hereon) cross-validation procedure (Imbens and Lemieux 2008). The idea behind the LM cross-validation procedure is to find the bandwidth h at which the local regression function provides the best fit for observations at the boundary of the local sub-sample of length h. To determine to what extent local regression models provide a good fit given a certain bandwidth h, the following procedure is followed separately for observations to the left and to the right of the cut-off point. At the left of the cut-off point ($\tilde{v}_{it-1} < 0$), the value of $\hat{Y}_i(\tilde{v}_{it-1})$ is predicted for each observation i by means of a local linear regression model which is estimated using observations between $\tilde{v}_{it-1} - h \leq \tilde{v} < \tilde{v}_{it-1}$. Note that the observation i itself is *not* used to estimate the regression function. The same procedure is then followed for observations to the right of the cut-off point.

After the predicted value $\hat{Y}(\tilde{v}_{it-1})$ is calculated for each observation, a goodness-of-fit indicator can be determined as the squared difference between the observed value Y_i and the predicted value $\hat{Y}(\tilde{v}_{it-1})$. The purpose of the cross-validation procedure is to find the bandwidth h at which the mean squared error is minimised. Formally, the optimal bandwidth using this cross-validation procedure can be calculated as:

$$h_{CV}^{opt} = \arg\min_{h} \frac{1}{n} \sum_{i}^{n} (Y_i - \hat{Y}(\tilde{v}_{it-1}))^2$$

Although computationally intensive, this LM cross-validation optimal bandwidth can be easily estimated with statistical software. Following Imbens and Lemieux (2008), only the predicted values of $\hat{Y}(\tilde{v}_{it-1})$ with $\tilde{v}_{it-1}$ above (below) the median value of $\tilde{v}_{it-1}$ to the left (right) of the cut-off point will be used in the cross-validation procedure, since the observations located further away from the cut-off point are unlikely to be very informative for point estimation at the boundary of the cut-off point.

Alternative methods to determine the optimal bandwidth are the so-called 'plug-in' methods, where the optimal bandwidth is mathematically expressed as a function of a number of characteristics of the data. The objective of these methods is to find a bandwidth that optimises the bias-variance trade-off. Imbens and Kalyanaraman (2012, IK from hereon) proposed a now widely used formula for an optimal bandwidth in RD setting that attempts to minimise the asymptotic mean squared error (AMSE) (see also Imbens and Kalyanaraman 2009). In the Appendix, I discuss the IK bandwidth selector in more detail and I will present the steps used to compute the optimal bandwidth.

More recently, Calonico, Cattaneo, and Titiunik (2014, from hereon CCT) have proposed a novel mean squared error (MSE) optimal approach for bandwidth selection, noting that earlier methods often led to too 'large' bandwidth choices (Calonico, Cattaneo, and Farrell 2020). Complementary to this MSE-optimal bandwidth procedure which results in MSE-optimal *point estimates* of the treatment effect, Calonico, Cattaneo, and Farrell (2018) have also introduced a procedure to determine the coverage error rate (CER) optimal bandwidth, which is specifically constructed for *inference*—but not for point estimation (Cattaneo, Idrobo, et al. 2019, p. 72). In Section 4.3, inference will be discussed in more detail. Both the MSE-optimal and the CER-optimal bandwidths are easily calculated using the `Stata` or `R` software packages developed by Calonico, Cattaneo, Farrell, and Titiunik (2017).[5]

4.2.2 Kernel functions

In the context of a RD design, it can be useful to adopt a weighting scheme for the local regression models to ensure that more weight is given to the observations closest to the cut-off point (Cattaneo, Idrobo, et al. 2019, p. 41). Such weighting function is known as the kernel $K(\cdot)$. With a given kernel function, it is possible to estimate a local regression model which algebraically equals a Weighted Least Squares (WLS) estimator. The optimisation problem for a local linear regression model with bandwidth h is:

$$\arg\min_{\alpha,\tau,\beta_l,\beta_r} \sum_{i}^{N} K(\frac{\tilde{v}_{it-1}}{h})(Y_i - \hat{Y}_i(\alpha,\tau,\beta_l,\beta_r))^2$$

with

$$\hat{Y}_i(\alpha,\tau,\beta_l,\beta_r) = \alpha + \tau D_i + \beta_l \tilde{v}_{it-1} + \beta_r D_i \tilde{v}_{it-1}$$

The most simple local linear estimate can be obtained by using a uniform kernel, which gives equal weight to all observations with $|\tilde{v}_{it-1}| \leq h$, and zero weight if $|\tilde{v}_{it-1}| > h$—in this case, the estimator boils down to OLS.

However, Cattaneo, Idrobo, et al. (2019) recommend using a triangular kernel, where:

$$K\left(\frac{\tilde{v}_{it-1}}{h}\right) \equiv \begin{cases} 1 - |\frac{\tilde{v}_{it-1}}{h}| & \text{if} \quad |\frac{\tilde{v}_{it-1}}{h}| \leq 1 \\ 0 & \text{if} \quad |\frac{\tilde{v}_{it-1}}{h}| > 1 \end{cases}$$

With such triangular kernel, the weights for observations are inversely related to the distance between the observation and the cut-off point $\tilde{v} = 0$. Intuitively, this weighting scheme is appealing because observation further removed from the cut-off point are less informative in the estimation of the boundary point $\mathbb{E}[Y_i|\tilde{v}_{it-1} = 0]$. M.-Y. Cheng

[5]See https://rdpackages.github.io/.

et al. (1997) have also shown that boundary estimates have optimal properties when triangular kernels are used in combination with local polynomial regression models (see also Fan and Gijbels 1996), and Armstrong and Kolesár (2020) note the high asymptotic efficiency of triangular kernels. For these reasons, Cattaneo, Idrobo, et al. (2019) recommend using triangular kernels in combination with their proposed MSE-optimal bandwidth. In this thesis, I will present the results of local regression models using both uniform and triangular kernels, but the preferred estimates will be based on triangular kernels due to their favourable properties.

4.3 Inference and standard errors

The final important methodological aspect is inference. In commonly used 'guides to practice', it is recommended to estimate standard errors using standard least square methods (Imbens and Lemieux 2008; Lee and Lemieux 2010). In the context of this thesis, this would mean computing heteroscedasticity-consistent (HC) standard errors clustered at the electoral district–representation order level to account for correlation of the errors over time within electoral districts.

These inference methods are perfectly suitable for the global parametric approach, yet this is not necessarily the case for the local non-parametric approach. Although Lee and Lemieux (2010) state that 'the usual standard errors can be used with local linear regressions', Calonico, Cattaneo, and Titiunik (2014) have more recently warned against the use of standard errors and confidence intervals constructed using standard OLS, because such conventional methods contain the implicit assumption that the functional specification in the non-parametric approach is correct. Since the main idea behind the non-parametric approach is to *approximate* the underlying regression function, this assumption will be violated in most cases (see also Cunningham 2021, sec. 6.2.6).

Due to the presence of misspecification in local regression models, Calonico, Cattaneo, and Titiunik (2014) recommend to combine their MSE-optimal bandwidth selector with a robust bias-corrected approach to inference (see also Calonico, Cattaneo, and Farrell 2020; Cattaneo, Idrobo, et al. 2019). This approach centers the confidence intervals at a bias-corrected estimate of the treatment effect $\hat{\tau} - \hat{\mathscr{B}}$, where $\hat{\mathscr{B}}$ is the estimated bias term of $\hat{\tau}$ as also used to calculate the MSE-optimal bandwidth. Furthermore, this robust bias procedure uses a non-conventional (and larger) estimate of the standard error which takes to variability added due to the bias estimation step into account.

Hyytinen et al. (2018) provide empirical evidence for this robust bias-corrected approach by comparing experimental estimates with RD estimates of the same treatment effect. Analysing a Finnish dataset on municipal elections where some elections could be exploited with a RD design, while other seats were assigned randomly because of tied votes, Hyytinen et al. (2018) find that the experimental estimates of a personal incumbency effect can be recovered with RD estimates, but only when Calonico, Cattaneo, and Titiunik's (2014) robust bias-corrected approach to inference is used. Conventional inference methods—*i.e.* heteroscedasticity-robust and clustered standard errors—indicated a statistically significant treatment effect where the experimental evidence showed there was none. Furthermore, Hyytinen et al. (2018) observe that the robust bias-corrected approach is less sensitive to the choice of bandwidth in their study.

Lastly, as already mentioned in Section 4.2.1, Calonico, Cattaneo, and Farrell's (2018) CER-optimal bandwidth is constructed specifically for inference. Although estimates based on CER-optimal bandwidth are not very suitable to investigate the magnitude of the treatment effect, the confidence intervals produced by the CER-optimal bandwidth can provide additional evidence on the question *whether* there is a statistically significant treatment effect.

5 Data and sample selection

The data used in this thesis comes from the Canadian Library of Parliament. For the candidates that have run for office in elections for the Canadian House of Commons since 1867, data is available on their political affiliation, previous employment, the electoral results, and their gender.[6] In total, the dataset contains information on 44442 candidates, competing in 11908 elections. For 16 candidates, information on their gender was missing. This information has been manually added based on Google searches. In the next sections, I will describe how the analysis sample is selected.

5.1 Sample selection

Because universal suffrage was introduced in Canada in 1960 when Canada's indigenous peoples were granted an unconditional right to vote in federal elections, observations prior to the general election of the 25th parliament on 18 June 1962 have been dropped.

To conduct the RD analysis, data is needed for each electoral district on the election t and $t-1$. Since the borders of electoral districts in Canada are redrawn approximately once a decade, it is not possible to have a set of two-period observations for each election—*i.e.* for elections that took place just after electoral boundaries have been redrawn, there is (often) no information on the previous election in the same district, because the district might not have existed before. The elections for which no

[6] The data has been retrieved on 25 June 2021 from the Library of Parliament's online tool Parlinfo https://lop.parl.ca/sites/ParlInfo/default/en_CA/ElectionsRidings/Elections.

observation on the previous election is available due to redistricting have been dropped from the dataset.

Of the remaining data, not all observations are useful for the purpose of this thesis. Only the elections in which the two strongest candidates are a combination of a woman and a man will be informative for this study, since only for these mixed-gender elections, treatment status is discontinuous around the cut-off point. Selecting only the mixed-gender elections leaves 786 observations where 4280 candidates, of which 1456 are women, competed for a seat in Canada's House of Commons. In Table 1, the different steps of the sample selection are presented.

5.2 Descriptive statistics

In Table 2, some descriptive statistics are presented for variables available in the main dataset. The mean margin of victory in the main sample is 0.21, indicating that certainly not all elections were highly competitive. The forcing variable—*i.e.* the normalised female margin of victory—has an average of -0.04 and an standard deviation of 0.26. In Figure 2 a histogram of the forcing variable is shown. The relative abundance of the observations with a forcing variable close around the cut-off point is promising for this thesis, since the local non-parametric estimation techniques discussed in Section 4.2 have better performance when the number of observations close around the cut-off point is large.

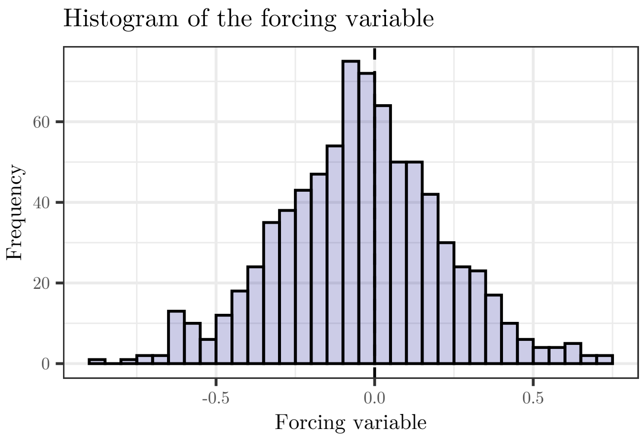

Figure 2: Histogram of the forcing variable $\tilde{v}_{it-1}$.

In the sub-sample of mixed-gender elections ($n = 786$), on average, there are 5.45 candidates for the House of Commons, of which 1.85 are women. The fact that on average, 0.77 of the candidates was an incumbent Member of Parliament shows that there is a large tendency to seek re-election.

The descriptive statistics are also provided for the full population of all elections since the introduction of universal suffrage ($n = 5688$). With the exemption of the variable *Margin of victory*, there seem to be substantial differences between the sub-sample and the population. In the mixed-gender elections, there are more candidates, more female candidates, and fewer incumbents on average. Furthermore, the mean number of total cast votes is larger for the mixed-gender elections, indicating that on average, these elections took place in larger districts, or that the mixed-gender elections correlate with relatively high voter turnouts.

In the last column of Table 2, the p-values of Welch's t-test are presented. This test investigates the equality of means. With the exception of the variable *Margin of victory*, the p-values show that the means of the sub-sample and the population are not equal. Due to these differences, the results of the RD design do not necessarily provide information on the treatment effect for the full population of elections, since the sub-sample used to estimate the RD treatment effect is not representative for the full population. Consequentially, the RD estimate of the treatment effect is not only local because it only provides an estimate for the effect in elections close around the cut-off point, but it is also local because it provides an estimate for a non-representative sub-sample of the full population.

In Table 3, an overview of the party affiliation of the winners and runner-ups of the mixed-gender sample is presented. The four major political parties in Canada (Liberal, Conservative/Progressive Conservative, Bloc Québécois, and New Democratic) are the largest suppliers of winner- and runner-up candidates in the close-gender elections, but the percentage of female candidates varies substantially per party.

5.3 Additional data sources

For the covariate balance tests and the complementary geographic analysis, additional data is used from Canada's Censuses. Every five years, Canada conducts a large-scale national census in which a wide array of data is gathered, including demographic, economic, and educational information. Since 1996, census data are available on the federal electoral district level. Because elections are held more frequently, census data is not available on the election level.

The main purpose of the covariate balance test, as discussed in section 3.3.2, is to investigate whether covariates determined *prior* to the assignment of treatment show a discontinuity around the cut-off point. Therefore, the absence of census data on the election level does not make the census data irrelevant for the RD design. To ensure that none of the covariates could themselves be affected by the treatment variable, data collected before the elections took place have been matched to each observation. For example, the elections that took place in 2004 and 2006 have been matched to data from the 2001 Census. Because no Census data is available on the federal electoral district level before 1996, the covariate-augmented sub-sample ($n = 599$) is smaller than the main sample used in the RD analysis ($n = 786$).

Table 1: Sample selection.

	Elections			Candidates		
Data sample	Total	Female MPs	Percentage	Total	Female	Percentage
Full sample	11908	803	7%	44442	5765	13%
Universal suffrage	5688	783	14%	29233	5 586	19%
Redistricting	3559	442	12%	17457	3119	18%
Mixed-gender	786	333	42%	4280	1456	34%

Note: In the final row, the data sample used to estimate the main regression models is described.

Table 2: Descriptive statistics.

	Mixed-gender elections ($n = 786$)				Universal suffrage ($n = 5688$)				
Variable	Mean	Min	Max	SD	Mean	Min	Max	SD	Welch t-test
Forcing variable	-0.04	-0.86	0.74	0.26					
Margin of victory	0.21	0.00	0.86	0.16	0.21	0.00	1.00	0.16	0.814
Total candidates	5.45	2.00	11.00	1.45	5.14	1.00	13.00	1.63	0.000
Female candidates	1.85	1.00	6.00	0.88	0.98	0.00	9.00	1.10	0.000
Incumbents	0.77	0.00	2.00	0.43	0.83	0.00	2.00	0.40	0.000
Total votes	45696	7721	75500	11934	41352	4804	183443	14255	0.001

Note: To test the equality of means, the Welch's t-test is used to accommodate for the unequal variances and unequal sample sizes of the sub-sample of mixed-gender elections and the full population of elections after the introduction of universal suffrage.

Where possible, Census data has been matched based on the name of the federal electoral district. Due to differences in the codification of the names of electoral districts and changing names of the electoral districts, not all data could be matched based on the district name automatically. In a number of cases, matching has been done manually. For all these manual matches, name changes have been checked using the list of 'Legal measures governing changes in federal electoral districts' published online by the Library of Parliament.[7]

Finally, data from the 2019 Canadian Election Study (Stephenson et al. 2020) have also been matched with the main data. The 2019 Canadian Election Study is a very large sample survey ($n = 37822$) conducted among Canadians during the 2019 general election. In this study, Canadians are asked about their political behaviour and attitudes. Although Canadian Election Studies have been conducted since 1965, the 2019 edition is for the first time designed to allow for constituency-level analysis. On average, there are 110 respondent for each of the 338 federal electoral districts. The smaller sample sizes make the previous editions unfortunately unsuitable for the purpose of this thesis. The data from the 2019 Canadian Election Study will be used to construct two alternative outcome variables in one of the extensions presented in section 8.

[7] https://lop.parl.ca/sites/ParlInfo/default/en_CA/legislation/legalMeasuresDistricts.

6 Main results

6.1 Global parametric approach

In its simplest form, the treatment effect τ can be estimated by running two separate linear regression models at both sides of the cut-off point, to then calculate the difference between the estimated intercepts for both models. This simple method provides an estimate of the LATE of 0.489, which suggests that compared to the districts where a woman had barely lost the previous election, on average, there are 0.489 more female candidates in the districts where a woman won the previous election by a narrow margin.

As discussed above, a more flexible way to obtain estimates of the treatment effect τ—and the corresponding standard errors—is to estimate the following pooled regression model:

$$Y_{it} = \alpha + \tau D_{it-1} + f(\tilde{v}_{it-1}) + D_{it-1} f(\tilde{v}_{it-1}) + \epsilon_{it}$$

where the functional form of $f(\tilde{v}_{it-1})$ is a polynomial of degree p:

$$f(\tilde{v}_{it-1}) = \sum_{n=0}^{p} \beta_n \tilde{v}_{it-1}^{n}$$

In Table 4, the estimated treatment effect $\hat{\tau}$ is shown for the pooled regression model with a degree of polynomial p ranging from 0 to 6. Regardless of the functional specification of $f(\tilde{v}_{it-1})$, the estimated treatment effect is positive and statistically significant at the 0.05 level—

Table 3: Political affiliation of the winners and runner-ups.

	Winners			Runner-ups		
Party	Total	Women	Percentage	Total	Women	Percentage
Liberal Party of Canada	315	159	50%	319	208	65%
Conservative Party of Canada	183	50	27%	158	58	37%
Bloc Québécois	90	41	46%	34	16	47%
Progressive Conservative Party	90	28	31%	85	46	54%
New Democratic Party	85	49	58%	160	110	69%
Reform Party of Canada	16	5	31%	15	7	47%
Social Credit Party of Canada	3	0	0%	7	2	29%
Independent	2	0	0%	1	0	0%
Green Party of Canada	1	1	100%	5	5	100%
Liberal Labour Party	1	0	0%	0	0	0
No affiliation to a recognised party	0	0	0	1	0	0%
Rhinoceros Party of Canada	0	0	0	1	1	100%
Totals	786	333	42%	786	453	58%

and in most cases the p-value < 0.01.[8] Interestingly, the estimated treatment effect with $p = 0$ is of a somewhat comparable size as most other estimates—suggesting that the relationship between the forcing variable and the outcome variable might be of a relatively modest magnitude.

Although $\hat{\tau}$ is positive and statistically significant for all functional specifications, it is important to notice that the size of the effect varies substantially for the different degrees of polynomials. Where the estimated treatment effect is 0.376 with $p = 2$, it is almost three times as large (0.998) with $p = 6$. Clearly, the point estimates of the treatment effect are sensitive to the degree of polynomial.

Table 4: RD estimates from global polynomial regression models.

p	$\hat{\tau}$	s.e.	95% CI	p-value	AIC
0	0.549	0.087	[0.379, 0.720]	0.000	2452
1	0.489	0.138	[0.219, 0.759]	0.000	2455
2	0.376	0.176	[0.031, 0.722]	0.033	2457
3	0.589	0.219	[0.160, 1.019]	0.007	2454
4	0.677	0.245	[0.196, 1.158]	0.006	2456
5	0.820	0.312	[0.208, 1.433]	0.009	2459
6	0.998	0.336	[0.339, 1.657]	0.003	2461

Note: The degree of the polynomial is indicated by p. All standard errors are heteroskedasticity consistent (HC3), clustered on the electoral district–representation order level.

According to the AIC presented in Table 4, the most appropriate functional specification would be a constant with $p = 0$ and the second 'best' specification is a cubic function with $p = 3$. However, as already mentioned, the purpose of RD designs is to obtain local estimates at the cut-off point, and model selection criteria which depend on global goodness-of-fit indicators, such as that AIC, are not necessarily meaningful in RD settings.

In principle, graphical inspection of the data could provide a better indication of the relationship between the forcing variable and the outcome variable. In Figure 3, four different plots of the RD design are shown. Two observations are noteworthy. First, the plots with quantile-spaced bins—bins all containing a similar number of observations—clearly show that a large majority of observations is located within 0.5 points from the cut-off point. Secondly, within this range, none of the plots suggest a clear non-linear relationship between the forcing variable and the outcome variable. Even though the plotted lines of a fourth degree polynomial regression model seem to provide an excellent global fit for the data in the evenly-spaced panels (A & B), this is far less obvious in the quantile-spaced panels (C & D).

The sensitivity of the estimated treatment effect to the functional specification in combination with a lack of clear evidence on the true functional relationship between the forcing variable and the outcome variable calls for a cautious interpretation of the global polynomial estimates. Nonetheless, all global polynomial regression model do suggest that an important first conclusion can be made with some degree of confidence: the election of a woman to the Canadian House of Commons seems to have a positive and statistically significant effect on the number of female candidates in the following election within an electoral district. However, the magnitude of this effect remains far from certain.

6.2 Local non-parametric approach

Given the sensitivity of $\hat{\tau}$ to the functional specification in the global parametric approach, it is even more important to complement this analysis with local non-parametric approaches. In comparison to the global approach, a local

[8] For comparison, in Table 18, the non-clustered standard errors have also been presented.

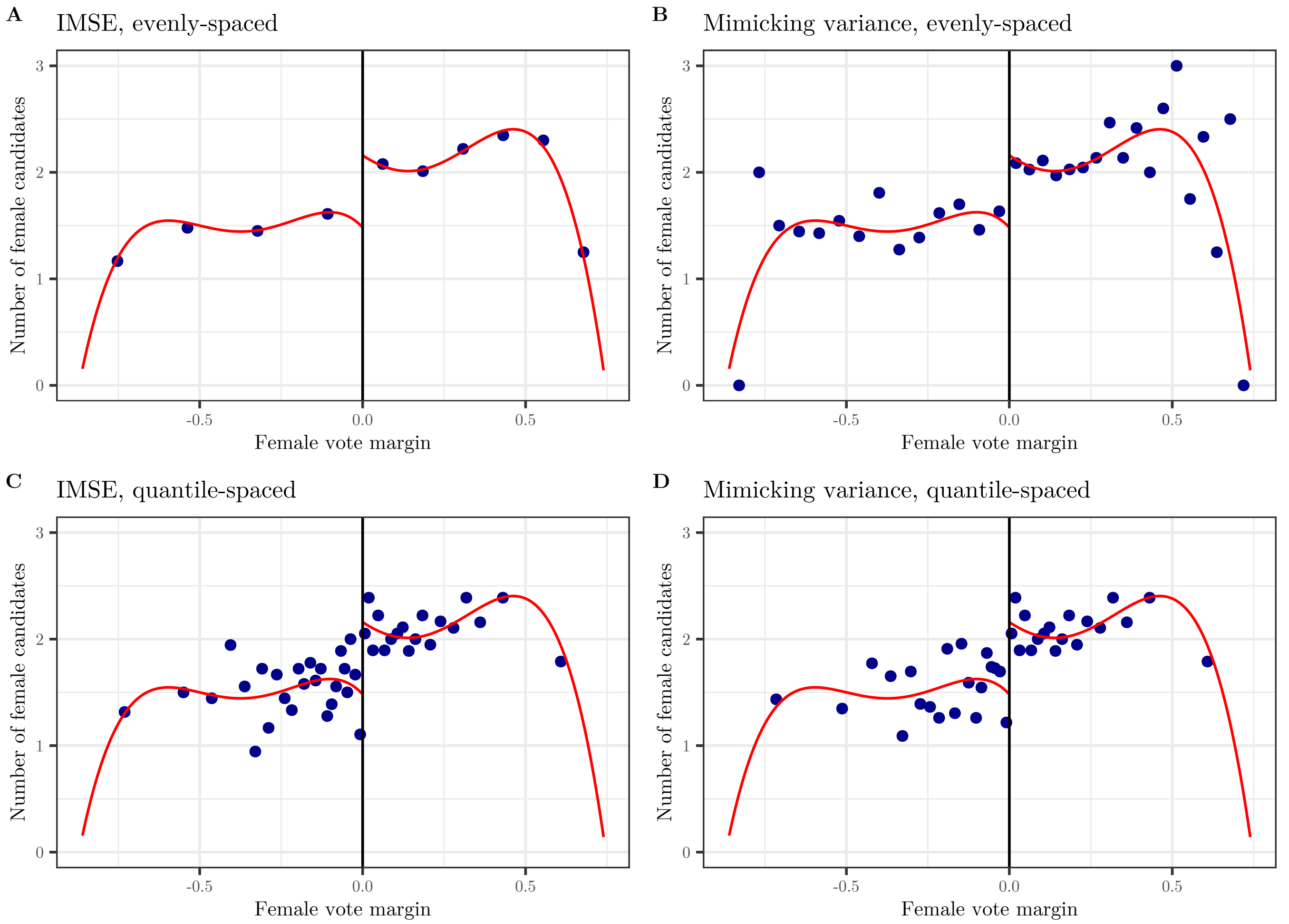

Figure 3: Graphical inspection of the RD design. The red lines are plots of polynomial regression models with $p = 4$. Bins are evenly-spaced (panel A & B) or quantile-spaced (panel C & D). The number of bins is calculated with Cattaneo, Idrobo, et al.'s (2019) IMSE-optimal (panel A & C) or the mimicking variance methods (panel B & D).

approach is less prone to unreliable and biased estimates of the boundary points. Since local regression model focus on approximating the regression function only near the cut-off point, the obvious drawback is a reduction of the sample size and an accompanying loss of precision.

In Figure 4, this bias-variance trade-off can be observed. In panel A and B of this figure, $\hat{\tau}$ is plotted for all possible bandwidths h, as estimated by a linear (A) and quadratic (B) regression model. For example, at $h = 0.1$, the regression functions are estimated using only observations where the normalised female vote margin $\tilde{v}_{it-1}$ in the range $[-0.1, 0.1]$. The shaded areas correspond to the the 95% confidence interval of $\hat{\tau}$. The regression models are estimated with uniform kernels. In the Appendix, the same figure is replicated using triangular kernels (Figure 17).

Interestingly, these local regression models suggest that the global regression models are affected by a downward bias. Although $\hat{\tau}$ estimated using very small bandwidths ($h < 0.05$) are sometimes smaller than the global estimates ($h = 1.00$), and with the quadratic model, $\hat{\tau}$ is even estimated to be negative for some very small bandwidths, it is important to note that these small-bandwidth estimates have a low precision. In panel C, it can be seen that with $h < 0.05$, the sample consists of fewer than 150 observations.

However, for both the local linear and quadratic regression models, there exists a range of bandwidths were $\hat{\tau}$ is estimated to be substantially larger than the global estimate with a relatively large degree of precision. For example, at $h = 0.06$, the treatment effect is estimated to be $\hat{\tau} = 1.075$ with a 95% confidence interval of $[0.486, 1.663]$ in the local linear regression model, whereas the treatment effect was estimated to $\hat{\tau} = 0.489$ using a global linear regression model.

Although Figure 4 thus suggests that the global estimates suffer from a (slight) downward bias, it remains unclear which estimate is the most credible. It could be the local linear estimate of $\hat{\tau} = 1.075$ at $h = 0.06$, but it could also be the local linear estimate at $h = 0.12$, in which case the treatment effect would be $\hat{\tau} = 0.532$.

6.2.1 Bandwidth selection

As discussed in Section 4.2.1, an important aspect of the local non-parametric approach to a RD design is deter-

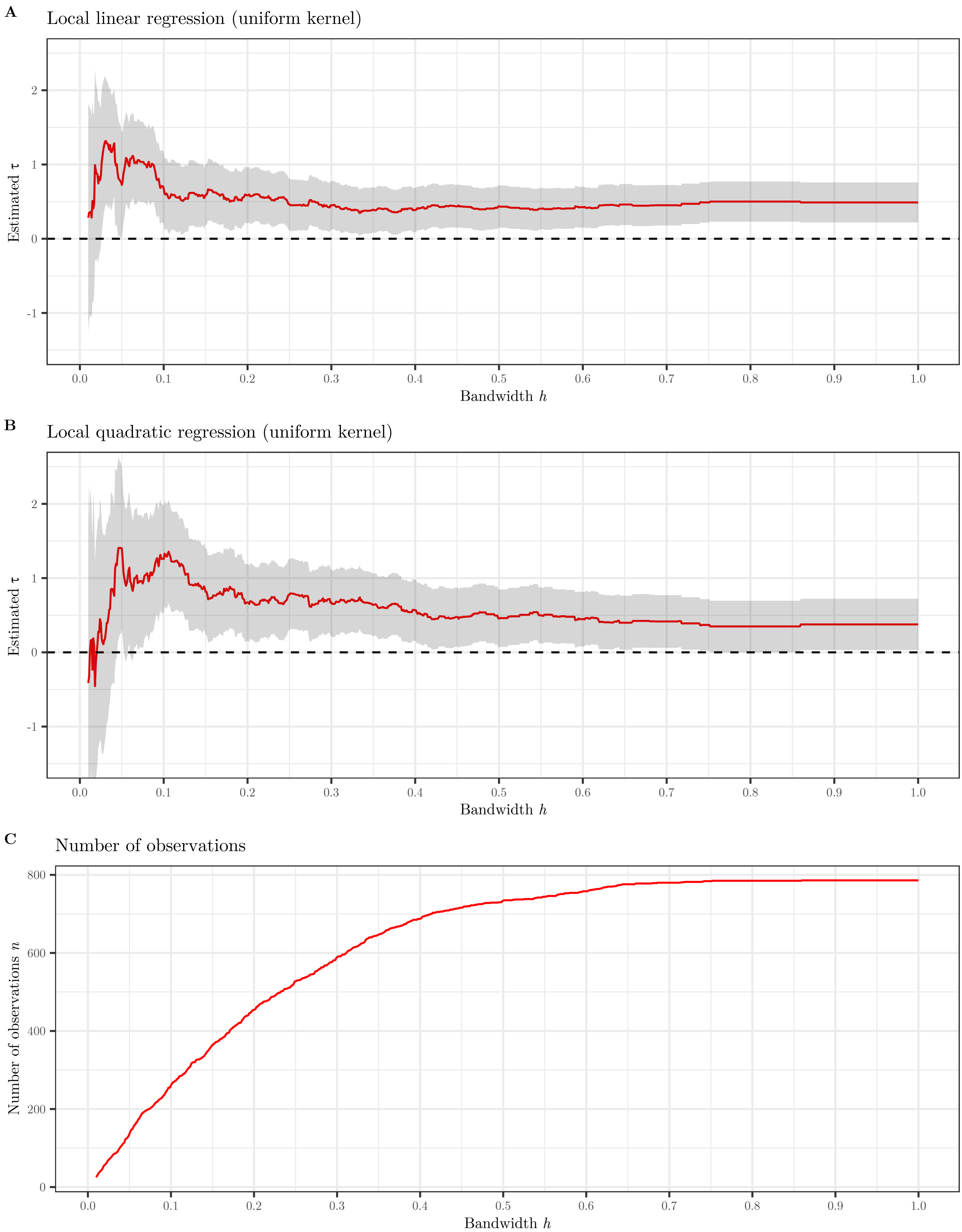

Figure 4: Graphical representation of the bias-variance trade-off of local linear regression techniques. Panel A and B show the estimated treatment effect $\hat{\tau}$ of local linear and quadratic regression models with bandwidths ranging from 0.01 to 1.00. The shaded areas represent the 95% confidence interval, calculated using heteroskedasticity consistent (HC3) standards errors clustered on the electoral district–representation order level. Panel C shows the number of observations.

mining the bandwidth h used to estimate the treatment effect. A first inspection of Panel C of Figure 4 suggest to choose a bandwidth smaller than 0.5, since the amount of observations with $\tilde{v}_i > 0.5$ is limited. Intuitively, any substantial reduction in variance and the accompanying increase in precision from using a bandwidth larger than 0.5 therefore unlikely.

In Table 5, a number of 'optimal' bandwidths as calculated by the procedures discussed in Section 4.2.1 are shown.[9] The corresponding estimated treatment effect $\hat{\tau}$ and the 95% confidence intervals based on local linear regression models are shown as estimated with both a uniform kernel as well as a triangular kernel. In the final row of the table, the global parametric estimate is also included for comparison.

Table 5 shows a number of interesting results. First, both the IK bandwidth (0.414) and the LM cross-validation bandwidth (0.835) are surprisingly large—especially considering that 700 of the 786 observations (89%) fall within the IK bandwidth of 0.414. Consequentially, the corresponding estimated $\hat{\tau}$ and the 95% confidence intervals are highly comparable to the global estimate.

The MSE-optimal bandwidth (and the CER-optimal bandwidth) are substantially smaller than the LM and IK bandwidths. Because local regression models are based on the idea that functions can be approximated increasingly well by linear (or low-order polynomial) function as the bandwidth decreases, these smaller bandwidth will be used as the preferred bandwidths from now on. The choice for these smaller bandwidth as the preferred bandwidths is also be backed by the fact that the reduction in precision does not affect inference: the treatment effect remains statistically significant.

Somewhat surprising is the similarity between the global parametric estimate of $\hat{\tau}$ and the local non-parametric estimate $\hat{\tau}$ with the MSE-optimal bandwidth of 0.118 using a uniform kernel. However, once the treatment effect is estimated with a triangular kernel—which is optimal for the estimation of boundary points in combination with the MSE bandwidth—the difference between the global estimate and the MSE-optimal point estimate grows substantially. This provides additional evidence for the suggestion that the treatment effect estimated with a global approach suffer from a (slight) downward bias—or at least when low-order polynomials are used.

Note that the estimated $\hat{\tau}$ based on the CER-optimal bandwidth has little value in itself, since this bandwidth is constructed specifically for inference and *not* for point estimates. However, the fact that the confidence intervals are of similar length for the estimates based on the MSE- and CER-optimal bandwidths does suggest that inference based on MSE-optimal point estimates and conventional standard errors is unlikely to be substantially biased.

[9] See also Appendix A.

6.2.2 Inference

As discussed in Section 4.3, more sophisticated inference procedures that take into account the fact that local regression models are assumed to only approximate the underlying regression function are readily available. In Table 6, the results of Calonico, Cattaneo, and Titiunik's (2014) robust bias-corrected inference methods based on a local linear regression model using the MSE-optimal bandwidth ($h = 0.146$) and a triangular kernel are shown. The most important estimates are printed bold.

The first thing to notice is that the conventional point estimate $\hat{\tau} = 0.697$ is the same as the estimate presented in Table 5. As already mentioned, this estimate is MSE-optimal, and thus remains the preferred estimate for the magnitude of the treatment effect. In the second row, a bias-correction of the treatment effect is applied: $\hat{\tau} - \hat{\mathscr{B}} = 0.764$, and the confidence interval is centered around this bias-corrected estimate. The final row incorporates the robust standard error, which is slightly larger than the conventional standard error. As can be seen from the p-value $= 0.003$ and the 95% confidence interval, the robust bias-corrected inference approach does not invalidate the conclusion that in an electoral district, the election of a female politician leads to an increase in the average number of female candidates in the next election by approximately 0.697.

The plot of this local linear RD model with the preferred bandwidth $h = 0.146$ is shown in Figure 5. The local linear regression model is estimated with a triangular kernel. The blue dots are binned local averages constructed with Cattaneo, Idrobo, et al.'s (2019) quantile-spaced mimicking variance method—i.e. each bin contains a similar amount of observations and the variability shown by the bins approximates the variability of the raw data. The discontinuity in the linear functions at the cut-off point represents the estimated treatment effect $\hat{\tau} = 0.697$.

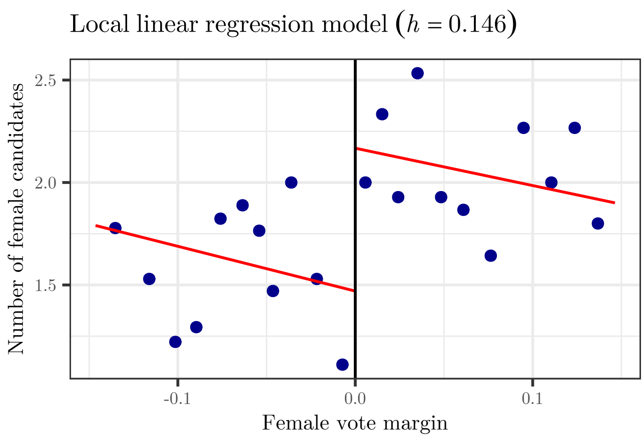

Figure 5: Graphical representation of the local linear regression model with $h = 0.146$. The bins are quantile-spaced, and the number of bins is calculated with Cattaneo, Idrobo, et al.'s (2019) Mimicking Variance method, which approximates the variability in the raw data. The graph is rendered using the R command `rdplot`.

Table 5: Optimal bandwidths for local linear regression models.

	Uniform kernel					Triangular kernel				
Procedure	h^{opt}	$\hat{\tau}$	s.e.	p-value	95% CI	h^{opt}	$\hat{\tau}$	s.e.	p-value	95% CI
LM	0.835	0.501	0.138	0.000	[0.230, 0.771]	0.835	0.452	0.137	0.001	[0.184, 0.719]
IK	0.414	0.409	0.150	0.007	[0.115, 0.703]	0.414	0.454	0.157	0.004	[0.146, 0.761]
MSE	0.118	0.559	0.236	0.018	[0.097, 1.022]	0.146	0.697	0.222	0.002	[0.263, 1.132]
CER	0.085	1.012	0.261	0.000	[0.500, 1.523]	0.106	0.885	0.249	0.000	[0.397, 1.374]
Global	1.000	0.489	0.138	0.000	[0.219, 0.759]	1.000	0.461	0.135	0.001	[0.196, 0.727]

Note: Optimal bandwidths are determined by the different procedures discussed in Section 4.2.1. The calculation of the LM cross-validation- and the IK AMSE bandwidths is discussed in Appendix A. For the CCT MSE- and CER-optimal bandwidths, which are determined using the `R` package `rdrobust`, different optimal bandwidths are available for the uniform and triangular kernel (Calonico, Cattaneo, Farrell, and Titiunik 2017). The treatment effects $\hat{\tau}$ are estimated using (local) linear regression models with uniform and triangular kernels. All standard errors are heteroskedasticity consistent (HC3). Standard errors are clustered on the electoral district–representation order level.

Table 6: Local linear regression estimates with robust bias-corrected inference methods

Method	$\hat{\tau}$	s.e.	p-value	95% CI
Conventional	**0.697**	0.225	0.002	[0.256, 1.139]
Bias-Corrected	0.764	0.225	0.001	[0.322 , 1.205]
Robust	0.764	**0.259**	**0.003**	**[0.256 , 1.271]**

Note: Results of the robust bias-corrected methods introduced by (Calonico, Cattaneo, and Titiunik 2014) are calculated using the `R` package `rdrobust`. The standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

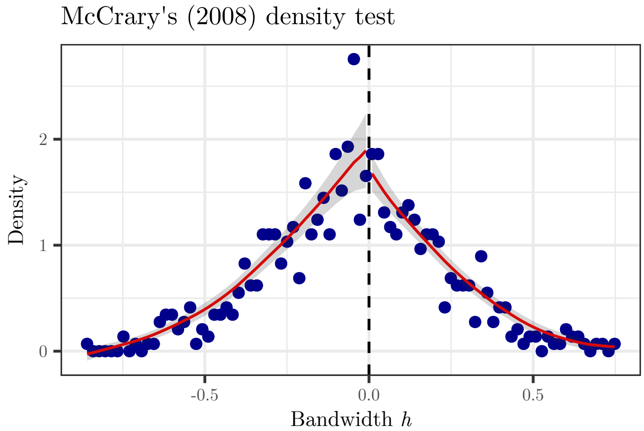

Figure 6: Graphical representation of the McCrary (2008) density test.

7 Validity and robustness checks

7.1 Forcing variable density test

The first validity tests is the forcing variable density test. In Figure 6, the McCrary (2008) density test is graphically presented, as constructed using the `R` package `rdd`. The blue points are local density estimates of the forcing variable $\tilde{v}_{it-1}$ and the red line is the local linear smoothing of this density, estimated separately at each side of the cut-off point using a triangular kernel. The shaded areas represent the 95% confidence interval of the smoothed density estimate. The fact that the smoothed density estimates are not far apart at each side of the cut-off point and that the confidence intervals are for a large part overlapping already suggests that there is no discontinuity in the density of the forcing variable.

A more formal investigation of the McCrary (2008) density test confirms this conclusion. The log difference in heights of the smoothed density estimate is -0.115 with a standard error of 0.139. The corresponding p-value of 0.408 is not small enough to reject the null hypothesis of no discontinuity in the density of the forcing variable.

In Figure 7, the Cattaneo, Jansson, et al. (2020) density test is graphically presented, as constructed using the `R` package `rddensity`. The blue histogram represents the density of the forcing variable and the red lines and corresponding shaded areas represent a local quadratic density estimator, as estimated with a triangular kernel and robust bias-corrected confidence intervals (Calonico, Cattaneo, and Titiunik 2014). The graphical representation of this density test, with its overlapping confidence intervals, provides a similar overall idea of the density of the running variable as Figure 6—although the difference in the estimated density just to the left and the right of the cut-off point is somewhat larger.

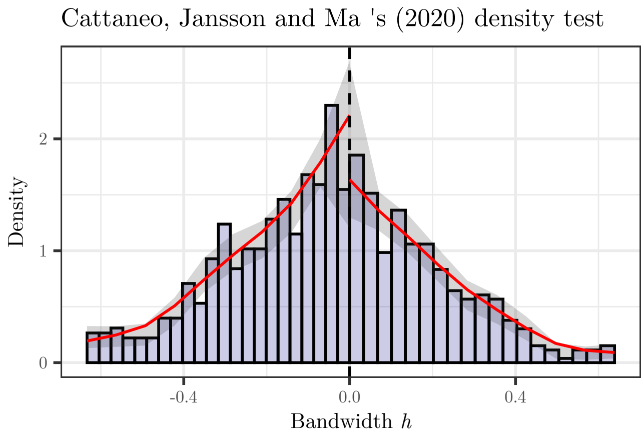

Figure 7: Graphical representation of the Cattaneo, Jansson, et al. (2020) density test.

As with the McCrary (2008) test, the Cattaneo, Jansson, et al. (2020) density test does not provide any evidence to reject the null hypothesis of a continuous density of the forcing variable. Neither the p-values of the test without ($p = 0.095$) and with robust bias-correction ($p = 0.948$) provide sufficient evidence to reject the null hypothesis of a continuous density of the forcing variable. As expected in an electoral setting, there seems to be no precise manipulation of the voting shares obtained by the candidates for political office. Consequentially, there is a (refutable) expectation that the potential outcomes Y_{1it} and Y_{0it} are continuous around the cut-off point.

7.2 Covariate balance test

As discussed in Section 3.3.2, a second widely used test to determine the validity of a RD design is the covariate balance test. Implementing this test is relatively straightforward: by applying the same statistical techniques used to determine the existence (and magnitude) of a discontinuity of the outcome variable of interest Y_i to a number of covariates which should not be affected by the treatment, it can be tested whether the 'as good as random' distribution of the treatment variable holds around the cut-off point.

The main dataset already contains a number of variables can be used as covariates in this balance test. The variables used are (A) the total number of votes cast, (B) the total number of candidates, (C) the number of female candidates, and (D) the number of incumbent candidates running for re-election. To ensure that these variables are not affected by the treatment, the variables from the election $t-1$ are used. From the additional Census data discussed in Section 5.3, a number of variables that are not directly related to the elections will also be used to conduct the covariate balance test. Because Census data is not available for the elections before 1996, the sub-sample for the Census-based covariate balance test ($n = 599$) is smaller than the main sample ($n = 786$).

Since all the covariates used in the balance test are determined prior to the determination of the treatment status, statistically significant discontinuities in the covariates at the cut-off point would suggest that on average, the electoral districts just to the left and right of the cut-off point were *not* similar when the treatment status was determined. The consequences of such discontinuities is that a violation of the identifying assumption of the RD design becomes more probable, since the discontinuities in pre-treatment characteristics makes it less obvious that the potential outcomes would still be continuous around the cut-off point.

In Figure 8, the relationships between these covariates and the running variable are graphically presented in a similar fashion as Figure 3. A first observation from these figures is that there seesm to be no clear discontinuity in any of the covariates at the cut-off point. Although the 4th degree global polynomials in Panels E, K, or L do seem to suggest a discontinuity around the cut-off point, it is important to remember that such high-order polynomials can provide poor estimates at the boundary of a function. Since the binned local averages in these Panels do not showcase a clear discontinuity, these figures do not provide substantial evidence to believe that the covariates are discontinuous around the cut-off point. Fortunately, formal tests can bring more clarity.

In Table 7, the treatment effect as estimated with a local linear regression model is shown for each of the covariates. All models are estimated using CCT MSE-optimal bandwidths and triangular kernels. Both the conventional- and robust bias-corrected standard errors and p-values are shown. In the Appendix, the same table is replicated with uniform kernels (Table 19).

Surprisingly, for the covariate *Female Candidates*, the null hypothesis of no discontinuity must be rejected with a significance level $\alpha = 0.05$ when the robust bias-corrected inference method is used. The local linear regression models thus seem to provide some evidence that during election $t-1$, the average number of female candidates in electoral district just to the left of the cut-off point is lower in comparison to the district just to the right of the cut-off point.

However, two observations cast some doubt whether this statistical significance can truly be interpreted as evidence for a 'treatment effect' on the covariate *Female Candidates*. First, as shown in Figure 18 in the Appendix, choosing almost any bandwidth $h \neq 0.095$ results in the disappearance of the statistical significance. Secondly, when the CER-optimal bandwidth is used, the estimated treatment effect on the covariate *Female Candidates* also loses its significance with a conventional $p = 0.061$ and a robust bias-corrected $p = 0.053$.

Nonetheless, even if one of the twelve covariates used in the balance test does truly show a statistically significant discontinuity at the 5% significance level, this cannot be interpreted as definitive evidence for the invalidity of the RD design. A simple calculation shows that, with a significance level $p = 0.05$ and $n = 12$ executed statistical tests, the chance of exactly $k = 1$ false positive is: $P(X = 1) = \binom{12}{1} \cdot 0.05^1 \cdot 0.95^{11} = 0.341$. In other words, given the number of covariates used in the balance test, it is not very surprising that in one of those test, the null hypothesis of no discontinuity must be rejected (Eggers, Fowler, et al. 2015, p. 271).

7.3 Sensitivity to covariates

With these covariates, an additional sensitivity check can be performed. Since causal inference in a RD design is not based on control variables or matching techniques, the addition of covariates to the local linear regression models should have no substantial impact on the estimate of the treatment effect—although it could lead to increased

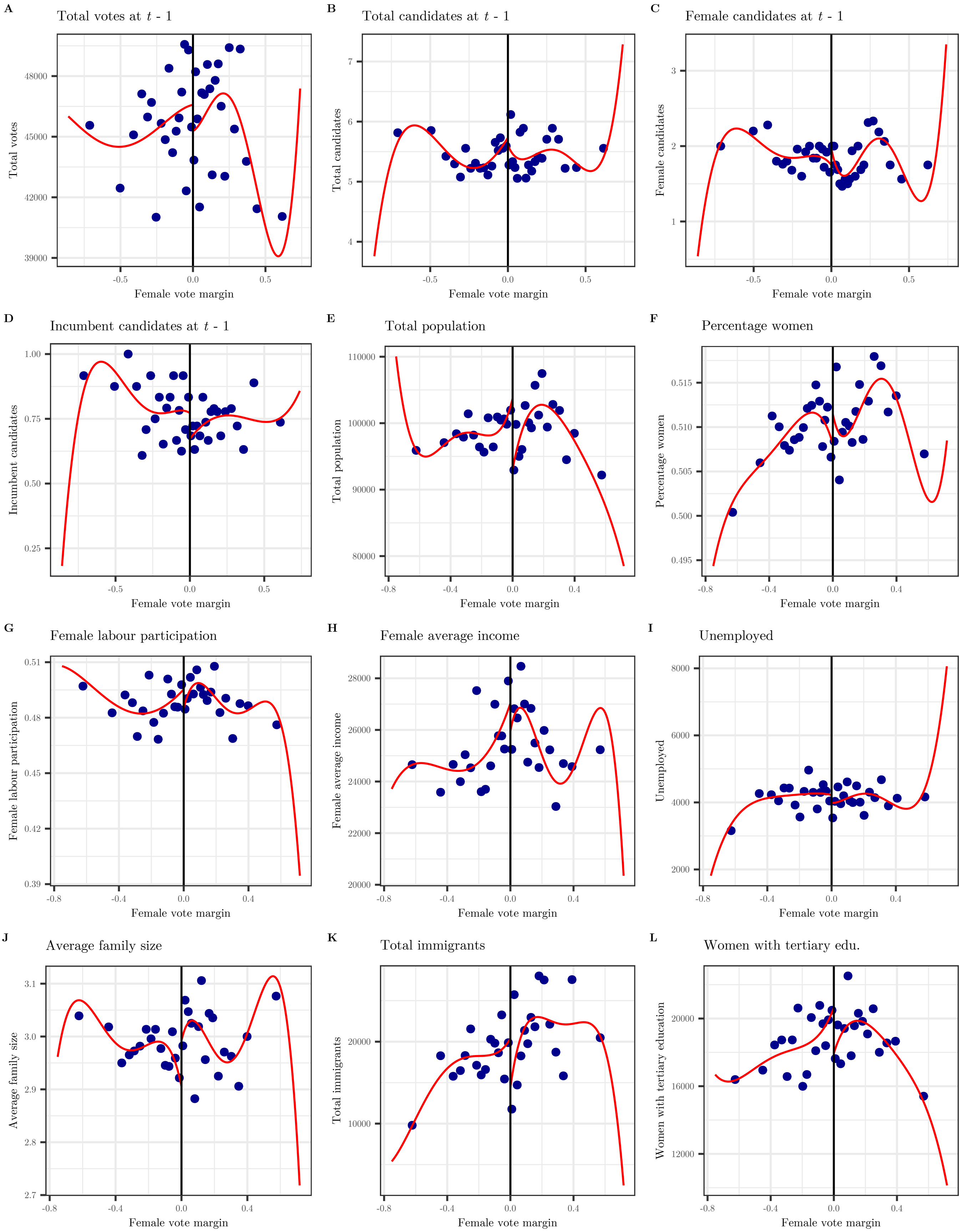

Figure 8: Graphical representation of covariate balance tests. Graphs are produced using the `R` package `rdplot`. The red lines are plots of polynomial regression models with $p = 4$. Bins are quantile-spaced, and the number of bins is calculated with Cattaneo, Idrobo, et al.'s (2019) IMSE-optimal methods—which is best suited to 'trace out' the underlying relationship between the forcing variable and the covariate.

Table 7: Covariate balance tests.

Covariate	h^{opt}	$\hat{\tau}$	Conventional s.e.	Conventional p-value	Robust s.e.	Robust p-value
Total votes	0.235	-1396.827	2332.974	0.549	2736.264	0.561
Total candidates	0.162	-0.134	0.314	0.669	0.374	0.784
Female candidates	0.095	0.470	0.243	0.053	0.274	0.038
Incumbent candidates	0.170	-0.109	0.095	0.252	0.114	0.300
Total population	0.179	-7587.341	4398.189	0.085	5145.727	0.097
Percentage women	0.143	0.003	0.003	0.188	0.003	0.111
Female labour participation	0.154	-0.005	0.013	0.718	0.016	0.640
Female average income	0.164	-701.274	1829.213	0.701	2156.658	0.566
Unemployed	0.165	-306.450	368.629	0.406	443.452	0.471
Average family size	0.156	0.090	0.055	0.099	0.066	0.148
Total immigrants	0.172	-2495.230	4307.129	0.562	5197.876	0.618
Women with tertiary education	0.168	-2133.928	1546.002	0.167	1820.053	0.171

Note: For each covariate, the MSE-optimal bandwidth h^{opt} is determined. Using this bandwidth, the estimated treatment effect $\hat{\tau}$ and the corresponding (robust bias-corrected) standard errors and p-values are calculated using the `R` package `rdrobust`. The standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

precision (e.g. Lee 2008). Following Calonico, Cattaneo, Farrell, and Titiunik (2019), the covariates are added to the local linear regression model as additive-separable with a linear functional specification (without assuming the functional specification to be actually linear). This gives the following covariate-adjusted local linear regression model:

$$\hat{Y_i t} = \alpha + \tau D_{it-1} + \beta_l \tilde{v}_{it-1} + \beta_r D_{it-1}\tilde{v}_{it-1} + \gamma \boldsymbol{Z_i}$$

where $\boldsymbol{Z_i}$ is a vector of added covariates.

The only additional assumption necessary to ensure that this covariate-adjusted local linear regression provides a consistent estimate of the treatment effect, is that the treatment has no effect itself on the covariates (Calonico, Cattaneo, Farrell, and Titiunik 2019, p. 445-446). With the exception of the variable *Female candidates*, the results of Table 7 showed that this condition is likely to be satisfied. In the absence of discontinuities in the covariates around the cut-off point, the estimated treatment effect should therefore in principle not be (highly) sensitive to the addition of these pre-determined covariates.

Since the covariates are not available for all observations used in this RD analysis, the treatment effect estimated without the addition of covariates using the subset of observation for which the covariates are available is first presented in Table 8 as model (1). This MSE-optimal estimated treatment effect $\hat{\tau} = 0.735$ is slightly larger that the effect estimated using the full mixed-gender sample ($\hat{\tau} = 0.697$). This could indicate that the magnitude of the treatment effect has grown over the years, since only the earliest observations (prior to 1996) have been dropped. In the second row of Table 8, the p-value of the estimate is shown, as calculated by Calonico, Cattaneo, and Titiunik's (2014) robust bias-corrected inference methods. Unsurprisingly, the p-value is larger than the p-value for the treatment effect as estimated with the full mixed-gender sample—since fewer observations reduce the precision.

Far more surprising is the sensitivity of the estimated treatment effect to the addition of covariates to the local linear regression model. In model (2), the covariate *Total votes* is added to the regression model—denoted by the X—which causes the estimated treatment effect to increase in magnitude to 0.886. With the addition of the variable *Total candidates* in model (3), the treatment effect is estimated to be even larger, with $\hat{\tau} = 0.961$. The further addition of the other covariates leaves the estimate somewhere in the range $[0.886, 0.983]$. As can be seen in the last three rows, this changing estimate seems not to be (solely) attributable to the only slightly changing MSE-optimal bandwidth h, which is calculated separately for each covariate-adjusted regression model.

This sensitivity to the addition of the pre-determined covariates is remarkable, especially given the results of Table 7, which suggested almost all covariates to be balanced around the cut-off. So, even though the electoral districts just to the left and just to the right of the cut-off point were found not to systematically differ in terms of the total votes, or the total amount of candidates, the addition of these covariates increases the size of the estimated treatment effect by almost one third.

Calonico, Cattaneo, Farrell, and Titiunik (2019) have formally shown that the addition of covariates as linear variables in an additive-separable way results in consistent estimates for the treatment effect $\hat{\tau}$ as long as:

$$\lim_{\tilde{v}_{it-1}\downarrow 0} \mathbb{E}[Z_i|\tilde{v}_{it-1}] = \lim_{\tilde{v}_{it-1}\uparrow 0} \mathbb{E}[Z_i|\tilde{v}_{it-1}]$$

Given the earlier conclusion from the covariate balance test that this condition holds for all covariates—with the

Table 8: Covariate-adjusted local linear regression models

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Treatment effect $\hat{\tau}$	0.735	0.886	0.961	0.897	0.904	0.907	0.949	0.948	0.972	0.979	0.983	0.975	0.976	1.051
p-value	0.010	0.002	0.001	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000
Total votes		X	X	X	X	X	X	X	X	X	X	X	X	X
Total candidates			X	X	X	X	X	X	X	X	X	X	X	X
Female candidates				X	X	X	X	X	X	X	X	X	X	
Incumbent candidates					X	X	X	X	X	X	X	X	X	X
Total population						X	X	X	X	X	X	X	X	X
Percentage women							X	X	X	X	X	X	X	X
Female labour part.								X	X	X	X	X	X	X
Female ave. income									X	X	X	X	X	X
Unemployed										X	X	X	X	X
Average family size											X	X	X	X
Immigrants												X	X	X
Tertiary edu. women													X	X
Total observations	599	599	599	599	599	599	599	599	599	599	599	599	599	599
MSE-optimal h	0.148	0.141	0.141	0.141	0.141	0.140	0.140	0.140	0.135	0.137	0.137	0.135	0.135	0.135
N left	139	132	131	132	132	131	131	131	128	128	128	128	127	127
N right	127	122	122	122	122	121	121	121	118	119	119	118	118	118

Note: In the first two rows, the treatment effect and the corresponding p-value are presented as estimated using the robust bias-corrected methods introduced by (Calonico, Cattaneo, and Titiunik 2014). In the following rows, the *X* indicates whether to covariate has been added to the local linear regression models. In the last four rows, the total number of observations, the MSE-optimal bandwidth and the used number of observations to the left and the right of the cut-off point are shown.

exception of the variable *Female candidates*—there seems to be little evidence to substantiate a claim that the estimates presented in Table 8 are *not* consistent estimates for the RD treatment effect. Furthermore, when the covariate *Female candidates* is removed from the regression in model (14), the estimated treatment effect even further increases in magnitude to $\hat{\tau} = 1.051$.

Without taking any decisive conclusions on the true size of the local average treatment effect, most evidence thus suggests the estimate obtained with a simple local linear regression model without any covariates is more likely to be an underestimation than an overestimation of the true treatment effect.

7.4 Placebo discontinuities test

A final test for the validity of the RD design is the placebo discontinuities test, in which the existence of a statistically significant 'treatment effects' is investigated at values of the forcing variable $\tilde{v}_{it-1}$ where there is no discontinuity in treatment. The null hypothesis of this test is that the observed outcome variable Y_i is continuous at $\tilde{v}_i \neq 0$. A rejection of this null hypothesis would call the identifying assumption of the RD design—the continuity of the *potential* outcomes Y_{1it} and Y_{0it} at $\tilde{v}_{it-1} = 0$—into question.

This test can be conducted at both sides of the true cut-off point, but in order to ensure that the tests are not affected by the actual change in treatment status occurring at the true cut-off point, only observations to the left or the right of the cut-off points should be used. For example, with the placebo cut-off point $\tilde{v} = -0.1$, only observations with $\tilde{v}_{it-1} < 0$ are used to estimate the placebo treatment effect. As with the covariate balance test, the same techniques used to estimate the local linear regression models with the true cut-off point can be used to perform the placebo discontinuities test.

In Table 9, the results of local linear regressions models estimated for a number of cut-off points ranging from $\tilde{v}_i = -0.25$ to $\tilde{v}_i = 0.25$ are presented. For each of the placebo—and the actual—cut-off points, the MSE-optimal bandwidth is determined, based on which the 'treatment effect' $\hat{\tau}$ is estimated using a triangular kernel. Standard errors and the corresponding p-values are shown for both conventional inference methods as well as Calonico, Cattaneo, and Titiunik's (2014) robust bias-corrected methods.

The results of the placebo discontinuities test are comparable to the results of the covariates balance test: there is one case, at the placebo cut-off point $\tilde{v} = 0.05$, for which the null hypothesis of the continuity of the outcome variable must be rejected with a significance level $\alpha = 0.05$. The statistical significance remains present when the Calonico, Cattaneo, and Farrell (2018) CER-optimal bandwidth is used ($p = 0.040$).

As with the covariates balance test, a single rejection of the null hypothesis does not necessarily invalidate the RD design, because such rejection could very well have happened by chance when multiple statistical tests are

Table 9: Placebo discontinuities tests.

			Conventional		Robust			
cut-off	h^{opt}	$\hat{\tau}$	s.e.	p-value	s.e.	p-value	N left	N right
-0.25	0.089	-0.229	0.460	0.619	0.542	0.707	71	82
-0.20	0.079	0.422	0.390	0.279	0.456	0.370	60	81
-0.15	0.050	-0.484	0.789	0.539	0.934	0.518	48	56
-0.10	0.052	0.097	0.407	0.812	0.473	0.764	55	80
-0.05	0.058	-0.222	0.306	0.468	0.374	0.210	85	72
0.00	0.146	0.697	0.225	0.002	0.259	0.003	192	161
0.05	0.067	-1.336	0.626	0.033	0.894	0.034	64	66
0.10	0.043	0.190	0.596	0.750	0.727	0.831	38	45
0.15	0.059	0.349	0.406	0.390	0.475	0.396	58	50
0.20	0.082	-0.154	0.428	0.719	0.493	0.775	75	46
0.25	0.056	-0.801	0.477	0.093	0.545	0.072	34	27

Note: The tests for discontinuities at the placebo cut-off point are conducted with local linear regression models and Calonico, Cattaneo, and Titiunik's (2014) robust bias-corrected inference methods, implemented with the `R` package `rdrobust`. The MSE-optimal bandwidth is represented by h^{opt}, and the corresponding MSE-optimal point estimates is $\hat{\tau}$. Both the conventional and the robust bias-corrected standard errors are presented. The standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level. The final two columns show the number of observations used in the estimation of the local linear regression models.

performed. Furthermore, Figure 19 in the Appendix does indicate that the statistical significance of this estimate disappears for certain choices of bandwidth h.

8 Extensions

The results of section 6 showed that female success in competitive mixed-gender elections for the Canadian House of Commons raises the number of female candidates in the following election. This effect is estimated to be statistically significant at a 5% confidence level for all functional specifications with the global parametric approach and almost all bandwidths with the local non-parametric approach.

The estimated magnitude of the treatment effect did show some sensitivity to (i) the functional specification, (ii) the chosen bandwidth in the local regression models, and (iii) the inclusion of covariates. In the preferred local linear regression model, with an MSE-optimal bandwidth $h = 0.146$ using a triangular kernel, the treatment effect was estimated to be 0.697. Most available evidence suggest that if any bias is present, it is most likely (slight) downward bias. Both the inclusion of covariates, as well as the reduction of the bandwidth size increased the magnitude of the estimated treatment effect.

In section 7, evidence was presented for the validity of the RD design. First, no evidence for manipulation of the forcing variable could be found, as the density of the forcing variable was found to be continuous around the cut-off point. The results of the covariate balance test and the placebo discontinuities test were slightly ambiguous, since for both validity tests, the null hypothesis had to be rejected in one case. However, as more than 10 different models were estimated for both tests, it is far from certain that a single rejection of the null hypothesis should invalidate the RD design. Overall, the results of the validity tests thus point in favour of the fulfilment of the identifying assumption: the potential outcomes seem to be continuous around the cut-off point.

The conclusion that the election of a woman in a competitive mixed-gender election increases the number of female candidates during the following election does however not necessarily mean that a role-model effect is present. After all, if the increased number of women on the ballot can be fully explained by the tendency among incumbent members of parliament to seek re-election, it would be incorrect to speak of a role-model effect. To investigate more in depth whether the positive treatment effect can be (partly) attributed to an increased inflow of non-incumbent female candidates, four alternative hypothesis will be explored in this section.

8.1 Female challengers

First, the dependent variable of the main analysis as performed in section 6 can be swapped for an alternative outcome variable: the number of female *challengers* during the next election—*i.e.* non-incumbent female candidates. In the dataset used to estimate the treatment effect, the average number of candidates for each election was 5.45, of which 4.67 were not incumbent members of parliament running for re-election. If a discontinuity in the number of female challengers can be observed, this would indicate that the election of a woman to the Canadian House of Commons causes an increase in the inflow of other female candidates running for federal election.

Note however that the opposite is not necessarily true: the absence of a treatment effect does not provide conclusive evidence against the existence of a role-model effect.

It is important to realise that given the tendency among incumbent members of parliament to run for re-election, the number of female challengers within an electoral district will be predominantly determined by the number of female candidates from non-incumbent political parties. Therefore, estimates based on this alternative model with the number of female challengers as the outcome variable will most likely provide evidence on the existence and magnitude of a *cross-party* role-model effect or contagion effect. To determine whether there are *within-party* role-model effects, the current analysis is less suitable.

Based on the data available, an indicator for the number of female challengers can be easily constructed. On average, there are 1.41 female challengers in the elections following a mixed-gender election. For the alternative outcome variable *Female challengers*, the results of global regression models with a polynomial of degrees ranging from 0 to 6 are presented in Table 10. Interestingly, the treatment effect is estimated to be negative and statistically significant in the models with polynomials of degrees less than $p < 3$. This negative effect could for example resonate Reingold and Harrell's (2010) finding that in the US, the election of female politicians increase male political engagement within the opposing party. Alternatively, these negative estimates could indicate that women (and/or political parties) are less inclined to engage in intra-gender competition—because it could potentially split the gendered votes (Bhalotra et al. 2018, p. 1865).

Before theorising on the possible causal mechanisms behind these negative estimates, it is important to stress that these global estimates must be interpreted with caution since the true functional relationship between the outcome variable and the forcing variables is unknown—although graphical inspection of Figure 9 does suggest the functional relationship between the forcing variable and the alternative outcome variable to be close to linear.

Table 10: Global polynomial estimates for the outcome variable: *Female challengers*.

p	$\hat{\tau}$	s.e.	95% CI	p-value	AIC
0	-0.316	0.084	[-0.481, -0.150]	0.000	2431
1	-0.399	0.132	[-0.657, -0.141]	0.002	2434
2	-0.479	0.168	[-0.809, -0.150]	0.004	2437
3	-0.334	0.209	[-0.743, 0.076]	0.111	2437
4	-0.205	0.240	[-0.675, 0.265]	0.393	2439
5	-0.123	0.275	[-0.662, 0.415]	0.654	2442
6	0.077	0.308	[-0.527, 0.680]	0.803	2444

Note: The degree of the polynomial is indicated by p. All standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

Furthermore, as can be seen in the graphs presented in Figure 10, the estimated treatment effect is not only sensitive to the order of polynomials in global regression models, but it is also sensitive to the choice of bandwidth h. This figure shows that, when estimated with a triangular kernel, the treatment effects loses its statistical significance around $h < 0.30$ in a linear model (panel A), and around $h < 0.60$ in a quadratic model (panel B).

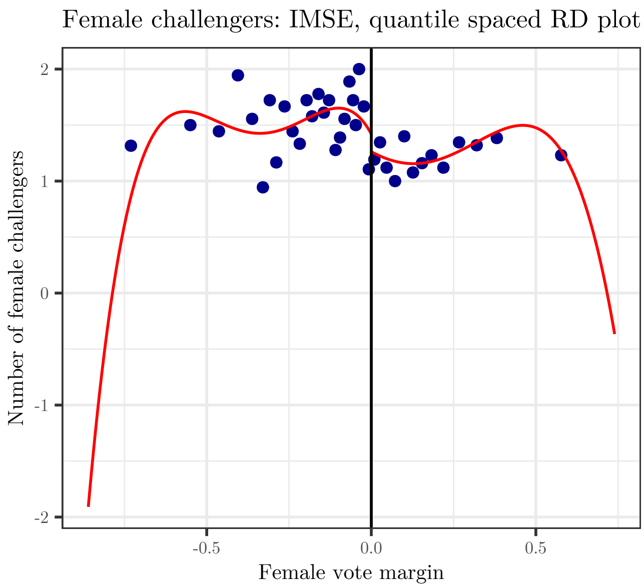

Figure 9: Graphical representation of the relationship between the forcing variable and the number of female challengers during the next election. The graph is produced using the `R` package `rdplot`. The red lines are plots of polynomial regression models with $p = 4$. Bins are quantile-spaced, and the number of bins is calculated with Cattaneo, Idrobo, et al.'s (2019) IMSE-optimal method—which is best suited to 'trace out' the underlying functional relationship.

A local non-parametric approach provides further evidence that based on the available data, the most appropriate conclusion is that there is no discontinuity in the number of female challengers around the cut-off point. In Table 11, the results of a local linear regression model using the CCT MSE-optimal bandwidth of $h = 0.148$ are shown. Although the estimated treatment effect is negative, it is far from statistically significant. Estimates using the CER-optimal bandwidth (shown in Appendix Table 20) back the same conclusion: around the cut-off point, the election of a female politicians has no statistically significant effect on the number of female challengers during the next election.

Table 11: Local linear estimates for the outcome variable: *Female challengers*.

Method	$\hat{\tau}$	s.e.	p-value	95% CI
Conventional	**-0.220**	0.222	0.321	[-0.656, 0.215]
RBC	-0.161	**0.256**	**0.529**	**[-0.663, 0.341]**

Note: Results of the robust bias-corrected (RBC) methods introduced by (Calonico, Cattaneo, and Titiunik 2014) are calculated using the `R` package `rdrobust`. The standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

8.2 Female candidates in nearby districts

Following Broockman (2014), I will also investigate whether the election of a female politician has a spillover

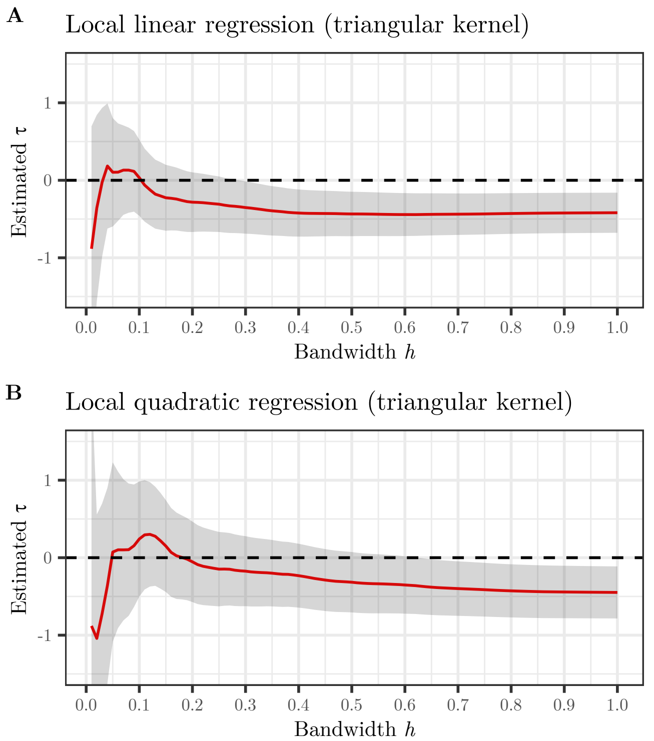

Figure 10: Graphical representation of the sensitivity of the estimated treatment effect using the number of *female challengers* as outcome variable. Panel A and B show the estimated treatment effect $\hat{\tau}$ of local linear and quadratic regression models with bandwidths ranging from 0.01 to 1.00. The shaded areas represent the 95% confidence interval, calculated using heteroskedasticity consistent (HC3) standards errors clustered on the electoral district–representation order level.

effect on the candidacy of women in nearby districts during the next election. The existence of such spillover effects would suggest that female politicians generate role-model affects that transcend the boundaries of their electoral district. To conduct this analysis, additional data is needed.

Using geographic datasets on the federal electoral districts for the Representation Orders of 1987,[10] 1996,[11] 2003,[12] and 2013,[13] the distance between each electoral district was calculated as the distance between the geometric centres of the districts.[14] Using these calculated distances, for each district, the five most nearby districts were selected, and based on the number of female candidates in those nearby districts $j = \{1, ..., 5\}$ during the election following female electoral success, a new outcome variable was constructed: $Y_{it}^{nearby} = \sum_j^5 Y_{jt}$, where Y_{jt} denotes the number of female candidates in district j during election t. Districts with by-elections have been removed from the sample. This leaves a sample of 656 mixed-gender elections from 1988 until 2015. In Appendix B, an example of the construction of this alternative outcome variable is given.

In Table 12, the result of global polynomials regression models with the number of female candidates in the five most-nearby districts as the outcome variable are presented. For none of the polynomial specifications, the treatment effect is statistically significant.

Table 12: Global polynomial estimates for the outcome variable: *Nearby female candidates*.

p	$\hat{\tau}$	s.e.	95% CI	p-value	AIC
0	0.016	0.275	[-0.524, 0.556]	0.953	3481
1	-0.065	0.438	[-0.923, 0.792]	0.881	3484
2	0.379	0.595	[-0.788, 1.546]	0.525	3486
3	0.690	0.741	[-0.762, 2.141]	0.352	3488
4	0.414	0.954	[-1.456, 2.285]	0.664	3490
5	-0.047	1.217	[-2.432, 2.338]	0.969	3494
6	-0.342	1.486	[-3.253, 2.570]	0.818	3498

Note: The degree of the polynomial is indicated by p. All standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

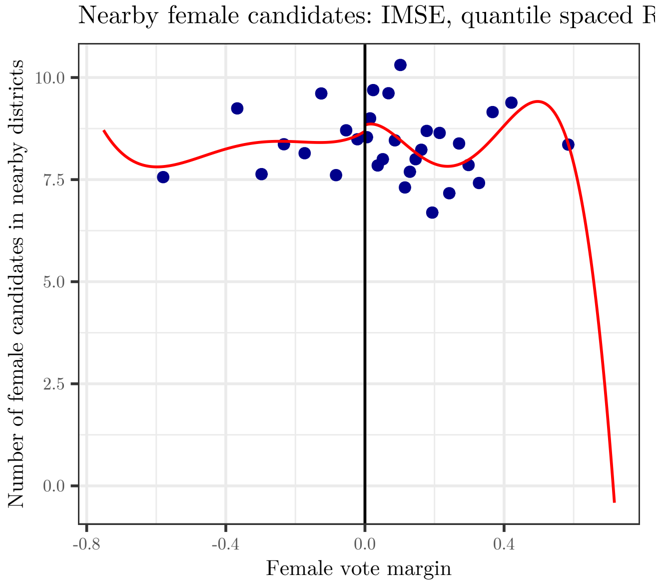

Figure 11: Graphical representation of the relationship between the forcing variable and the number of female candidates in nearby districts during the next election. The graph is produced using the R package `rdplot`. The red lines are plots of polynomial regression models with $p = 4$. Bins are quantile-spaced, and the number of bins is calculated with Cattaneo, Idrobo, et al.'s (2019) IMSE-optimal method—which is best suited to 'trace out' the underlying functional relationship.

A quick inspection of Figure 11 supports the conclusion that the election of a woman to the Canadian House of Commons has no effect on the number of female candidates in nearby electoral districts during the next election. At least at first sight, there seems to be no jump around the cut-off point. The estimates of a local linear regression model with the CCT MSE-optimal bandwidth $h = 0.163$ using a triangular kernel, shown in Table 13, also do no lead to a rejection of the null hypothesis of no effect. Sim-

[10] Statistics Canada (1997a).
[11] Statistics Canada (1997b).
[12] Statistics Canada (2019a).
[13] Statistics Canada (2019b).
[14] Both the geometric centres and the distance between the centers have been determined in R with the commands `st_centroid` and `st_distance` form the package `sf`.

Table 13: Local linear estimates for the outcome variable: *Nearby female candidates*.

Method	$\hat{\tau}$	s.e.	p-value	95% CI
Conventional	**0.261**	0.847	0.758	[-1.399, 1.921]
RBC	0.117	**1.018**	**0.909**	**[-1.878, 2.112]**

Note: Results of the robust bias-corrected (RBC) methods introduced by (Calonico, Cattaneo, and Titiunik 2014) are calculated using the `R` package `rdrobust`. The standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

ilar to Broockman (2014), no evidence can be found for the presence of spillover role-model effects across the boundaries of electoral districts.

8.3 Female candidates in open seat elections

Another complementary inquiry into the existence of the role-model effect can be conducted by analysing the sub-sample of elections where the incumbent politician did *not* seek re-election. Unfortunately the number of such open seat elections is small. Of the full sample of 786 mixed-gender elections, only 117 were followed by an open seat election. This substantially reduced sample size makes inference difficult as it deteriorates the statistical power and thus increases the probability of a false negative (type II error). Therefore, the absence of a statistically significant treatment effect cannot be interpreted as definitive evidence that a treatment effect is truly absent—especially if the true treatment effect is small.

The results of global regression models estimated with the sub-sample of mixed-gender and open seat elections are shown in Table 14. Surprisingly, the model with a polynomial of degree 0—*i.e.* a constant model—suggests the treatment effect to be negative and statistically significant. Since this models assumes the absence of a relationship between the forcing variable and the outcome variable, $\hat{\tau}$ is simply the difference in the mean values of the outcome variable at the left and right of the cut-off point. Thus, on average, there are 0.589 fewer female candidates in the districts with an open seat election where a woman had won the previous mixed-gender election, compared to the open seat districts where a man had won the election. With a p-value of $p = 0.017$, a simple two-tailed t-test indeed suggests that there is a statistically significant difference in means.

Less surprising is the fact that the introduction of a covariate that captures a relationship between the forcing variable and the outcome variable renders the treatment effect statistically insignificant—although for most specifications, the point estimate remains negative. Figure 12 shows how a smoothed function with a polynomial of degree 4 provides a negative point estimate of the treatment effect. However, the low number of bins already reveals the underlying scarcity of data, while the smoothed function also showcases the seemingly erratic boundary be-

Table 14: Global polynomial estimates for sub-sample of open seat elections.

p	$\hat{\tau}$	s.e.	95% CI	p-value	AIC
0	-0.589	0.243	[-1.065, -0.113]	0.017	392
1	0.011	0.384	[-0.742, 0.763]	0.978	391
2	-0.019	0.531	[-1.060, 1.022]	0.972	394
3	-0.852	0.783	[-2.387, 0.684]	0.279	396
4	-0.801	1.064	[-2.887, 1.286]	0.454	400
5	-1.195	1.468	[-4.072, 1.683]	0.418	403
6	-1.052	2.212	[-5.388, 3.284]	0.635	407

Note: The degree of the polynomial is indicated by p. All standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

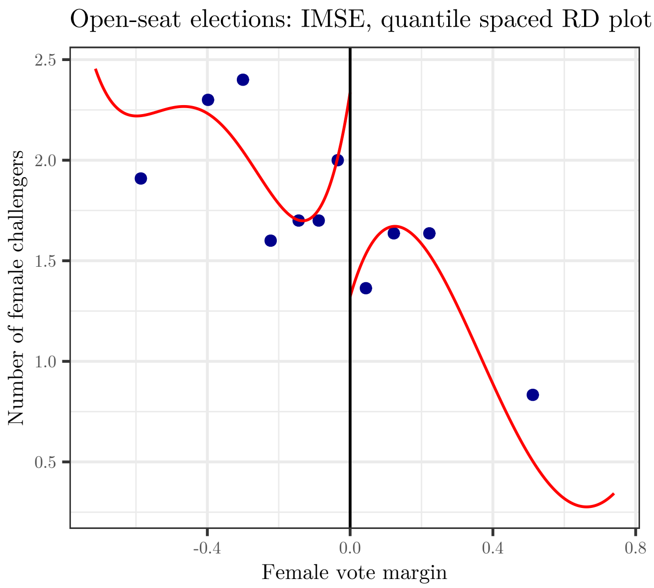

Figure 12: Graphical representation of the relationship between the forcing variable and the number of female candidates in in open-seat elections following a mixed-gender election. The graph is produced using the `R` package `rdplot`. The red lines are plots of polynomial regression models with $p = 4$. Bins are quantile-spaced, and the number of bins is calculated with Cattaneo, Idrobo, et al.'s (2019) IMSE-optimal method—which is best suited to 'trace out' the underlying functional relationship.

haviour which can render global parametric approaches to RD designs unreliable.

Finally, in Table 15, the estimates of a local linear regression model with the CCT MSE-optimal bandwidth $h = 0.227$ ($n = 38 + 31$) are shown. Again, the null hypothesis of no effect cannot be rejected based on this local linear model.

The available evidence thus points to the absence of a treatment effect for open seat districts. However, it must be stressed that the reduced sample size makes this evidence weak. Based on this analysis, it would thus be inappropriate to decisively conclude that the election of a woman does not induce other women to become candidates during the next election. Hopefully, an increase in the amount of data in the future could enable larger-sample RD analyses to shed further lights on these results.

Table 15: Local linear estimates for sub-sample of open seat elections.

Method	$\hat{\tau}$	s.e.	p-value	95% CI
Conventional	**-0.649**	0.731	0.375	[-2.082, 0.784]
RBC	-0.875	**0.873**	**0.316**	**[-2.585, 0.836]**

Note: Results of the robust bias-corrected (RBC) methods introduced by (Calonico, Cattaneo, and Titiunik 2014) are calculated using the R package `rdrobust`. The standard errors are heteroskedasticity consistent (HC3) and clustered on the electoral district–representation order level.

8.4 Political interest and self-efficacy

A final extension that can shed lights on the causal mechanism behind the main results focusses on the intermediating role of the levels of political interest and self-efficacy in the hypothesised role-model effect. As discussed in section 2.1, exposure to female politicians is often found to be positively associated with female interest in politics, and political self-efficacy among women. In turn, these increased levels of political interest and self-efficacy could stimulate the emergence of female political candidates. Especially if the hypothesised positive effect of the election of a woman on the emergence of female political candidates predominantly materialises in the long run, a positive treatment effect on these intermediating outcome variables in the short run could suggest a role-model effect to be present.

Using the 2019 Canadian Election Study (Stephenson et al. 2020), two alternative outcome variables can be constructed. First, respondents were asked to rate their interest in politics in general on a scale from 0 to 10. For each electoral district, the average score to this question among female respondents is used as a measure on *Female political interest*. Secondly, the respondents were asked whether they agreed with the statement: 'Sometimes, politics and government seem so complicated that a person like me can't really understand what's going on'. For each electoral district, the percentage of women answering 'strongly disagree' or 'somewhat disagree' to this question has been used as a measure on *Female political self-efficacy*. Per electoral district, an average of 64.83 female respondents filled in the 2019 Canadian Election Study. Assuming no systemic measurement errors are present, these alternative outcome variables should provide an accurate representation of female political interest and female political self-efficacy in the electoral districts.

For both these alternative outcome variables, it can be tested whether they are affected by the election of a female MP in 2015 using a RD design. However, because only the 2019 edition of the Canadian Election Study is sufficiently large to perform analyses on the level of the federal electoral district, the sample size of this extension is limited ($n = 141$). As with the previous extension in which the small sample of open seat elections was analysed, the results of these RD models must therefore be interpreted with caution.

In Table 16, the global polynomial estimates are presented for the outcome variables *Female political interest* and *Female political self-efficacy*. Somewhat surprisingly, all point estimates are negative, which suggests that around the cut-off point, the election of a female politician in 2015 led to a reduced level of interest in politics among women and decreased the likelihood that women reported to consider politics understandable during the 2019 election. However, with the exemption of the constant model ($p = 0$) for the outcome variable *Female political interest*, none of these negative point estimates provide statistical evidence for the rejection of the null hypothesis of no effect with a significance level $\alpha = 0.05$.

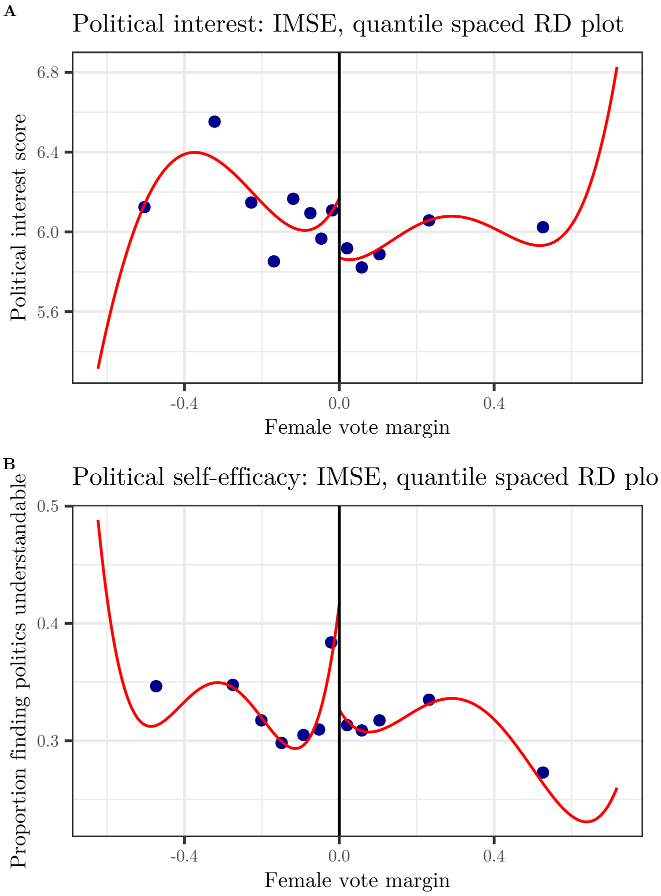

Figure 13: Graphical representation of the relationship between female political interest and political self-efficacy and the female vote margin. The graph is produced using the R package `rdplot`. The red lines are plots of polynomial regression models with $p = 4$. Bins are quantile-spaced, and the number of bins is calculated with Cattaneo, Idrobo, et al.'s (2019) IMSE-optimal method.

In Figure 13, the the RD plots for for these two alternative outcome variables are shown. Although the smoothed fourth-degree polynomial functions do seem to suggest a discontinuity in the outcome variables at the cut-off point, the existence of such discontinuity is less clear when looking at the local binned averages. Especially in panel B of this figure, which shows the RD plot for the outcome variable *Female political self-efficacy*, the suggested discontinuity seems to be largely the result of the surprisingly high score of the outcome variable just

Table 16: Global polynomial estimates for the variables *Female political interest* and *Female political self-efficacy*.

	Female political interest					*Female political self-efficacy*				
p	$\hat{\tau}$	s.e.	95% CI	p-value	AIC	$\hat{\tau}$	s.e.	95% CI	p-value	AIC
0	-0.184	0.074	[-0.330, -0.038]	0.014	166	-0.021	0.014	[-0.047, 0.006]	0.133	-309
1	-0.171	0.104	[-0.375, 0.032]	0.102	167	-0.000	0.020	[-0.040, 0.039]	0.984	-307
2	-0.149	0.145	[-0.434, 0.136]	0.307	170	-0.041	0.027	[-0.095, 0.012]	0.132	-308
3	-0.254	0.198	[-0.643, 0.135]	0.202	173	-0.062	0.037	[-0.134, 0.011]	0.099	-306
4	-0.427	0.255	[-0.927, 0.073]	0.096	172	-0.084	0.045	[-0.172, 0.004]	0.063	-309
5	-0.168	0.339	[-0.832, 0.496]	0.620	172	-0.098	0.059	[-0.213, 0.018]	0.100	-305
6	0.009	0.415	[-0.805, 0.822]	0.984	174	-0.097	0.078	[-0.249, 0.055]	0.214	-302

Note: The degree of the polynomial is indicated by p. All standard errors are heteroskedasticity consistent (HC3). The number of observations $n = 141$. All models are estimated with uniform kernels.

Table 17: Local linear estimates for the outcome variables *Female political interest* and *Female political self-efficacy*.

	Female political interest				*Female political self-efficacy*			
Method	$\hat{\tau}$	s.e.	p-value	95% CI	$\hat{\tau}$	s.e.	p-value	95% CI
Conventional	**-0.210**	0.203	0.300	[-0.608, 0.187]	**-0.080**	0.044	0.068	[-0.166, 0.006]
RBC	-0.180	**0.243**	**0.459**	**[-0.656, 0.296]**	-0.092	**0.052**	**0.078**	**[-0.195, 0.010]**

Note: Results of the robust bias-corrected (RBC) methods introduced by (Calonico, Cattaneo, and Titiunik 2014) are calculated using the `R` package `rdrobust`. The standard errors are heteroskedasticity consistent (HC3).

before the cut-off point. However, given the low amount of observations, this 'abnormal' local average is relatively imprecise and could very well be the product of chance.

The local linear estimates for the outcome variables *Female political interest* and *Female political self-efficacy* are presented in Table 17. As with the global parametric models, the treatment effects are estimated to be negative, but at the conventional significance level $\alpha = 0.05$, the null hypothesis of no treatment cannot be rejected for both outcome variables. As noted before, it is important to remember that the samples used to estimate these models are small. Especially in these local linear models, the number of observations is small with CCT MSE-optimal bandwidths of $h = 0.158$ ($n = 45 + 38$) and $h = 0.117$ ($n = 34 + 32$) respectively.

It must be noted however, that the p-values for the point estimate of the treatment effect of female electoral success on the measure on female political self-efficacy are only slightly larger than $p = 0.05$ in both the local linear regression models, as well as some of the global polynomial models. Figure 14 also shows that the estimated treatments effects for the outcome variable *Female political self-efficacy* remain close to being statistically significant for quite a large range of chosen bandwidths—especially when local quadratic regression models are used.

Although the current results do not provide any decisive evidence on the existence of a treatment effect, the only marginally statistical insignificant negative point estimates of the causal effect of female electoral success on the measure of female political self-efficacy is remarkable and rather unexpected.

9 Discussion and Conclusion

In this thesis, a RD design has been applied to the first-past-the-post electoral system used in Canadian parliamentary elections to obtain a credible estimate of the local average treatment effect of a woman winning a competitive mixed-gender election for a seat in the Canadian House of Commons on the number of female candidates in the subsequent election within the federal electoral district.

The main results presented in section 6 provide credible evidence to conclude with a large degree of confidence that the election of a woman to the Canadian House of Commons does have a positive and statistically significant causal effect on the number of female candidates in the next election within an electoral district. The overall conclusion is neither sensitive to different functional specifications in the global polynomial approach, nor do any reasonable choices of bandwidths in the local non-parametric models provide evidence for alternative conclusions. Finally, both the CER-optimal bandwidths and the robust bias-corrected inference methods introduced by Calonico, Cattaneo, and Titiunik (2014) indicated the main inference to be valid: the treatment effect is positive and statistically significant.

Although the main results unambiguously indicate that the treatment effect is positive and statistically significant, the main results provide less clarity on the size of the treatment effect. The preferred point estimate, based on a local linear regression model with a triangular kernel and a MSE-optimal bandwidth ($h = 0.146$) indicates that the election of a woman causes an increase in the number

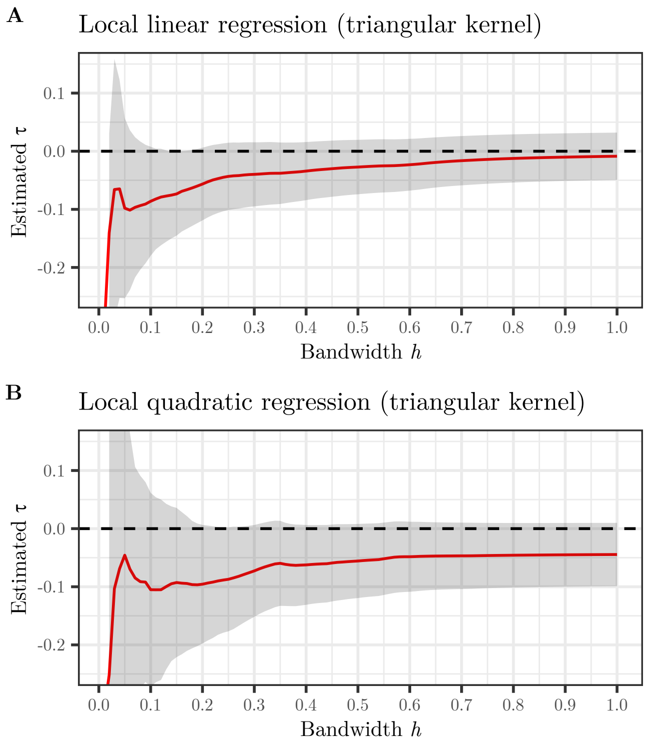

Figure 14: Graphical representation of the sensitivity of the estimated treatment effect using the measure on *Female political self-efficacy* as outcome variable. Panel A and B show the estimated treatment effect $\hat{\tau}$ of local linear and quadratic regression models with bandwidths ranging from 0.01 to 1.00. The shaded areas represent the 95% confidence interval, calculated using heteroskedasticity consistent (HC3) standards errors clustered on the electoral district–representation order level.

of female candidates during the subsequent election of $\hat{\tau} = 0.697$. Additional evidence from regression models with smaller bandwidths and the covariate-adjusted local linear regression models suggest that this point estimate is more likely to be an underestimation than an overestimation. Both the selection of smaller bandwidths, as well as covariate adjustment result in larger point estimates.

The causal mechanisms behind the main results remain somewhat unclear. As some other authors have pointed out, the increase in female candidates might simply be attributable to the tendency of incumbent candidates to seek re-election (Bhalotra et al. 2018; Broockman 2014). To investigate whether (part of) the treatment effect can be attributed to a role-model effect, four alternative regression models have been explored in section 8. All these alternative models aim to provide insight into the question whether the election of a woman induces *other* women to run for election.

In the first two extensions, the number of female challengers, and the number of female candidates in nearby electoral districts have been used as alternative outcome variables. In the third extension, only a sub-sample of open seat elections—*i.e.* elections where the incumbent member of parliament did not participate—has been used to estimate the treatment effect. In the fourth extension, data from the from the 2019 Canadian Election Study was used to construct two alternative outcome variables measuring the levels female political interest and political self-efficacy within the electoral district.

In none of these extensions, the null hypothesis of no treatment effect could be rejected. These result can be interpreted as follows: first, the fact that the election of a woman does not increase the number of female challengers suggests that there is no *cross-party* role-model or contagion effect. For example, the election of a woman from the Liberal Party does not seem to increase the likelihood that the nominee of the Conservatives is female during the next election. Secondly, the absence of an increase in female candidates in nearby electoral districts indicates that female electoral success does not generate a spillover role-model effect that transcends the boundaries of the electoral district. Finally, both the lack of a statistically significant treatment effect when only the subsample of open seat elections is used in estimation, as well as the lack of a statistically significant treatment effect on female political interest and political self-efficacy must be interpreted with caution. For these two extensions, the low sample-size substantially increases the likelihood of a false negative, and at most, the absence of a statistically significant effect indicates that there is no *large* role-model effect.

Overall, the available evidence thus suggests that the election of a female candidates does increase the number of female candidates during the next election, but most likely, this increase is not attributable to a role-model effect. Instead, the increase in female candidates seems to be largely the result of the tendency among incumbents to run for re-election. The implication of this result is that policy measures such as (mandatory) gender quota in politics can obviously result in mechanical increases in the level of female political representation, but such increases are unlikely to improve other women's political participation in Canada—or at least in the short run. Perhaps because the share of female MPs has already reached a certain 'threshold' level in Canada, biases, barriers, and disincentives faced by aspiring female politicians can remain unaffected by an increased presence of female role-models (Gilardi 2015).

The results of this thesis do show some new interesting areas open for exploration. First, the methods used in this thesis can be easily applied to other countries which use first-past-the-post voting systems to investigate whether similar conclusions can be drawn outside Canada. Importantly, the result obtained in the analysis of a relatively well-scoring country in terms of gender equality such as Canada might not be applicable in countries where the gender gaps in politics have remained even more pronounced, such as Malaysia.

Secondly, and perhaps more interestingly, it is important to realise that this study only investigates a very specific type of role-model effect: a *horizontal* role-model effect. Although most evidence presented in this thesis suggests that the election of a woman does not increase the inflow of female candidates in Canadian *federal* election, it might be possible that exposure to high-level female politicians increases the number of women entering politics at lower levels—*e.g.* in municipal or provincial elected bodies. The investigation of such *vertical* role-model effects could provide more insight in the dynamics of female political representation, especially since local politics can pave the way to higher office as it allows potential candidates to build networks and to gain political experience (Buckley et al. 2015; Standing Committee on the Status of Women 2019).

Furthermore, in this thesis, I have already made an initial attempt to use a RD design to explore have a wider range of aspects of political participation among women is affected by the exposure to female politicians. In the fourth extension, I aimed to follow the examples of D. E. Campbell and Wolbrecht (2006) and Beaman, Duflo, et al. (2012) to investigate whether women in districts where a woman won the election show increased levels of political interest and self-efficacy. The methods used in this thesis can be easily applied to study any of the hypothesized effects of exposure to female politicians in more detail. For example, it could be investigated whether the political attitudes and behaviours of adolescent girls are affected by the presence of a female MP, or surveys among the influential local Electoral District Associations could be conducted to analyse whether local political elites are impacted by their exposure to women occupying political office.

The obvious drawback of all these follow-up studies is that it requires major efforts for data collection, as data on the gender of provincial and municipal political candidates is not readily available, nor is there an abundance of publicly available data on a wider range of measures of female political engagement at the level of federal electoral districts—with the exception of the 2019 Canadian Election Study. However, as more election are held in the future, more data might hopefully become available.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity*. *Q. J. Econ.*, *115*(3), 715–753. https://doi.org/10.1162/003355300554881

Alexander, A. C. (2012). Change in Women's Descriptive Representation and the Belief in Women's Ability to Govern: A Virtuous Cycle. *Polit. Gend.*, *8*(04), 437–464. https://doi.org/10.1017/S1743923X12000487

Anastasopoulos, L. (2016). Estimating the gender penalty in House of Representative elections using a regression discontinuity design. *Elect. Stud.*, *43*, 150–157. https://doi.org/10.1016/j.electstud.2016.04.008

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics*. Princeton University Press. https://doi.org/10.2307/j.ctvcm4j72

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *J. Econ. Perspect.*, *24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3

Armstrong, T. B., & Kolesár, M. (2020). Simple and honest confidence intervals in nonparametric regression. *Quant. Econom.*, *11*(1), 1–39. https://doi.org/https://doi.org/10.3982/QE1199

Atkeson, L. R. (2003). Not All Cues Are Created Equal: The Conditional Impact of Female Candidates on Political Engagement. *J. Polit.*, *65*(4), 1040–1061. https://doi.org/10.1111/1468-2508.t01-1-00124

Atkeson, L. R., & Carrillo, N. (2007). More is Better: The Influence of Collective Female Descriptive Representation on External Efficacy (2007/03/01). *Polit. Gend.*, *3*(01), 79–101. https://doi.org/10.1017/S1743923X0707002X

Atkeson, L. R., & Krebs, T. B. (2008). Press Coverage of Mayoral Candidates. *Polit. Res. Q.*, *61*(2), 239–252. https://doi.org/10.1177/1065912907308098

Baltrunaite, A., Casarico, A., Profeta, P., & Savio, G. (2019). Let the voters choose women. *J. Public Econ.*, *180*, 104085. https://doi.org/10.1016/j.jpubeco.2019.104085

Barnes, T. D., & Burchard, S. M. (2012). "Engendering" Politics. *Comp. Polit. Stud.*, *46*(7), 767–790. https://doi.org/10.1177/0010414012463884
doi: 10.1177/0010414012463884

Bashevkin, S. (2011). Women's representation in the House of Commons: A stalemate? *Can. Parliam. Rev.*, *34*(1), 17–22.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful Women: Does Exposure Reduce Bias? *. *Q. J. Econ.*, *124*(4), 1497–1540. https://doi.org/10.1162/qjec.2009.124.4.1497

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science (80-.).*, *335*(6068), 582–586.

Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2018). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *Rev. Econ. Stud.*, *86*(1), 191–239. https://doi.org/10.1093/restud/rdy032

Bhalotra, S., Clots-Figueras, I., & Iyer, L. (2018). Pathbreakers? Women's Electoral Success and Future Political Participation. *Econ. J.*, *128*(613), 1844–1878. https://doi.org/10.1111/ecoj.12492

Black, D. A., Galdo, J., & Smith, J. A. (2007). Evaluating the Worker Profiling and Reemployment Services System Using a Regression Discontinuity Approach. *Am. Econ. Rev.*, *97*(2), 104–107. https://doi.org/10.1257/aer.97.2.104

Black, J. H., & Erickson, L. (2003). Women candidates and voter bias: do women politicians need to be better? *Elect. Stud.*, *22*(1), 81–100. https://doi.org/10.1016/S0261-3794(01)00028-2

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs about Gender. *Am. Econ. Rev.*, *109*(3), 739–773. https://doi.org/10.1257/aer.20170007

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

Broockman, D. E. (2014). Do female politicians empower women to vote or run for office? A regression discontinuity approach. *Elect. Stud.*, *34*, 190–204. https://doi.org/10.1016/j.electstud.2013.10.002

Buckley, F., Mariani, M., McGing, C., & White, T. (2015). Is Local Office a Springboard for Women to Dáil Éireann? *J. Women, Polit. Policy*, *36*(3), 311–335. https://doi.org/10.1080/1554477X.2015.1050912

Bühlmann, M., & Schädel, L. (2012). Representation matters: The impact of descriptive women's representation on the political involvement of women. *Representation*, *48*(1), 101–114. https://doi.org/10.1080/00344893.2012.653246
doi: 10.1080/00344893.2012.653246

Burns, N., Schlozman, K. L., & Verba, S. (2001). *The Private Roots of Public Action: Gender, Equality, and Political Participation*. Harvard University Press.

Butler, D. M. (2009). A regression discontinuity design analysis of the incumbency advantage and tenure in the U.S. House. *Elect. Stud.*, *28*(1), 123–128. https://doi.org/10.1016/j.electstud.2008.07.010

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

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

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

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Rev. Econ. Stat.*, *101*(3), 442–451. https://doi.org/10.1162/rest_a_00760

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

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal Data-Driven Regression Discontinuity Plots. *J. Am. Stat. Assoc.*, *110*(512), 1753–1769. https://doi.org/10.1080/01621459.2015.1017578

Campbell, D. E., & Wolbrecht, C. (2006). See Jane Run: Women Politicians as Role Models for Adolescents. *J. Polit.*, *68*(2), 233–247. https://doi.org/10.1111/j.1468-2508.2006.00402.x

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handb. res. teach.* Chicago, Rand McNally.

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *Q. J. Econ.*, *125*(3), 1101–1144. https://doi.org/10.1162/qjec.2010.125.3.1101

Carty, R. K., Eagles, D. M., & Sayers, A. (2003). Candidates and Local Campaigns: Are there Just Four Canadian Types? *Party Polit.*, *9*(5), 619–636. https://doi.org/10.1177/13540688030095006

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press. https://doi.org/10.1017/9781108684606

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple Local Polynomial Density Estimators. *J. Am. Stat. Assoc.*, *115*(531), 1449–1455. https://doi.org/10.1080/01621459.2019.1635480

Caughey, D., & Sekhon, J. S. (2011). Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008 (2017/01/04). *Polit. Anal.*, *19*(4), 385–408. https://doi.org/10.1093/pan/mpr032

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

Cheng, C., & Tavits, M. (2011). Informal Influences in Selecting Female Political Candidates. *Polit. Res. Q.*, *64*(2), 460–471. https://doi.org/10.1177/1065912909349631
Copyright - Copyright SAGE PUBLICATIONS, INC. Jun 2011; Document feature - Tables; ; Last updated - 2018-10-06

Cheng, M.-Y., Fan, J., & Marron, J. S. (1997). On Automatic Boundary Corrections. *Ann. Stat.*, *25*(4), 1691–1708. https://doi.org/https://doi.org/10.1214/aos/1031594737

Chong, A., & Ferrara, E. L. (2009). Television and Divorce: Evidence from Brazilian Novelas. *J. Eur. Econ. Assoc.*, *7*(2-3), 458–468. https://doi.org/10.1162/JEEA.2009.7.2-3.458

Clayton, A., O'Brien, D. Z., & Piscopo, J. M. (2019). All Male Panels? Representation and Democratic Legitimacy. *Am. J. Pol. Sci.*, *63*(1), 113–129. https://doi.org/https://doi.org/10.1111/ajps.12391

Cross, W. P. (2016). The Importance of Local Party Activity in Understanding Canadian Politics: Winning from the Ground Up in the 2015 Federal Election (2016/11/04). *Can. J. Polit. Sci.*, *49*(4), 601–620. https://doi.org/10.1017/S0008423916000962

Cross, W. P., & Pruysers, S. (2019). The Local Determinants of Representation: Party Constituency Associations, Candidate Nomination and Gender. *Can. J. Polit. Sci.*, *52*(3), 557–574. https://doi.org/10.1017/S0008423919000064

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. http://www.jstor.org/stable/j.ctv1c29t27

Dolan, K. (2006). Symbolic Mobilization? The Impact of Candidate Sex in American Elections. *Am. Polit. Res.*, *34*(6), 687–704. https://doi.org/10.1177/1532673X06289155
doi: 10.1177/1532673X06289155

Dovi, S. (2002). Preferable Descriptive Representatives: Will Just Any Woman, Black, or Latino Do? *Am. Polit. Sci. Rev.*, *96*(4), 729–743. http://www.jstor.org/stable/3117507

Dovi, S. (2018). Political Representation. In E. N. Zalta (Ed.), {*stanford*} *encycl. philos.* (Fall 2018). Metaphysics Research Lab, Stanford University.

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

Eggers, A. C., & Hainmueller, J. (2009). MPs for Sale? Returns to Office in Postwar British Politics (2009/10/30). *Am. Polit. Sci. Rev.*, *103*(4), 513–533. https://doi.org/10.1017/S0003055409990190

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Chapman & Hall/CRC. https://doi.org/10.1201/9780203748725

Ferguson, S. J. (2016). *Women and Education: Qualifications, Skills and Technology*. https://www150.statcan.gc.ca/n1/pub/89-503-x/2015001/article/14640-eng.htm

Ferreira, F., & Gyourko, J. (2014). Does gender matter for political leadership? The case of U.S. mayors. *J. Public Econ.*, *112*, 24–39. https://doi.org/10.1016/j.jpubeco.2014.01.006

Fox, R. L., & Lawless, J. L. (2004). Entering the Arena? Gender and the Decision to Run for Office. *Am. J. Pol. Sci.*, *48*(2), 264–280. https://doi.org/10.2307/1519882

Fox, R. L., & Lawless, J. L. (2011). Gendered Perceptions and Political Candidacies: A Central Barrier to Women's Equality in Electoral Politics. *Am. J. Pol. Sci.*, *55*(1), 59–73. https://doi.org/10.1111/j.1540-5907.2010.00484.x

Franks, C. E. S. (2007). Members and Constituency Roles in the Canadian Federal System. *Reg. Fed. Stud.*, *17*(1), 23–45. https://doi.org/10.1080/13597560701189602

Fridkin, K. L., & Kenney, P. J. (2014). How the Gender of U.S. Senators Influences People's Understanding and Engagement in Politics. *J. Polit.*, *76*(4), 1017–1031. https://doi.org/10.1017/S0022381614000589
doi: 10.1017/S0022381614000589

Galandy, J., & Tavcer, D. S. (2019). Improving gender representation in Canadian federal politics and parliament. *Can. Parliam. Rev.*, 15.

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

Gidengil, E., Giles, J., & Thomas, M. (2008). The Gender Gap in Self-Perceived Understanding of Politics in Canada and the United States (2008/11/25). *Polit. & Gend.*, *4*(4), 535–561. https : / / doi . org / DOI : 10 . 1017 / S1743923X08000469

Gilardi, F. (2015). The Temporary Importance of Role Models for Women's Political Representation. *Am. J. Pol. Sci.*, *59*(4), 957–970. https://doi.org/10.1111/ajps.12155
https://doi.org/10.1111/ajps.12155

Hahn, J., Todd, P., & der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Hartviksen, J. (2015). Interrogating Trudeau's Brand of Equality "Because it's 2015". https://blogs.lse.ac.uk/gender/2015/11/09/interrogating-trudeaus-brand-of-equality-because-its-2015/

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching As An Econometric Evaluation Estimator. *Rev. Econ. Stud.*, *65*(2), 261–294. https://doi.org/10.1111/1467-937X.00044

Heitshusen, V., Young, G., & Wood, D. M. (2005). Electoral Context and MP Constituency Focus in Australia, Canada, Ireland, New Zealand, and the United Kingdom. *Am. J. Pol. Sci.*, *49*(1), 32–45. https://doi.org/https://doi.org/10.1111/j.0092-5853.2005.00108.x

Holland, P. W. (1986). Statistics and Causal Inference. *J. Am. Stat. Assoc.*, *81*(396), 945–960. https://doi.org/10.2307/2289064

Hyytinen, A., Meriläinen, J., Saarimaa, T., Toivanen, O., & Tukiainen, J. (2018). When does regression discontinuity design work? Evidence from random election outcomes. *Quant. Econom.*, *9*(2), 1019–1051. https://doi.org/10.3982/QE864
https://doi.org/10.3982/QE864

Imbens, G. W., & Kalyanaraman, K. (2009). *Optimal Bandwidth Choice for the Regression Discontinuity Estimator* (Working Paper No. 14726). National Bureau of Economic Research. https://doi.org/10.3386/w14726

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

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

Imbens, G. W., & Rubin, D. B. (2010). Rubin causal model. *Microeconometrics* (pp. 229–241). Springer. https://doi.org/10.1057/9780230280816_28

Jankowski, M., Marcinkiewicz, K., & Gwiazda, A. (2019). The Effect of Electing Women on Future Female Candidate Selection Patterns: Findings from a Regression Discontinuity Design (2019/06/06). *Polit. Gend.*, *15*(2), 182–210. https://doi.org/10.1017/S1743923X19000096

Jensen, R., & Oster, E. (2009). The Power of TV: Cable Television and Women's Status in India. *Q. J. Econ.*, *124*(3), 1057–1094. https://doi.org/10.1162/qjec.2009.124.3.1057

Karp, J. A., & Banducci, S. A. (2008). When politics is not just a man's game: Women's representation and political engagement. *Elect. Stud.*, *27*(1), 105–115. https://doi.org/10.1016/j.electstud.2007.11.009

Kunze, A., & Miller, A. R. (2017). Women Helping Women? Evidence from Private Sector Data on Workplace Hierarchies. *Rev. Econ. Stat.*, *99*(5), 769–775. https://doi.org/10.1162/REST_a_00668

La Ferrara, E., Chong, A., & Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. *Am. Econ. J. Appl. Econ.*, *4*(4), 1–31. https://doi.org/10.1257/app.4.4.1

Ladam, C., Harden, J. J., & Windett, J. H. (2018). Prominent Role Models: High-Profile Female Politicians and the Emergence of Women as Candidates for Public Office. *Am. J. Pol. Sci.*, *62*(2), 369–381. https://doi.org/10.1111/ajps.12351

Latu, I. M., Mast, M. S., Lammers, J., & Bombari, D. (2013). Successful female leaders empower women's behavior in leadership tasks. *J. Exp. Soc. Psychol.*, *49*(3), 444–448. https://doi.org/10.1016/j.jesp.2013.01.003

Lawless, J. L. (2004). Politics of Presence? Congresswomen and Symbolic Representation. *Polit. Res. Q.*, *57*(1), 81–99. https://doi.org/10.1177/106591290405700107 doi: 10.1177/106591290405700107

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

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *J. Econ. Lit.*, *48*(2), 281–355. https://doi.org/10.1257/jel.48.2.281

Liu, S.-J. S., & Banaszak, L. A. (2017). Do Government Positions Held by Women Matter? A Cross-National Examination of Female Ministers' Impacts on Women's Political Participation. *Polit. Gend.*, *13*(01), 132–162. https://doi.org/10.1017/S1743923X16000490

Ludwig, J., & Miller, D. L. (2005). *Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design* (Working Paper No. 11702). National Bureau of Economic Research. https://doi.org/10.3386/w11702

Mansbridge, J. (1999). Should blacks represent blacks and women represent women? A contingent" yes". *J. Polit.*, *61*(3), 628–657. https://doi.org/10.2307/2647821

Mariani, M. (2008). A gendered pipeline? The advancement of state legislators to congress in five states. *Polit. Gend.*, *4*(2), 285–308. https://doi.org/10.1017/S1743923X08000196

Mariani, M., Marshall, B. W., & Mathews-Schultz, A. L. (2015). See Hillary Clinton, Nancy Pelosi, and Sarah Palin Run? Party, Ideology, and the Influence of Female Role Models on Young Women. *Polit. Res. Q.*, *68*(4), 716–731. http://www.jstor.org.eur.idm.oclc.org/stable/24637811

Matsa, D. A., & Miller, A. R. (2011). Chipping away at the Glass Ceiling: Gender Spillovers in Corporate Leadership. *Am. Econ. Rev.*, *101*(3), 635–639. https://doi.org/10.1257/aer.101.3.635

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

Norris, P., Lovenduski, J., & Campbell, R. (2004). *Gender and political participation*. The Electoral Commission.

Palmer, B., & Simon, D. M. (2005). When Women Run Against Women: The Hidden Influence of Female Incumbents in Elections to the U.S. House of Representatives, 1956–2002 (2005/09/09). *Polit. Gend.*, *1*(01), 39–63. https://doi.org/10.1017/S1743923X05050026

Paxton, P., Kunovich, S., & Hughes, M. M. (2007). Gender in Politics. *Annu. Rev. Sociol.*, *33*(1), 263–284. https://doi.org/10.1146/annurev.soc.33.040406.131651

Porter, C., & Serra, D. (2020). Gender Differences in the Choice of Major: The Importance of Female Role Models. *Am. Econ. J. Appl. Econ.*, *12*(3), 226–254. https://doi.org/10.1257/app.20180426

Reingold, B., & Harrell, J. (2010). The impact of descriptive representation on women's political engagement: Does party matter? *Polit. Res. Q.*, *63*(2), 280–294. https://doi.org/10.1177/1065912908330346 doi: 10.1177/1065912908330346

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.*, *66*(5), 688. https://doi.org/10.1037/h0037350

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Am. Stat. Assoc.*, *100*(469), 322–331. https://doi.org/10.1198/016214504000001880

Sanbonmatsu, K. (2006). Do Parties Know That "Women Win"? Party Leader Beliefs about Women's Electoral Chances (2006/11/28). *Polit. Gend.*, *2*(04), 431–450. https://doi.org/10.1017/S1743923X06060132

Sapiro, V. (1981). Research Frontier Essay: When Are Interests Interesting? The Problem of Political Representation of Women. *Am. Polit. Sci. Rev.*, *75*(3), 701–716. https://doi.org/10.2307/1960962

Shair-Rosenfield, S. (2012). The alternative incumbency effect: Electing women legislators in Indonesia. *Elect. Stud.*, *31*(3), 576–587. https://doi.org/10.1016/j.electstud.2012.05.002

Splawa-Neyman, J., Dabrowska, D. M., & Speed, T. P. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Stat. Sci.*, *5*(4), 465–472. https://doi.org/10.2307/2245382 Originally published in Polish in 1923

Standing Committee on the Status of Women. (2019). *Elect Her: A Roadmap for Improving the Representation of Women in Canadian Politics* (tech. rep.). House of Commons of Canada.

Tate, G., & Yang, L. (2015). Female leadership and gender equity: Evidence from plant closure. *J. financ. econ.*, *117*(1), 77–97. https://doi.org/10.1016/j.jfineco.2014.01.004

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

Thomas, M. (2012). The Complexity Conundrum: Why Hasn't the Gender Gap in Subjective Political Competence Closed? (2012/07/30). *Can. J. Polit. Sci.*, *45*(2), 337–358. https://doi.org/10.1017/S0008423912000352

Thomas, M., & Bodet, M. A. (2013). Sacrificial lambs, women candidates, and district competitiveness in Canada. *Elect. Stud.*, *32*(1), 153–166. https://doi.org/10.1016/j.electstud.2012.12.001

UN. (2014). *Beijing Declaration and Platform for Action*. UN Women.

UN Women. (2021). Facts and figures: Women's leadership and political participation. https://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures

van Kippersluis, H., O'Donnell, O., & van Doorslaer, E. (2011). Long-Run Returns to Education: Does Schooling Lead to an Extended Old Age? *J. Hum. Resour.*, *46*(4), 695–721. http://www.jstor.org/stable/41304839

Verba, S., Burns, N., & Schlozman, K. L. (1997). Knowing and Caring about Politics: Gender and Political Engagement. *J. Polit.*, *59*(4), 1051–1072. https://doi.org/10.2307/2998592

Wolbrecht, C., & Campbell, D. E. (2007). Leading by Example: Female Members of Parliament as Political Role Models. *Am. J. Pol. Sci.*, *51*(4), 921–939. https://doi.org/10.1111/j.1540-5907.2007.00289.x

World Economic Forum. (2021). *Global Gender Gap Report 2021*. World Economic Forum.

Data sources

Federation of Law Societies of Canada. (2018). 2018 Statistical Report of the Federation of Law Societies of Canada. https://flsc.ca/resources/

Inter-Parliamentary Union. (2021). Global and regional averages of women in national parliaments. https://data.ipu.org/women-averages

Statistics Canada. (1997a). Cartographic Boundary Files (CBF), 1996 Census - Federal Electoral Districts, 1987 Representation Order. http://geo.scholarsportal.info/%7B%5C#%7Dr/details/%7B%5C_%7Duri@=3577293255%7B%5C$%7DDLI%7B%5C_%7D1996%7B%5C_%7DCensus%7B%5C_%7DCBF%7B%5C_%7DEng%7B%5C_%7DNat%7B%5C_%7Dfed%7B%5C_%7D1987

Statistics Canada. (1997b). Cartographic Boundary Files (CBF), 1996 Census - Federal Electoral Districts, 1996 Representation Order. http://geo.scholarsportal.info/%7B%5C#%7Dr/details/%7B%5C_%7Duri@=3577293255%7B%5C$%7DDLI%7B%5C_%7D1996%7B%5C_%7DCensus%7B%5C_%7DCBF%7B%5C_%7DEng%7B%5C_%7DNat%7B%5C_%7Dfed%7B%5C_%7D1996

Statistics Canada. (2019a). 2006 Census - Boundary files. Federal electoral districts (2003 Representation Order). https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2006-eng.cfm

Statistics Canada. (2019b). 2016 Census - Boundary files. Federal electoral districts (2013 Representation Order). https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm

Stephenson, L. B., Harell, A., Rubenson, D., & Loewen, P. J. (2020). 2019 Canadian Election Study - Online Survey. https://doi.org/10.7910/DVN/DUS88V

A Optimal bandwidth calculation

A.1 LM cross-validation

In Figure 15, the results of the Ludwig and D. L. Miller (2005) cross-validation procedure, as calculated using the algorithm described in detail by Imbens and Lemieux (2008) are presented. The dashed vertical line indicates the optimal bandwidth at which the cross-validation criterion $\frac{1}{N}\sum_i^N (Y_i - \hat{Y}(\tilde{v}_i))^2$ takes its lowest value.

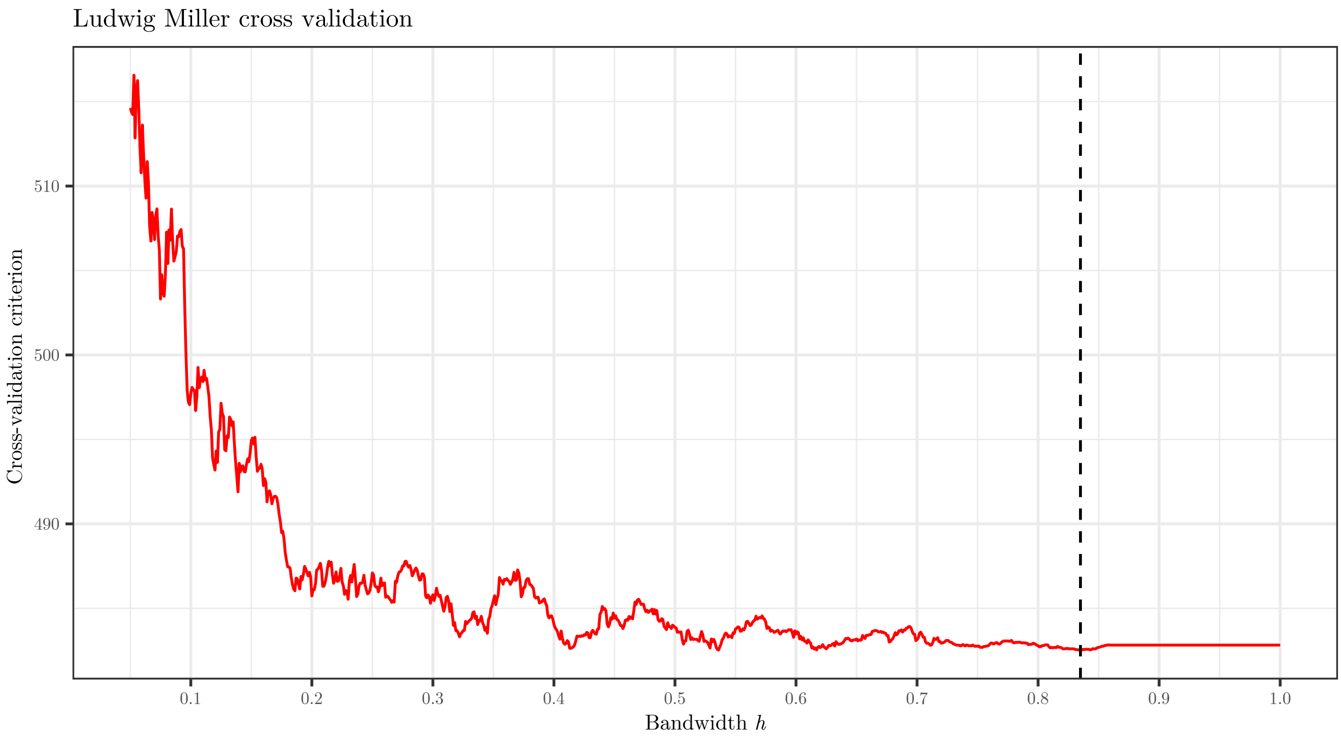

Figure 15: Plot of the Ludwig and D. L. Miller (2005) cross-validation criterion.

A.2 IK bandwidth

Imbens and Kalyanaraman (2012) propose the following optimal bandwidth:

$$\hat{h}_{opt} = C_K \Big(\frac{\hat{\sigma^2_-}(c) + \hat{\sigma^2_+}(c)}{\hat{f}(c) \cdot ((\hat{m}^{(2)}_+(c) - \hat{m}^{(2)}_-(c))^2 + (\hat{r}_+ - \hat{r}_-)} \Big)^{1/5} \cdot N^{-1/5} \tag{3}$$

where C_K is a constant that depends on the type of kernel used, $\hat{\sigma^2_-}(c)$ and $\hat{\sigma^2_+}(c)$ denote $\text{Var}(Y_i|\tilde{v}_i)$ at the cut-off point c from the left and the right, $\hat{f}(c)$ is the estimate of the density of the forcing variable $\tilde{v}_i$ at the cut-off point, $\hat{m}^{(2)}_+(c)$ and $\hat{m}^{(2)}_-(c)$ are the second derivatives of the conditional mean $\mathbb{E}(Y_i|\tilde{v}_i = c)$ estimated with a quadratic regression model separately to the left and right of the cut-off point, and $\hat{r}_+$ and $\hat{r}_+$ are regularisation terms.

Following the algorithm of Imbens and Kalyanaraman (2012), this optimal bandwidth can be easily computed for the RD design with normalised female margin of victory as the forcing variable $\tilde{v}_i$ with the cut-off point $c = 0$ and the number of female candidates in the next election as the outcome variable Y_i:

1. Calculate the pilot bandwidth h_1 to estimate $\hat{f}(0)$ and the conditional variances $\hat{\sigma^2_-}(0)$ and $\hat{\sigma^2_+}(0)$:

   $$h_1 = 1.84 \cdot S_{\tilde{v}_i} \cdot N^{1/5} = 1.84 \cdot 0.067 \cdot 786^{1/5} = 0.1255$$

2. Calculate $\hat{f}(0)$ using the number of observation within the pilot bandwidth N_{h_1-} and N_{h_1+}:

   $$\hat{f}(0) = \frac{N_{h_1-} + N_{h_1+}}{2 \cdot N \cdot h_1} = \frac{175 + 144}{2 \cdot 786 \cdot 0.1255} = 1.616$$

3. Estimate the conditional variances:

$$\hat{\sigma}^2_-(0) = \frac{1}{N_{h_1-} - 1} \sum_{i:-h_1 \leq \tilde{v}_i < 0} (Y_i - \bar{Y}_{h_1-})^2 = 1.085$$

and

$$\hat{\sigma}^2_+(0) = \frac{1}{N_{h_1+} - 1} \sum_{i:0 < \tilde{v}_i \leq h_1} (Y_i - \bar{Y}_{h_1+})^2 = 1.330$$

4. Estimate $\hat{m}^{(3)}$ by fitting a global cubic regression model with a jump at the cut-off point

$$Y_i = \alpha + \gamma_1 \cdot D_i + \gamma_2 \cdot \tilde{v}_i + \gamma_3 \cdot \tilde{v}_i^2 + \gamma_4 \cdot \tilde{v}_i^3 + \epsilon_i$$

with

$$\hat{m}^{(3)} = 4 \cdot \gamma_4 = 4 \cdot -1.237 = -7.421$$

5. Calculate two new pilot bandwidths

$$h_{2-} = 3.56 \cdot \Big(\frac{\hat{\sigma}^2_-(0)}{\hat{f}(0) \cdot (\hat{m}^{(3)})^2}\Big)^{1/7} \cdot N_-^{-1/7} = 0.791$$

and

$$h_{2+} = 3.56 \cdot \Big(\frac{\hat{\sigma}^2_+(0)}{\hat{f}(0) \cdot (\hat{m}^{(3)})^2}\Big)^{1/7} \cdot N_+^{-1/7} = 0.851$$

6. Using these two pilot bandwidths, fit two quadratic models to the left and right of the cut-off point to estimate $\hat{m}^{(2)}_- = 0.7637$ and $\hat{m}^{(2)}_+ = -5.97$

7. Calculate the regularisation terms $\hat{r}_-$ and $\hat{r}_+$ as:

$$\hat{r}_- = \frac{2160 \cdot \hat{\sigma^2_-}(0)}{N_{2-} \cdot h^4_{2-}} = \frac{2160 \cdot 1.085}{452 \cdot 0.791^4} = 13.25$$

and

$$\hat{r}_+ = \frac{2160 \cdot \hat{\sigma^2_+}(0)}{N_{2+} \cdot h^4_{2+}} = \frac{2160 \cdot 1.33}{333 \cdot 0.851^4} = 16.45$$

8. With all these estimates, it is possible to calculate the IK optimal bandwidth:

$$\hat{h}_{opt} = 3.4375\Big(\frac{1.085 + 1.330}{1.616 \cdot (0.851 - 0.7637)^2 + (16.45 - 13.25)}\Big)^{1/5} \cdot 786^{-1/5} = 0.414$$

B Nearby districts

Figure 16 presents an example of how the alternative outcome variable on the number of female candidates in nearby federal electoral districts is constructed. This map shows the electoral district in the centre of Toronto for the Representation Order of 2003. For each district, the geometric centres are shown by the blue points. For the district *Toronto Centre*—shown in red—the five districts with the most nearby geometric centres are coloured blue. The alternative outcome variable for *Toronto Centre* is the sum of the female candidates in the elections of the blue districts.

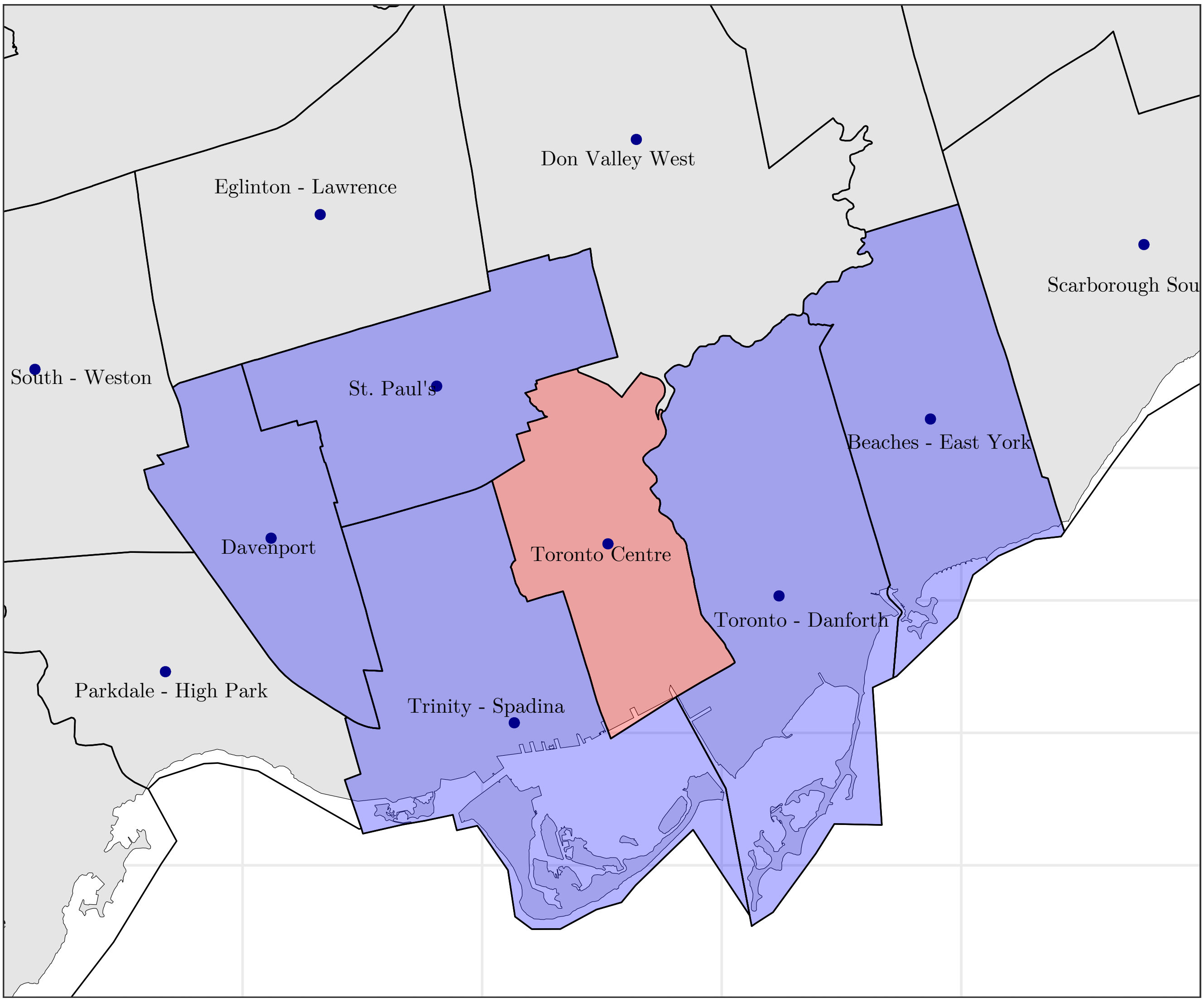

Figure 16: Map of electoral district in Toronto (Representation Order 2003) The five district in blue have geometric centers at the least distance from the district in red—Toronto Centre.

C Complementary graphs and figures

Table 18: RD estimates from global polynomial regression models.

		Nonclustered			Clustered			
p	$\hat{\tau}$	s.e.	95% CI	p-value	s.e.	95% CI	p-value	AIC
0	0.549	0.083	[0.387, 0.711]	0.000	0.087	[0.379, 0.720]	0.000	2452
1	0.489	0.132	[0.230, 0.748]	0.000	0.138	[0.219, 0.759]	0.000	2455
2	0.376	0.176	[0.032, 0.721]	0.033	0.176	[0.031, 0.722]	0.033	2457
3	0.589	0.220	[0.158, 1.021]	0.008	0.219	[0.160, 1.019]	0.007	2454
4	0.677	0.249	[0.190, 1.164]	0.007	0.245	[0.196, 1.158]	0.006	2456
5	0.820	0.314	[0.204, 1.437]	0.009	0.312	[0.208, 1.433]	0.009	2459
6	0.998	0.339	[0.335, 1.662]	0.003	0.336	[0.339, 1.657]	0.003	2461

Note: The degree of the polynomial is indicated by p. All standard errors are heteroskedasticity consistent (HC3). The clustering of standard errors took place on the electoral district–representation order level.

Table 19: Covariate balance tests conducted with uniform kernels.

			Conventional		Robust	
Covariate	h^{opt}	$\hat{\tau}$	s.e.	p-value	s.e.	p-value
Total votes	0.146	-2361.219	2651.616	0.373	3089.096	0.558
Total candidates	0.132	-0.226	0.325	0.486	0.378	0.614
Female candidates	0.075	0.514	0.253	0.042	0.281	0.033
Incumbent candidates	0.148	-0.101	0.092	0.270	0.107	0.255
Total population	0.129	-7853.480	4518.760	0.082	5141.291	0.083
Percentage women	0.090	0.002	0.003	0.496	0.003	0.352
Female labour participation	0.112	-0.003	0.014	0.850	0.016	0.737
Female average income	0.118	-741.927	1944.660	0.703	2279.114	0.597
Unemployed	0.134	-334.038	375.817	0.374	441.017	0.500
Average family size	0.112	0.096	0.058	0.102	0.070	0.161
Total immigrants	0.137	-3333.349	4269.068	0.435	4922.523	0.415
Women with tertiary education	0.183	-2153.200	1322.710	0.104	1488.766	0.089

Note: Covariate balance tests using uniform kernels in combination with the robust bias-corrected methods introduced by (Calonico, Cattaneo, and Titiunik 2014), as calculated by the `R` package `rdrobust`. The MSE-optimal bandwidth is represented by h^{opt}, and the corresponding MSE-optimal point estimates is $\hat{\tau}$. Both the conventional and the robust bias-corrected standard errors are presented. The standard errors are heteroskedasticity consistent (HC3), and they are clustered on the electoral district–representation order level.

Table 20: CER-optimal local linear estimates for the outcome variable: *female challengers*.

Method	$\hat{\tau}$	s.e.	p-value	95% CI
Conventional	**-0.036**	0.248	0.885	[-0.521, 0.450]
RBC	-0.003	**0.270**	**0.991**	**[-0.532, 0.525]**

Note: The CER-optimal bandwidth $h^{opt} = 0.106$. Results of the robust bias-corrected inference methods introduced by (Calonico, Cattaneo, and Titiunik 2014) are calculated using the `R` package `rdrobust`. The standard errors are heteroskedasticity consistent (HC3), and they are clustered on the electoral district–representation order level.

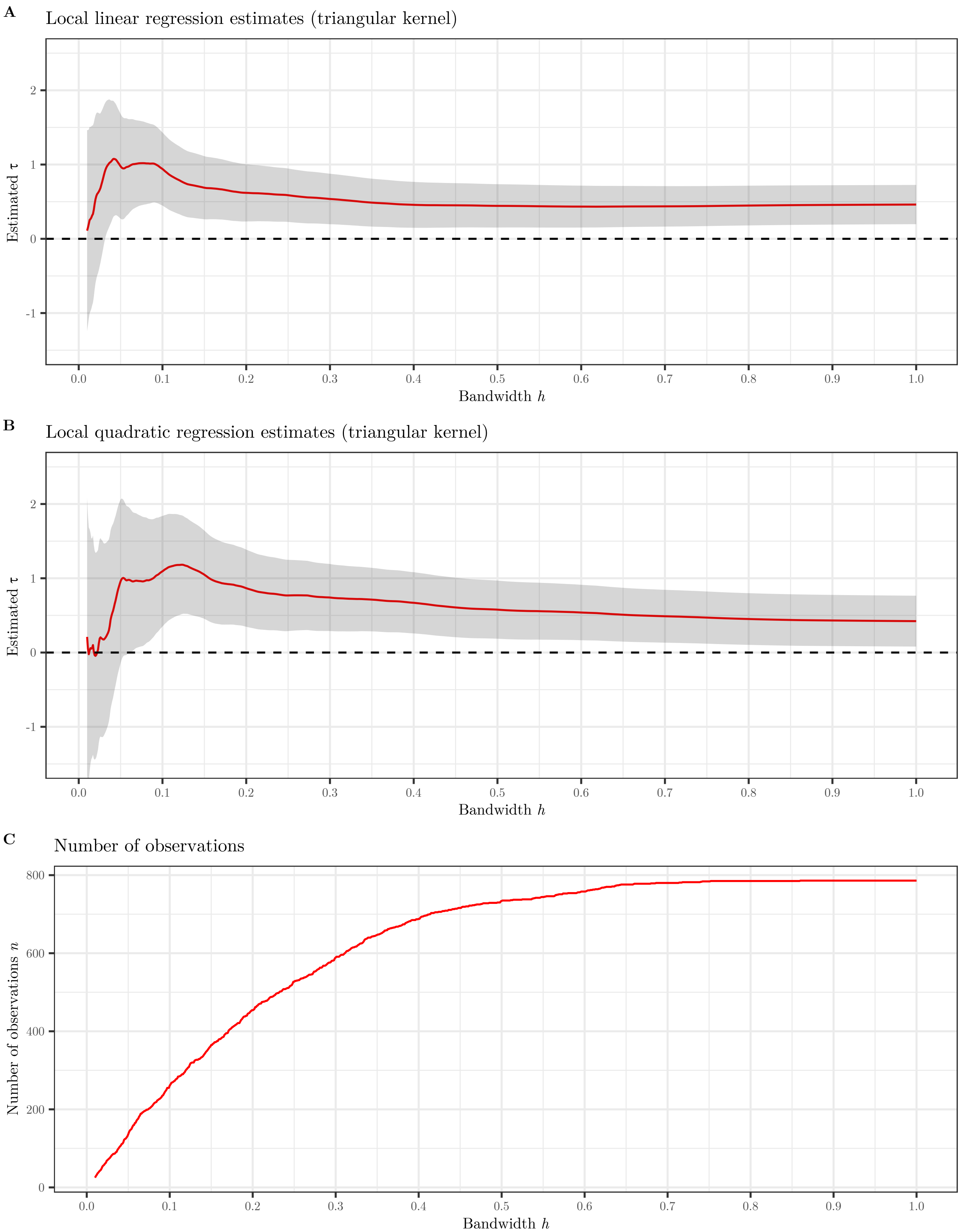

Figure 17: Graphical representation of the bias-variance trade-off of local linear regression techniques. Panel A and B show the estimated treatment effect $\hat{\tau}$ of local linear and quadratic regression models with bandwidths ranging from 0.01 to 1.00. Triangular kernels were used to weigh the observations. The shaded areas represent the 95% confidence interval, calculated using heteroskedasticity consistent (HC3) standards errors clustered on the electoral district–representation order level. Panel C shows the number of observations.

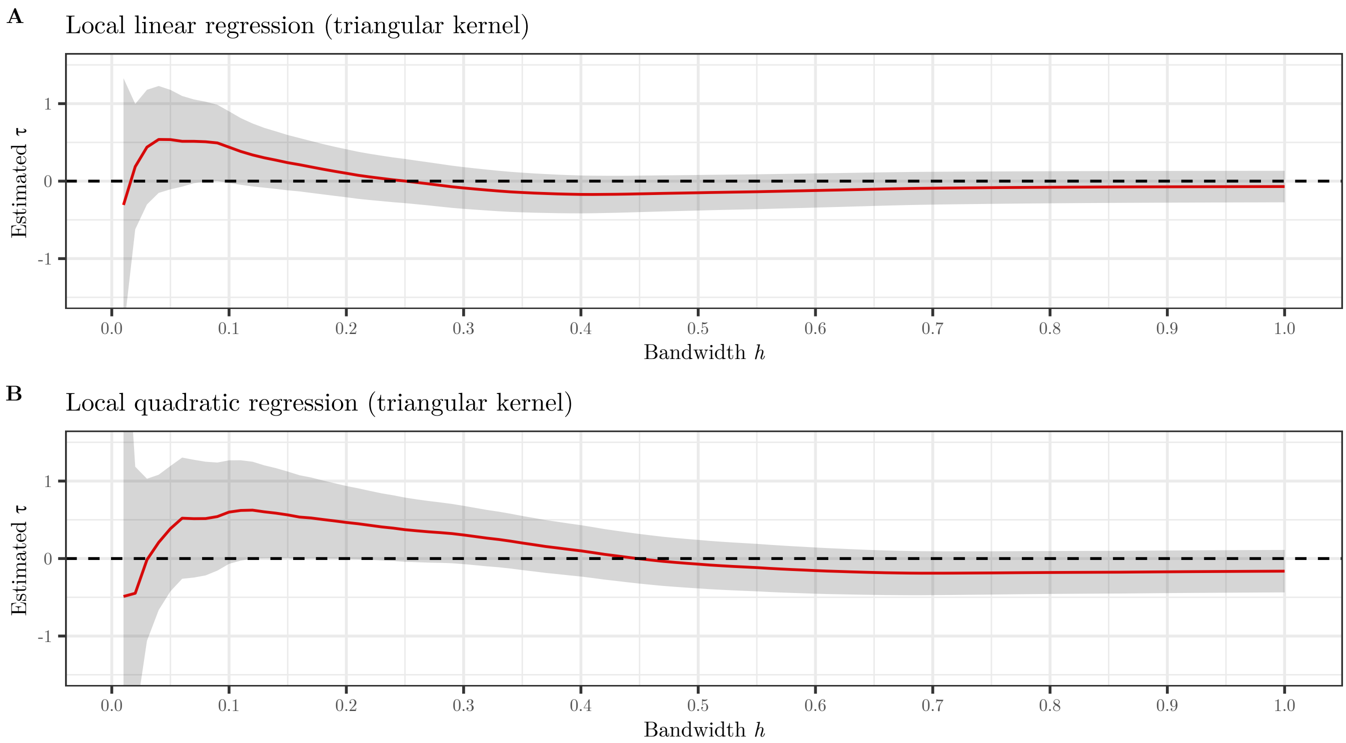

Figure 18: Graphical representation of the sensitivity of the estimated treatment effect for the covariate *Female candidates*. Panel A and B show the estimated treatment effect $\hat{\tau}$ of local linear and quadratic regression models with bandwidths ranging from 0.01 to 1.00. Triangular kernels were used to weigh the observations. The shaded areas represent the 95% confidence interval, calculated using heteroskedasticity consistent (HC3) standards errors clustered on the electoral district–representation order level.

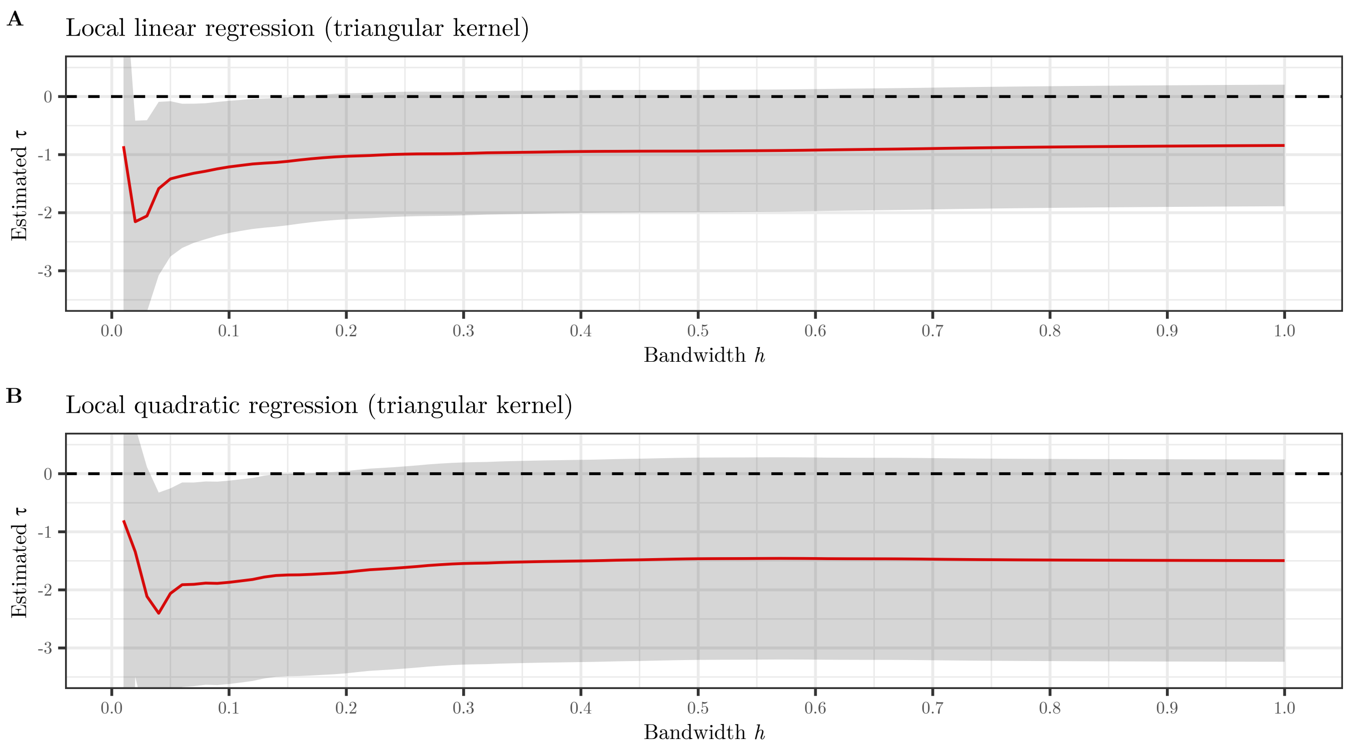

Figure 19: Graphical representation of the sensitivity of the estimated treatment effect for the placebo cut-off point $\tilde{v}_i = 0.05$. Panel A and B show the estimated treatment effect $\hat{\tau}$ of local linear and quadratic regression models with bandwidths ranging from 0.01 to 1.00. Triangular kernels were used to weigh the observations. The shaded areas represent the 95% confidence interval, calculated using heteroskedasticity consistent (HC3) standards errors clustered on the electoral district–representation order level.