Estimating Election Incumbency Bias in General Elections in England

Marwan Riach

CID: 01349928

Imperial College London

Supervised by: Dr. Christopher Hallsworth

June 8 2021

Abstract

In this report, we study the effects of incumbency of the party and of the individual in general elections in England. We begin by investigating omitted variable bias which helps us understand how models which exclude specific covariates can lead to inaccurate methods of prediction. The most common method of estimating personal incumbency known as sophomore surge in US elections is then critiqued and a simple, unbiased method is proposed. For party based incumbency, we use a regression discontinuity design approach combined with kernel density estimation, leaving us with the problem of optimal bandwidth choice. In the final section, we propose a new and hitherto untested method which merges the two types of incumbency bias so as to remove any omitted variable biases that may arise.
Key words: Omitted Variable Bias, Regression Discontinuity Design, Incumbency Advantage, Sophomore Surge, Optimal Bandwidth Selection

Acknowledgements

I would firstly like to thank my supervisor, Dr. Christopher Hallsworth, without whose support and attentiveness the completion of this project would not have been possible. Our weekly meetings throughout the course of the academic year ensured that any intermittent obstacles that appeared in the making of this report never went unnoticed and unaddressed.

Secondly, I would like to recognise Sir John Curtice FRSA from the University of Strathclyde for his contribution to this project by sharing with me a database of UK elections which was to become the source of almost all data used in the models.

Lastly, my gratitude goes to my mother for accommodating me in my final year at Imperial College and through what has been an extraordinary nine months.

Declaration

This is my own work except where otherwise stated.

Contents

1 Introduction

Incumbency advantage has long been a topic of wide discussion in the political community. Its effect is instrumental to pollsters and psephologists alike in the accurate prediction of general elections; when it is known that a candidate is set to run for re-election, the combination of swing in national polls and the incumbency factor can be used to predict elections on a constituency by constituency basis.

The first known study of incumbency effects started with questions surrounding the reelection rate of members of the United States House of Representatives in Erikson's [1] report and developed into a huge field of study in American politics, examining the direction and source of incumbents' electoral perquisites. In 1990, the methods for measuring incumbency advantage were further scrutinised by Gelman and King[2] and it was noted that all pre-existing methods used to compute incumbency advantage were biased. Gelman and King subsequently proposed an unbiased method which went on to show that being an incumbent candidate was a net electoral benefit in the House of Representatives in the first half of the twentieth century.

In the past decade, great strides have been made in the development of more sophisticated ways of measuring incumbency, one of which is the regression discontinuity design (RDD). The method of RDD is explained in this report and has allowed psephologists to measure incumbency of the political party as opposed to incumbency of the individual. Eggers and Spirling [3] use the technology of RDD in order to glean party based incumbency effects in countries where there are multiple political parties on the field, putting emphasis on close contests.

The first contribution of this paper is to investigate the effect of the omission of a variable that correlates both with the response variable and another covariate, in what is known as omitted variable bias. We see this phenomenon in a general setting and then offer a situation in which it could arise in the context of elections. The second contribution of this paper is to study the sophomore surge method and then tailor Gelman and King's unbiased model to a UK setting for the purposes of predicting personal incumbency bias. The third contribution is to implement Eggers and Spirling's method which uses RDD in order to deduce the party based incumbency bias. Finally, the last contribution of this paper is to devise a new method of estimating incumbency bias which combines both personal and party based incumbency. The merging of the two types of incumbency is motivated by the omitted variable bias that is incurred by only considering one of the two types of incumbency bias.

2 Omitted variable bias and its relationship with election modelling

When constructing a model, it is always preferable to abide by the law of parsimony which states that you should use no more variables than necessary: having too many variables often results in over-fitting of the model to the data set in use. In the other direction, using too few variables in a model can produce unreliable models via the phenomenon known as omitted variable bias (OVB).

OVB occurs when a relevant variable is left out and the bias results in the model attributing the effect of the missing variables to those that were included. A simple example of this is predicting a person's salary using only their level of education. The person's ability in the job has not been included in the model and, on average, individuals with more innate ability choose higher levels of education and consequently get paid more. The omission of the ability variable results in the estimate of the effect of education on the person's salary being too large.[4]

To see this in more detail, we will consider a theoretical explanation and then proceed with an example of OVB in simulated election data.

2.1 Theory behind OVB

Illustrating the theory behind OVB, we will consider the following, simple model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i \tag{1}$$

where y_i is dependent on two covariates, x_{1i} and x_{2i}, and an error term, $\epsilon_i \sim N(0, \sigma^2)$ with σ being fixed (a.k.a. homoscedasticity).

Let us consider another model where we ignore the confounding variate x_{2i}:

$$y_i = \beta_0^* + \beta_1^* x_{1i} + \epsilon_i^* \tag{2}$$

Ignoring the confounding variate x_{2i} can be illustrated by the following example:

Suppose that x_{2i} represents whether it is a rainy day or not. A data point from a rainy day is represented by the colour orange and a sunny day by the colour blue. We seek to estimate the number of ice creams sold on a particular day given the number of people visiting the park per day; it is assumed that ice creams are exclusively sold on the park. Figure 1 would suggest that fitting a model to the two clusters of data points would produce more accurate results than merging the two together and ignoring the weather information.

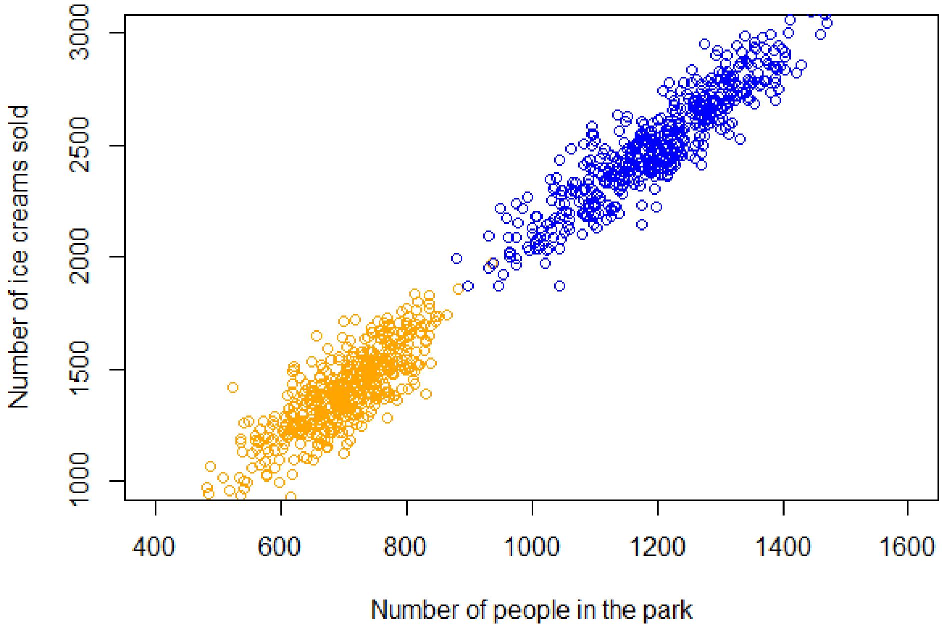


Figure 1: The discrete, confounding covariate is represented by the colour of the dot.

Ascertaining the relationship between models (1) and (2) will allow us to see the OVB. To allow ourselves to do this, let us consider the relationship between x_{1i} and x_{2i} as:

$$x_{2i} = \gamma_0 + \gamma_1 x_{1i} + \nu_i \tag{3}$$

Substitution of (3) into (1) yields:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2(\gamma_0 + \gamma_1 x_{1i} + \nu_i) + \epsilon_i$$

$$y_i = \beta_0 + \beta_2\gamma_0 + (\beta_1 + \beta_2\gamma_1)x_{1i} + \beta_2\nu_i + \epsilon_i \tag{4}$$

Equating coefficients of x_{1i} in (2) and (4) we find that:

$$\beta_1^* = \beta_1 + \beta_2\gamma_1 \tag{5}$$

and hence $\hat{\beta}_1^*$ is unbiased ($E(\hat{\beta}_1^*) = \beta_1$) if, and only if, $\beta_2\gamma_1 = 0$. We have used the fact that $\hat{\beta}_1$ is unbiased as it represents the true relationship and is therefore the ordinary least squares estimator, which we know is unbiased.

This requires that either there is no correlation between y and x_{2i} ($\beta_2 = 0$) or between x_{2i} and x_{1i} ($\gamma_1 = 0$). From this, we can conclude generally that for OVB to occur we require the omitted variable to be correlated with both the response variable and at least one other covariate. The direction of the bias is governed by the values of β_2 and γ_1 in this specific example.

This result is problematic and should be borne in mind when attempting to find an unbiased estimator of our betas, as this violates the strict exogeneity assumption in the classical linear regression model.

2.2 Vote share example

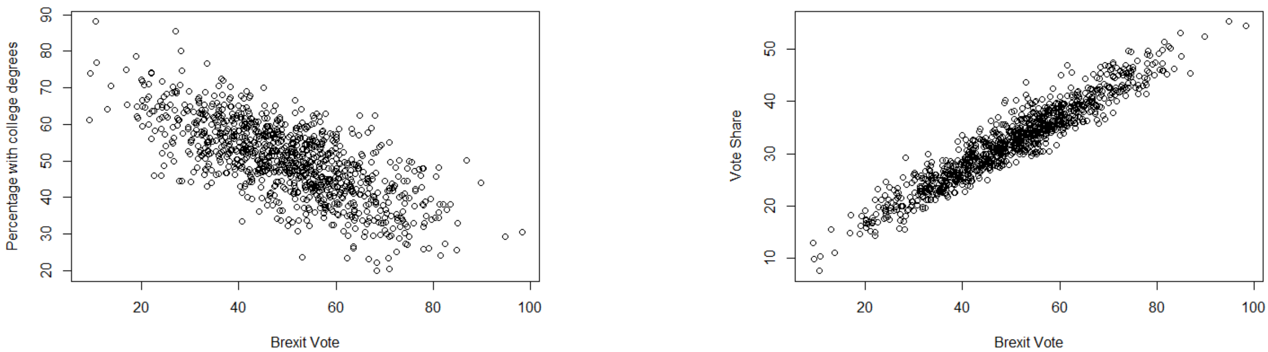


Figure 2: Brexit vote in simulated constituency vs percentage of population with degree (left) and Conservative vote share (right)

For the purposes of a numerical illustration of OVB, we will consider the following model and reduced model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \epsilon_i \tag{6}$$

$$y_i = \beta_0^* + \beta_1^* x_{1i} + \beta_2^* x_{2i} + \beta_3^* x_{3i} + \nu_i \tag{7}$$

where $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 are known and are equal to 20 ,2 ,-0.2 ,0.00007 and 0.4 respectively. y_i represents the vote share for the Conservative party and the coefficients respectively represent the intercept term, incumbency of the candidate (categorical), percentage of voters with college degrees, median income and the vote in the 2016 EU referendum in favour of leave. The response variables are constructed in such a way that they abide by the linear model in (6) with error terms added on in each simulation and we set up the data in such a way that the degree proportion and Brexit vote are purposefully negatively correlated. In the simulation, 1000 data points are used, representing 1000 hypothetical constituencies of varying demographics as encompassed by the four covariates.

In the model shown in (7), we omit the leave vote in the referendum. As can be seen from figure 2, the leave vote is positively correlated with vote share (the response variable) and negatively correlated with the percentage of voters with a university degree (another covariate). According to the theory derived previously, this should incur some OVB.

As we are dealing with simulated data and realisations of random variables, the best way that we can compute the bias of our $\hat{\beta^*}$ values (which relies on knowing the mean) is by appealing to the law of large numbers and performing a large number of simulations (in this case 10000) and taking the average across all simulations to emulate the expected value; the randomness comes from the 'noise' and so this is changed each simulation, with the response values y_i being kept constant.

The results of the simulation are as follows:

β_i^*	$E(\hat{\beta_i^*})$	s.e.
β_0^*	59.7	0.000824
β_1^*	2.02	0.000317
β_2^*	-0.593	$1.54 * 10^{-5}$
β_3^*	0.0000694	$7.71 * 10^{-9}$

As can be seen, $E(\hat{\beta_1^*})$ and $E(\hat{\beta_3^*})$ are virtually unchanged compared to the true values and this is due to the lack of correlation between income and incumbency with the leave vote (as per the set up). The coefficient estimate pertaining to the percentage of voters with college degrees, β_2^*, has its expectation significantly below the true coefficient, indicating a strong, negative bias; this agrees with the aforementioned theory as the bias is the product of a positive correlation coefficient and negative correlation coefficient, resulting in a negative bias overall. The intercept term is inflated and this coefficient is also biased due to the phenomenon that can be seen in (4).

To see how this relates to the result in (5) and verify it practically, we will approximate $E(\hat{\beta_2^*})$ as follows:

$$E(\hat{\beta_2^*}) = \beta_2 + \beta_4\gamma_1 \tag{8}$$

where γ_1 here is the non-intercept coefficient induced by fitting a linear regression with the Brexit vote as the response variable and percentage with college degrees as the regressor. We can omit β_1 and β_3 as their correlation with the Brexit vote is almost zero.

Table 1: Results of linear model with Brexit vote as response and degree proportion as regressor

	Dependent variable:
	Brexit Vote
Prop. of College Degrees	-0.984***
	(0.032)
Intercept	99.339***
	(1.628)
Observations	1,000
R^2	0.491
Adjusted R^2	0.490
Residual Std. Error	10.272 (df = 998)
F Statistic	961.082*** (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 1 shows us that $\gamma_1 = -0.984$, and hence according to (8), the value of $E(\hat{\beta_2^*})$ is:

$$E(\hat{\beta_2^*}) = -0.2 + 0.4 * (-0.984) = -0.594$$

which is indeed extremely close to the the expected value derived from the simulation of -0.593, supporting the theory behind OVB.

The vote shares, y_i, in each constituency will be used to determine the winner of the constituency, which is ultimately what we would want to predict in an election model. The probabilities of winning given the obtained vote share are given in the table below:

Percentage of Vote, y_i	Probability of Winning seat
50%+	1
45-50%	0.9
40-45%	0.7
35-40%	0.5
0-35%	0

This is simulated using a binomial distribution and attempts to replicate the distribution of votes to third parties. Logistic regression is used to model the probability of winning a constituency given the covariates.

This specific set up avoids any issues that may arise from a dataset which is perfectly separated. A complete separation in a logistic regression occurs when the outcome variable separates at least one predictor variable completely. In this instance, the likelihood grows with the size of the coefficient of the predictor variable which separates the outcome variable; this results in the coefficient becoming as large as possible culminating in infinite estimates.

To see this phenomenon, let us consider a covariate, x, and binary response variable, y, which takes its value depending on sign(x). Let us further suppose that the logistic regression model is given simply by:

$$p(x) = \frac{1}{1 + e^{-\beta x}}$$

Figure 3 shows us that as $\beta \to \infty$, $p(x)$ approaches the optimal model, thus solving the MLE estimate.

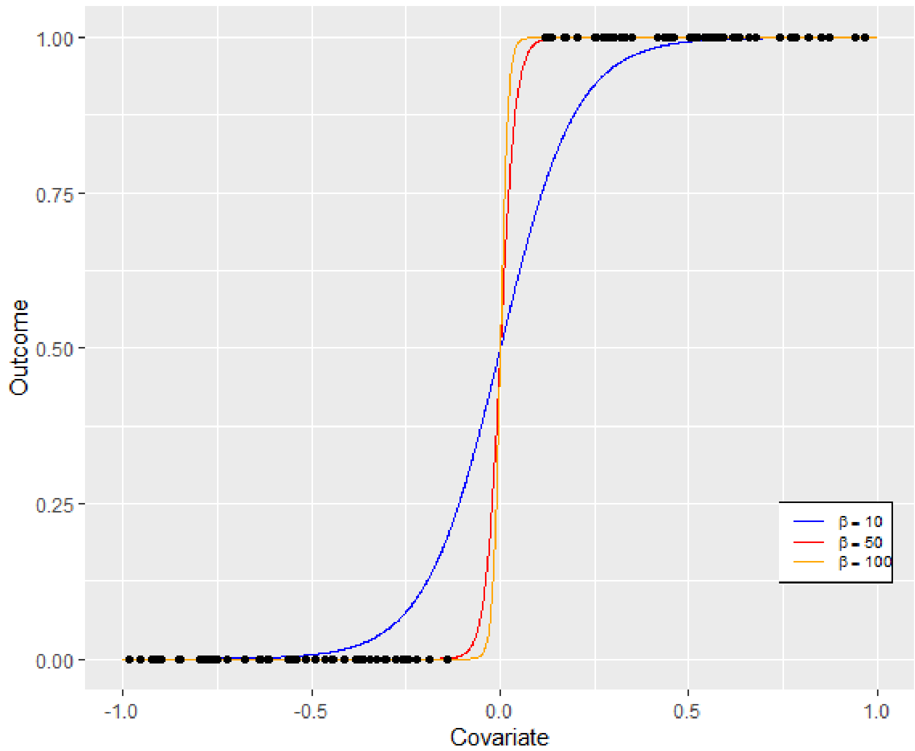


Figure 3: Perfectly separated and simulated points alongside three logistic regression curves to demonstrate that $\beta = \infty$ gives the optimal result.

Table 2: Results of logistic regression on complete data (left) and data with omitted variable (right)

	Dependent variable:	
	Seat Won by Conservative Party	
	(1)	(2)
Incumbency of Candidate	0.389^{*}	0.253
	(0.221)	(0.185)
Prop. of College Degrees	-0.072^{***}	-0.169^{***}
	(0.015)	(0.012)
Median Income	0.00002^{***}	0.00001^{***}
	(0.00000)	(0.00000)
Brexit Vote	0.172^{***}	
	(0.015)	
Intercept	-8.296^{***}	6.268^{***}
	(1.314)	(0.565)
Observations	1,000	1,000
Log Likelihood	-268.063	-374.492
Akaike Inf. Crit.	546.126	756.985
Residual Deviance	536.13	748.98
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

Table 2 shows the results of the two models which fit a logistic regression to the complete data set and data set which omits the Brexit vote. It can be seen that the model that includes the Brexit vote is superior to that which omits it; this is supported by the lower Akaike Information Criterion (AIC).

Another way we can test to see which of the two models is superior is to consider the deviance of both models. Let us call the model which omits the Brexit vote model A and the other which includes it model B. Here, we have two nested models. Suppose that model A is the true model. Then, by Wilks' theorem and under the null hypothesis, the difference between the two models' deviances follows an approximate chi-squared distribution with 1 degree of freedom. We therefore consider the following hypothesis:

$$H_0 : \beta_4 = 0 \text{ (i.e. model A is the true model)}$$

$$H_1 : \beta_4 \neq 0 \text{ (i.e. model B is the true model)}$$

Under the null hypothesis:

$$D = Deviance_A - Deviance_B$$

$$= 749 - 536 = 213$$

comes from a χ_1^2 distribution.

The p-value associated with associated with this statistic is well below 0.001 (as 0.001 is associated with 10.8) and hence we have significant evidence to reject the null hypothesis and deduce that model B is superior.

One feature of hypothesis tests that is of high importance is its power: the power of a hypothesis test is the probability that the test correctly rejects the null hypothesis (H_0) when a specific alternative hypothesis (H_1) is true. The above test is optimal by the Neyman-Pearson lemma which states that a likelihood ratio test has the highest power for any hypothesis test conducted with a predetermined type 1 error (also known as the false positive rate).

We perform a simulation of the above setting over different sample sizes, N, and fix the false positive rate $\alpha = 0.05$ in order to ascertain the power of the conducted hypothesis test.

The power of the chi-square test is formally given by:

$$1 - \beta = F_{k,\lambda}(x_{crit})$$

where F is the cumulative distribution function for the non-central chi-square distribution with k degrees of freedom and non-centrality parameter λ and x_{crit} is the critical value of a central chi-square distribution for a given value of α with k degrees of freedom.

The non-centrality parameter is given as $\lambda = w^2 N$, where w is the chi-square effect size. In the context of nested models, Cohen's effect size measure is widely used[6] and is defined as follows:

$$w = \sqrt{\frac{\Delta D}{N \cdot \Delta df}}$$

where ΔD and Δdf are the difference in the deviances and degrees of freedom of the two models respectively.

In our simulation, $\Delta df = k = 1$ as we are only working with the omission of one variable. Figure 4 shows us that for $N > 200$, the power of the hypothesis test is almost surely 1. Hence in our hypothesis test where $N = 1000$, the test is extremely powerful, buttressing the superiority of model B over model A.

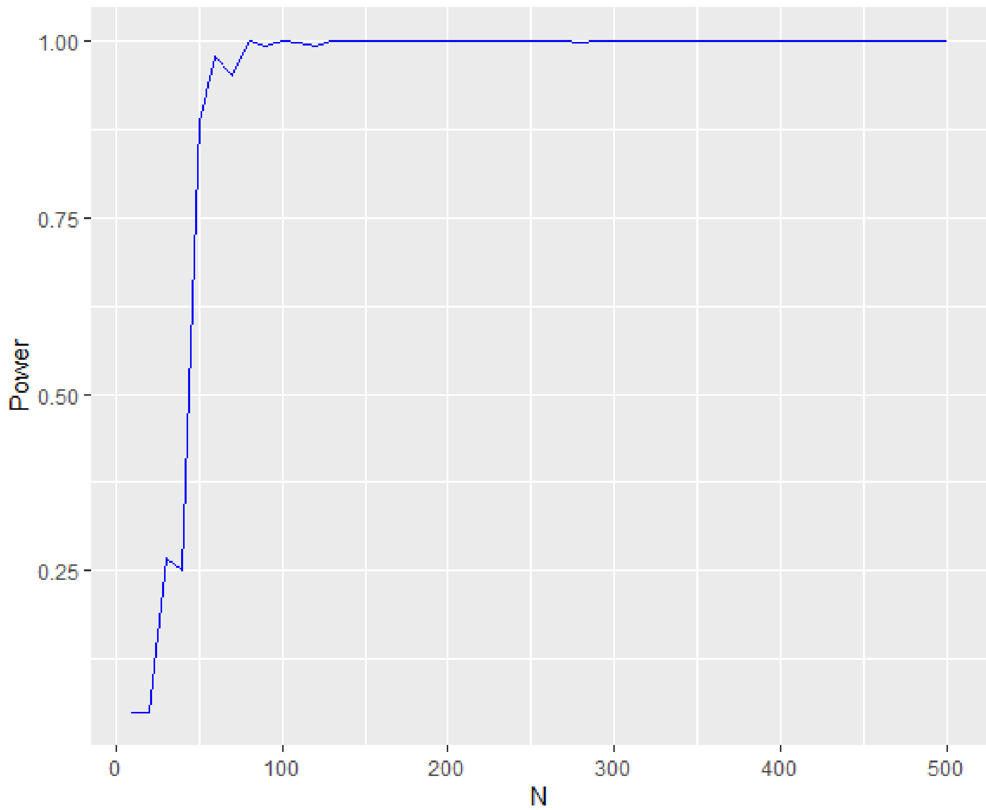


Figure 4: Simulation showing the power of the χ_1^2 test for the nested logistic regression models for different values of N

3 Estimating Incumbency Advantage by Individual

One phenomenon that is extensively discussed in studies of elections in the USA is personal incumbency bias, specifically in elections to the House of Representatives. This topic is scarcely covered in the UK and so this section seeks to estimate the effects of personal incumbency in UK elections from 2001, emulating and adapting the methodologies derived in American papers. We begin by discussing biased methods of measuring personal incumbency bias and studying their associated problems. We then implement an unbiased method from Gelman and King, adapted for UK elections.

3.1 The Sophomore Surge Method and its Associated Problems

Gelman and King[2] note that the most common method used to measure incumbency bias is the sophomore surge. This phenomenon is measured by computing the average increase in vote share for a candidate that is running for a second term; it is assumed that this is a fixed value, irrespective of which party holds the status of incumbency. We will demonstrate that this method is biased both theoretically and via hypothetical example. For the purposes of simplification, we will assume a two party system whereby two, and only two, candidates from the two main parties (which we will unimaginatively call party A and party B) stand in every constituency.

Prior to doing this, we will define the notation used:

- v_t^I is the vote share obtained by party A at time t ($t = 1, 2$) with an incumbent running. When the constituency is open, this is unobserved.

- v_t^O is the vote share obtained by party A at time t when the seat is open. When an incumbent is running, this is unobserved.

- I_t is equal to 1 if party A incumbent runs for reelection, 0 if no incumbent runs (i.e. open seat), and -1 if party B incumbent is seeking reelection.

- Ψ_t is the incumbency advantage granted to a candidate in election t. This is measured in percentage points.

- $f(v_t^O)$ is the probability density function of the vote share at time t in an open race, that generates v_t^O in election t.

For the purposes of simplicity, we will also assume that from one election to the next, the nationwide partisan swing will be δ in favour of party A from $t = 1$ to $t = 2$. Additional assumptions made here are that the incumbency effect across the pair of elections is constant and that the marginal density, f, is symmetric between elections 1 and 2, allowing for the swing of δ. The last assumption is equivalent to the following being true:

$$f(v_1, v_2 + \delta) = f(v_2, v_1 + \delta).$$

With the notation and assumptions laid out, we can now proceed to show the bias of the sophomore surge. Sophomore surge for party A is equal to the proportion of the vote a candidate for party A receives running for reelection at time 2 subtract the proportion they received in their first election at time 1 with necessary adjustments for swing being added in. Were the method to be unbiased, we would expect that $E(SS_A) = \Psi_2 + \delta$. Let us consider the true expectation of SS_A to see if any bias exists:

$$E(SS_A) = E(v_2^I - v_1^O | I_1 = 0, I_2 = 1) = E(v_2^O + \Psi_2 - v_1^O | I_1 = 0, I_2 = 1)$$

Conditioning on the incumbent winning the first open race and running as an incumbent at time $t = 2$ is equivalent to the candidate at time $t = 1$ obtaining more than 50% of the vote:

$$E(SS_A) = E(v_2^O + \Psi_2 - v_1^O | v_1^O > 0.5)$$
$$= \Psi_2 + \frac{1}{P(v_1 > 0.5)} \int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} (v_2 - v_1) f(v_1, v_2) dv_2$$

(Using the Kolmogorov definition of conditional probability)

The limits of our integrals should in theory go from $-\infty$ to ∞ (except for the lower bound of v_1). In practice, our limits of integration are determined by the support of the distribution. We know that the joint pdf is zero outside the range [0,1], as both v_1 and v_2 represent vote shares.

Introducing the change of variables $v_2 = u_2 + \delta$:

$$
\begin{aligned}
E(SS_A) &= \Psi_2 + \delta + \frac{1}{P(v_1 > 0.5)} \int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} (u_2 - v_1) f(v_1, u_2 + \delta) du_2 \\
&= \Psi_2 + \delta + \frac{1}{P(v_1 > 0.5)} \left(\int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} u_2 f(v_1, u_2 + \delta) du_2 - \int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} v_1 f(v_1, u_2 + \delta) du_2 \right) \\
&= \Psi_2 + \delta + \frac{1}{P(v_1 > 0.5)} \left(\int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} u_2 f(v_1, u_2 + \delta) du_2 - \int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} v_1 f(u_2, v_1 + \delta) du_2 \right)
\end{aligned}
$$

where we used the fact that the pdf f is symmetric about the partisan swing. The second term in the brackets can have its labels switched due to the symmetry of f which results in:

$$
\begin{aligned}
E(SS_A) &= \Psi_2 + \delta + \frac{1}{P(v_1 > 0.5)} \left(\int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} u_2 f(v_1, u_2 + \delta) du_2 - \int_{-\infty}^{\infty} dv_1 \int_{0.5}^{\infty} u_2 f(v_1, u_2 + \delta) du_2 \right) \\
&= \Psi_2 + \delta + \frac{1}{P(v_1 > 0.5)} \left(\int_{0.5}^{\infty} dv_1 \int_{-\infty}^{0.5} u_2 f(v_1, u_2 + \delta) du_2 - \int_{-\infty}^{0.5} dv_1 \int_{0.5}^{\infty} u_2 f(v_1, u_2 + \delta) du_2 \right)
\end{aligned}
$$

where the last line is obtained by elimination of common regions in the plane.

Given the limits with respect to u_2, replacing u_2 with 0.5 maximises the first integral in the brackets and minimises the second integral in the brackets. This therefore optimises the value of their difference and allows us to establish an upper bound:

$$
E(SS_A) \leq \Psi_2 + \delta + \frac{0.5}{P(v_1 > 0.5)} \left(\int_{0.5}^{\infty} dv_1 \int_{-\infty}^{0.5} f(v_1, u_2 + \delta) du_2 - \int_{-\infty}^{0.5} dv_1 \int_{0.5}^{\infty} f(v_1, u_2 + \delta) du_2 \right) = \Psi_2 + \delta
$$

where we realise that the value in brackets is equal to zero owing to the equality of the two double integrals; this directly follows from the symmetry of of f about the partisan swing. Thus unless $E(v_2^O - v_1^O | v_1^O > 0.5) = 0$, the sophomore surge method will always incur a negative bias. ■

To see this phenomenon in practice, we will work with simulated, hypothetical election data.

Let us set $\Psi_2 = 0.05$, $\delta = 0$, and fix the random variables in time 1 and time 2 as being independent (this implies $u_2 = v_2$). This then means that the marginal joint pdf factors into the product of the two individual pdfs of V_1 and V_2 i.e.:

$$
f_{V_1, V_2}(v_1, v_2) = f_{V_1}(v_1) f_{V_2}(v_2)
$$

Let us further take into account the assumption of symmetry for the marginal density:

$$f_{V_1,V_2}(v_1,v_2) = f_{V_1,V_2}(v_2,v_1) = f_{V_1}(v_2)f_{V_2}(v_1) = f_{V_1}(v_1)f_{V_2}(v_2)\ \forall v_1, v_2$$

We now have conditions that we must satisfy for our simulations. Note that a necessary condition for the symmetry of the marginal density above is that the supports of both pdfs are equal. The support is defined as:

$$\text{supp}(f_X) = \{x \in X \mid f_X(x) \neq 0\} = R_X$$

To see that the two supports of f_{V_1} and f_{V_2} must be equal, let us suppose that they are not. This means that $\exists v_1 \in R_{V_1}$ s.t. $v_1 \notin R_{V_2}$ or $\exists v_2 \in R_{V_2}$ s.t. $v_2 \notin R_{V_1}$.

Let us fix $v_2 \in R_{V_2}$ and suppose that $\exists v_1 \in R_{V_1}$ s.t. $v_1 \notin R_{V_2}$. This then implies that:

$$f_{V_1}(v_1)f_{V_2}(v_2) > 0$$
$$f_{V_1}(v_2)f_{V_2}(v_1) = f_{V_1}(v_2) \cdot 0 = 0$$
$$\Rightarrow f_{V_1}(v_1)f_{V_2}(v_2) \neq f_{V_1}(v_2)f_{V_2}(v_1)\perp$$

Therefore, we conclude that $R_{V_1} \subseteq R_{V_2}$.

Similarly, let us fix $v_1 \in R_{V_1}$ and suppose that $\exists v_2 \in R_{V_2}$ s.t. $v_2 \notin R_{V_1}$. This then implies that:

$$f_{V_1}(v_1)f_{V_2}(v_2) > 0$$
$$f_{V_1}(v_2)f_{V_2}(v_1) = 0 \cdot f_{V_2}(v_1) = 0$$
$$\Rightarrow f_{V_1}(v_1)f_{V_2}(v_2) \neq f_{V_1}(v_2)f_{V_2}(v_1)\perp$$

Therefore, we conclude that $R_{V_2} \subseteq R_{V_1}$.

Thus, $R_{V_1} = R_{V_2}$ is a necessary condition. ■

For our first simulation, we will let V_1 and V_2 be uniformly distributed on [0.25,0.75]. Theoretically, to see the bias that is incurred by this choice of distribution, let us consider $E(SS_A) - \Psi_2$:

$$
\begin{aligned}
&E(SS_A) - \Psi_2 \\
&= \frac{1}{P(v_1 > 0.5)} \int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} (v_2 - v_1) f(v_1, v_2) dv_2 \\
&= \frac{1}{0.5} \int_{0.5}^{0.75} dv_1 \int_{0.25}^{0.75} (v_2 - v_1) \cdot 2 \cdot 2 \cdot dv_2 \\
&= 8 \int_{0.5}^{0.75} dv_1 \int_{0.25}^{0.75} (v_2 - v_1) dv_2 \\
&= 8 \int_{0.5}^{0.75} \left[\frac{v_2^2}{2} - v_2 v_1 \right]_{0.25}^{0.75} dv_1 \\
&= 8 \int_{0.5}^{0.75} \frac{1}{4} - \frac{v_1}{2} dv_1 \\
&= 8 \left[\frac{v_1}{4} - \frac{v_1^2}{4} \right]_{0.5}^{0.75} \\
&= -\frac{8}{64} = -\frac{1}{8}
\end{aligned}
$$

This shows a negative bias. We illustrate this theory by simulation.

Using the aforementioned uniform distribution, we generate 1,000,000 hypothetical results from time 1 to time 2 and then retrospectively add an incumbency advantage of 5% to the result at time 2 conditional on the result at time 1 exceeding 50%. For the purposes of replicability of the simulation we set a seed. Figure 5 shows us the result of the simulation, supporting our findings in theory that the sophomore surge has a negative bias in its estimation of incumbency advantage. The large sample size is intended so as to make use of the law of large numbers.

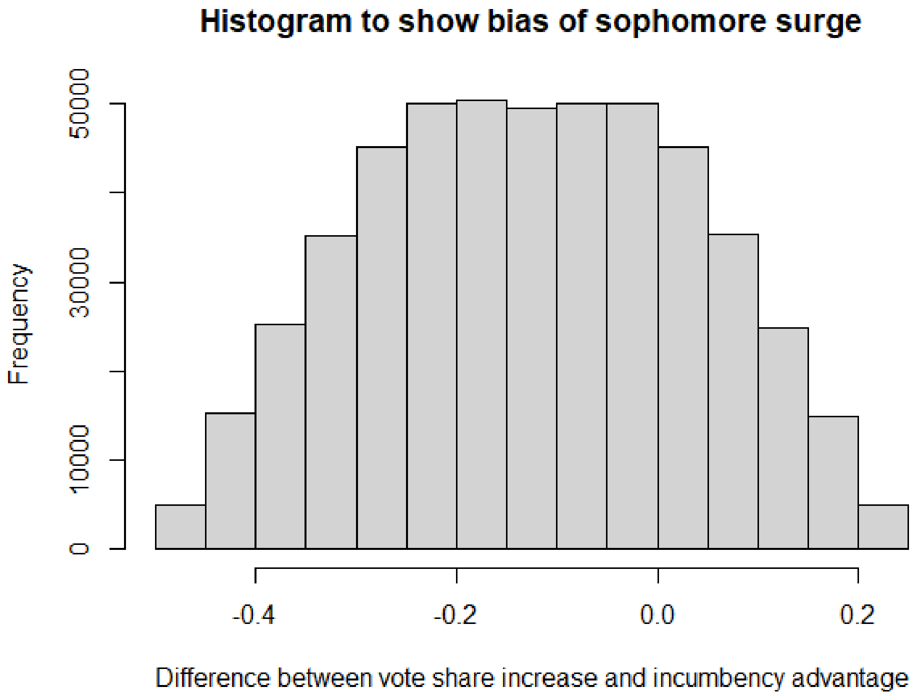


Figure 5: Result of simulation of sophomore surge using simulated data points demonstrated using a histogram. The mean of these simulated data points is -0.125 (to 3 s.f.), complementing the theoretical result of $-\frac{1}{8}$. The variance is 0.0261.

To see a more general result, we will see how the bias varies as the parameters of a uniform distribution are varied in our second simulation. Let V_1 and V_2 be i.i.d. $\sim U[\alpha, \beta]$ where we vary the two parameters with $\alpha \in [0.2, 0.45]$ and $\beta \in [0.55, 0.8]$.

To see the theoretical, incurred bias we will consider $E(SS_A) - \Psi_2$ in a similar way to as we did in the first example:

$$
\begin{aligned}
E(SS_A) - \Psi_2 &= \frac{1}{P(v_1 > 0.5)} \int_{0.5}^{\infty} dv_1 \int_{-\infty}^{\infty} (v_2 - v_1) f(v_1, v_2) dv_2 \\
&= \frac{\beta - \alpha}{\beta - \frac{1}{2}} \int_{0.5}^{\beta} dv_1 \int_{\alpha}^{\beta} (v_2 - v_1) \cdot \frac{1}{(\beta - \alpha)^2} dv_2 \\
&= \frac{1}{(\beta - \frac{1}{2})(\beta - \alpha)} \int_{0.5}^{\beta} dv_1 \int_{\alpha}^{\beta} (v_2 - v_1) dv_2 \\
&= \frac{1}{(\beta - \frac{1}{2})(\beta - \alpha)} \int_{0.5}^{\beta} \left[\frac{v_2^2}{2} - v_2 v_1 \right]_{\alpha}^{\beta} dv_1 \\
&= \frac{1}{(\beta - \frac{1}{2})(\beta - \alpha)} \int_{0.5}^{\beta} \frac{(\beta - \alpha)(\beta + \alpha)}{2} - (\beta - \alpha) v_1 dv_1 \\
&= \frac{1}{(\beta - \frac{1}{2})} \int_{0.5}^{\beta} \frac{(\beta + \alpha)}{2} - v_1 dv_1 \\
&= \frac{1}{(\beta - \frac{1}{2})} \left(\frac{\alpha\beta}{2} - \frac{(\alpha + \beta)}{4} + \frac{1}{8} \right) \\
&= \frac{\alpha}{2} - \frac{1}{4}
\end{aligned}
$$

Hence, we can see that the bias is negative, provided that the lower parameter in the uniform distribution does not exceed

0.5; this makes sense as we require there to be a possibility that the candidate at time 1 loses, which is possible if and only if $\alpha < 0.5$. The bias is also independent of our choice of β.

For each combination of α and β (using increments of 0.01) in the aforementioned ranges, we generate 10,000 results at time 1 and time 2 and then retrospectively add an incumbency advantage of 5% to the result at time 2 conditional on the result at time 1 exceeding 50%. The mean of SS_A is computed in order to calculate its empirical expected value and 0.05 is subtracted from this in order to compute the bias. Figure 6 shows us the results of the simulation.

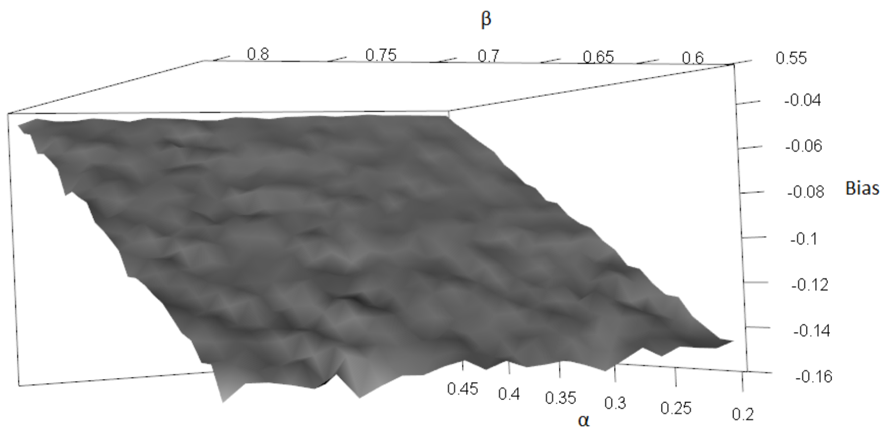


Figure 6: Simulation to show the bias of the sophomore surge method across parameters of α and β for a uniform distribution $U[\alpha, \beta]$. The bias is shown on the vertical axis and the plane perpendicular to this axis represents the different α and β values.

The biases of all points on the surface are well below 0, supporting the theory that the bias is negative. Furthermore, the surface only appears to change in the α direction, validating the fact that the bias is independent of our choice of β.

3.2 Unbiased Regression Based Method

Linear regression methods estimate incumbency advantage by comparing constituencies with incumbents running to open constituencies. This measures the incumbency advantage due to the individual running as opposed to the party. Gelman and King[2] conclude from their report that the following regression framework creates unbiased estimates for personal incumbency:

$$v_{i,t} = \beta_0 + \beta_1 v_{i,t-1} + \beta_2 P_{i,t} + \psi I_{i,t} + \epsilon_{i,t} \tag{9}$$

where $v_{i,t}$ is the vote share at time t in constituency i, $P_{i,t}$ represents the party that wins the election at time t in constituency i, and $I_{i,t}$ represents the winning status of the incumbent candidate (which is set to zero in an open race, 1 if the Conservative is the incumbent and -1 if Labour is the incumbent). The estimate of incumbency advantage is given by ψ. The vote share at time t is also adjusted in accordance with the swing from time $t-1$ to time t. This is to ensure that a surge in vote (or collapse) is not wrongly attributed to an incumbency effect. For the purposes of simplicity, we assume that the national swing is uniformly applied in each constituency. Figures 7 and 8 show that this assumption is potentially problematic, suggesting further scope for improvement (discussed in the conclusion).

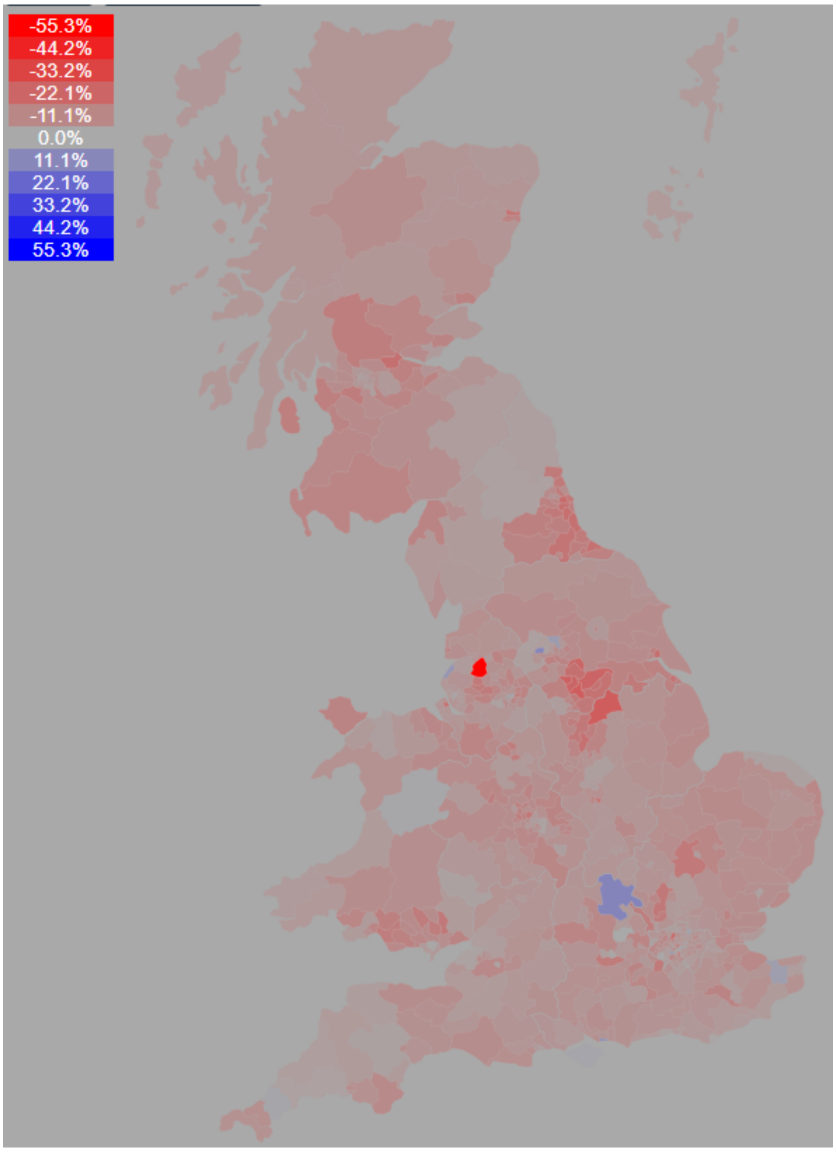


Figure 7: Vote share change in each constituency from 2017 to 2019 for the Labour party[5]

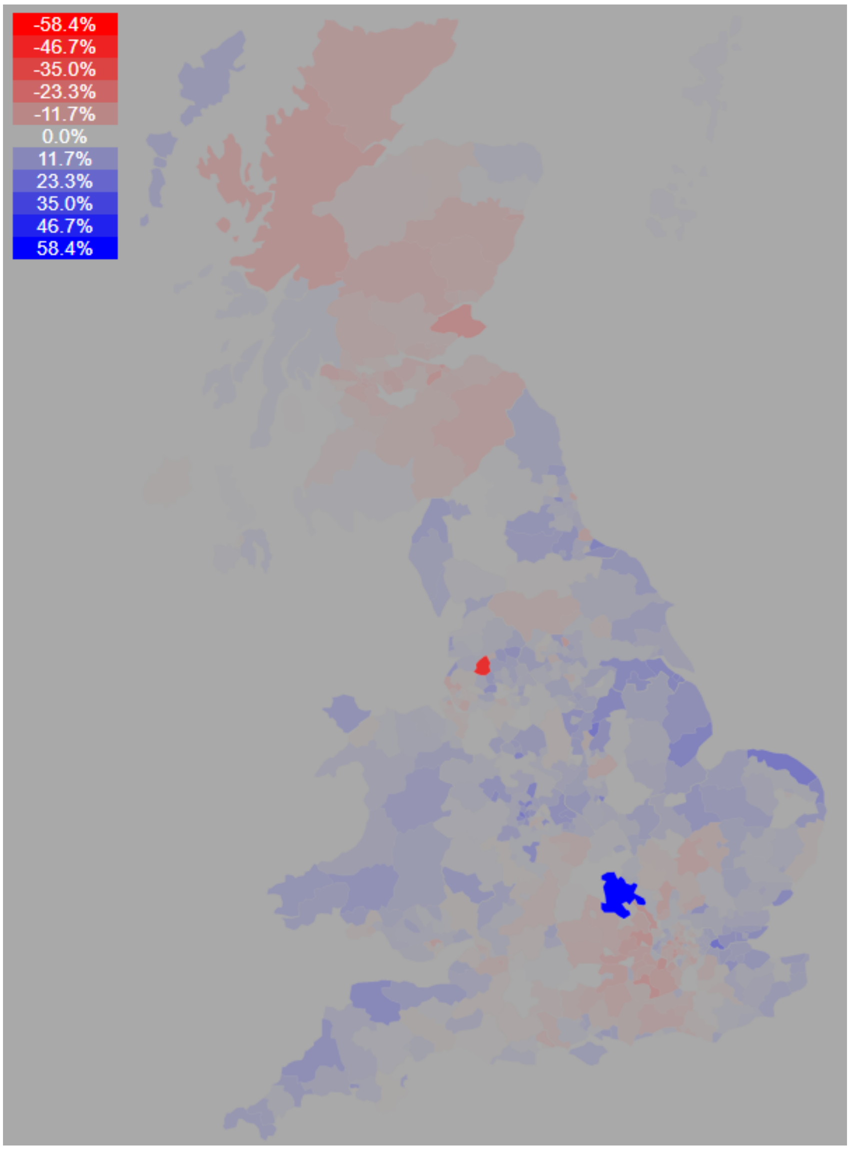


Figure 8: Vote share change in each constituency from 2017 to 2019 for the Conservative party[5]

In Gelman and King, the focus is on elections to congress with the parties forming a dichotomy of Republicans and Democrats. In the UK, the party system is more mutli-coloured, and so to transpose the method to the UK, it would seem sensible to consider only constituencies which are dominated by the Labour party and the Conservative party. This will be accounted for as follows: seats in which the Conservative party and Labour party have both taken up first and second place in every election since 1997 are seats that will be considered; there are 179 of these. The vote share that we will consider in this case is the Conservative vote share.

It is therefore important to note that any conclusion on incumbency bias in England for model (9) will focus specifically on seats in which both the Labour party and Conservative party dominate. This method is subject to some selection bias, however if the mean vote share in the chosen constituencies is compared to the mean vote share of both major parties in all elections from 1997, there is little change with the mean vote in selected seats being marginally lower but not significantly enough to be cause for concern (figure 9). Furthermore, the two main parties dominate in terms of parliamentary seats and so the seats which are being considered constitute the main battlegrounds in any general election. The issue of selection bias is discussed in Gelman and King, however this chiefly concerns the omission of uncontested seats; fortunately, this issue is never encountered in UK elections (except for in the rare case of election of the Speaker of the House which only accounts for 1 seat).

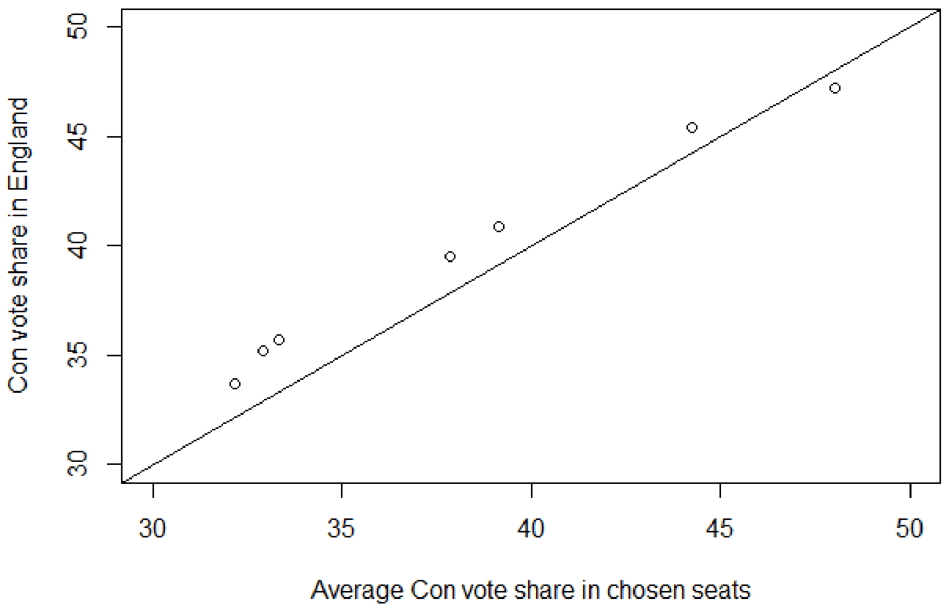


Figure 9: Conservative vote share across England vs average Conservative vote share in 179 selected constituencies[7]

3.3 Results

Fitting the model to elections in the UK between 2001 and 2019 inclusive produces the estimates given in figure 10. What can be seen is that all coefficients that are significant at the 95% level (2010, 2015, 2019) produce positive incumbency bias, with the bias being most pronounced in 2019. However, three of the coefficients are statistically insignificant and furthermore, the 95% confidence intervals are extremely wide (especially in 2001 and 2017). This suggests scope for improvement and that an alternative measure of incumbency should be considered in order to incorporate more than the 179 constituencies originally used.

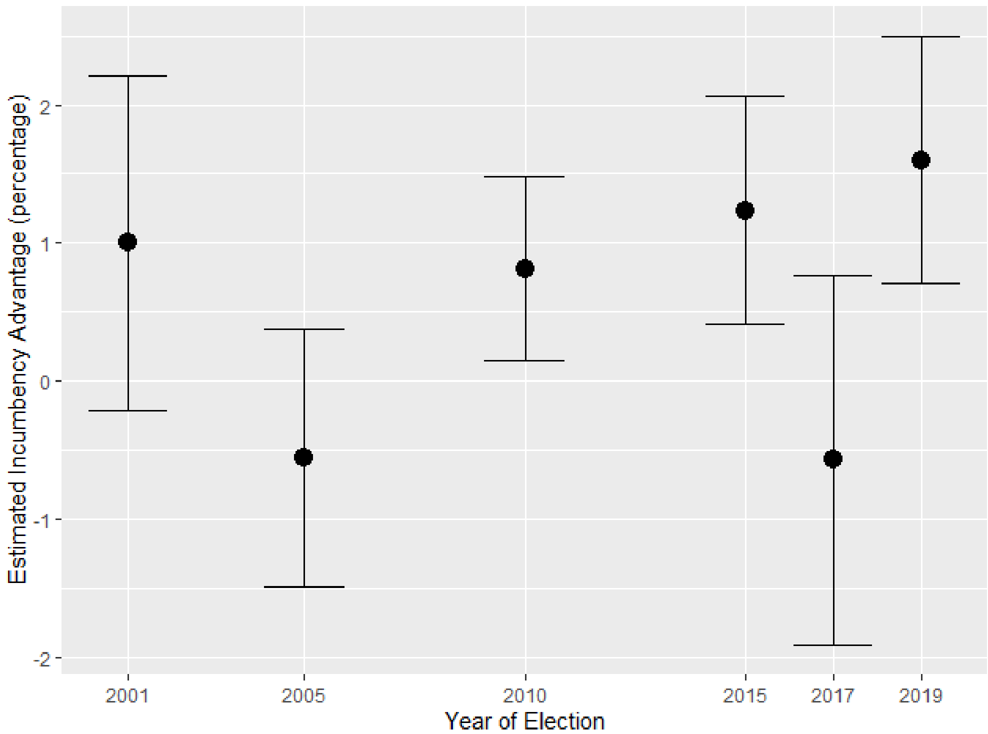
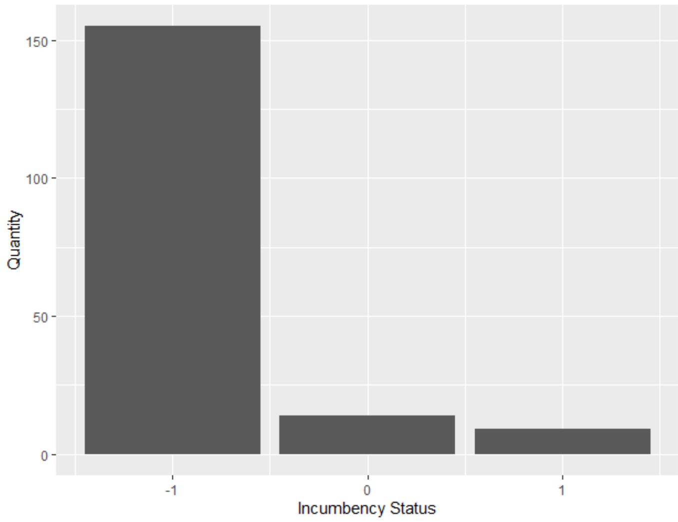
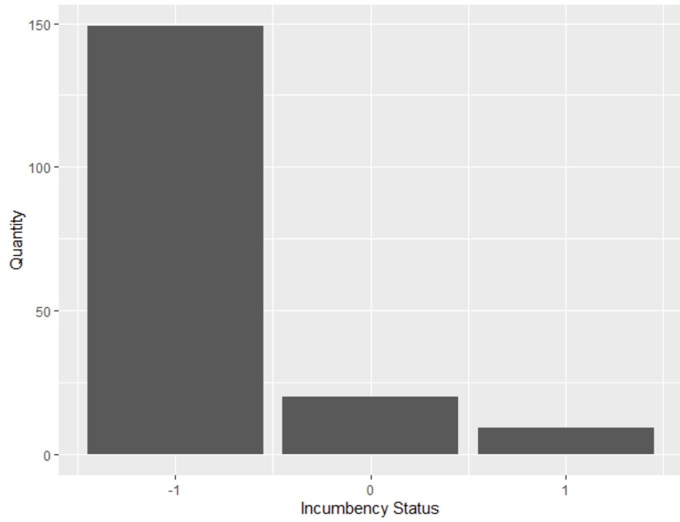


Figure 10: Estimates of incumbency advantage in each election from 2001-2019 with 95% confidence intervals

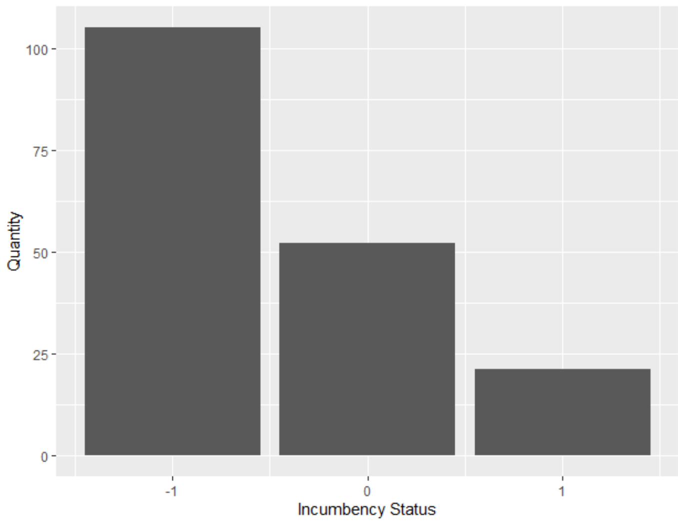
Figure 11 shows the distribution of incumbency status for each general election. 11c, 11d and 11f produce the most balanced data for incumbency status, supporting the fact that these also correspond to the narrowest confidence intervals. 2017 is particularly problematic as there is hardly any data which measures the effect of no incumbency, leading to the extremely wide confidence interval as can be seen in figure 10 caused by the large variance. Reducing the variance and width of the confidence intervals can be achieved by dealing with a larger sample (using as many constituencies as possible).



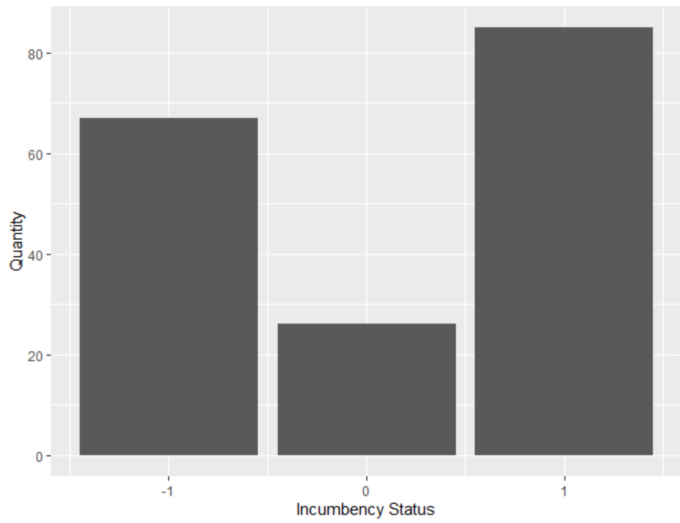
(a) 2001



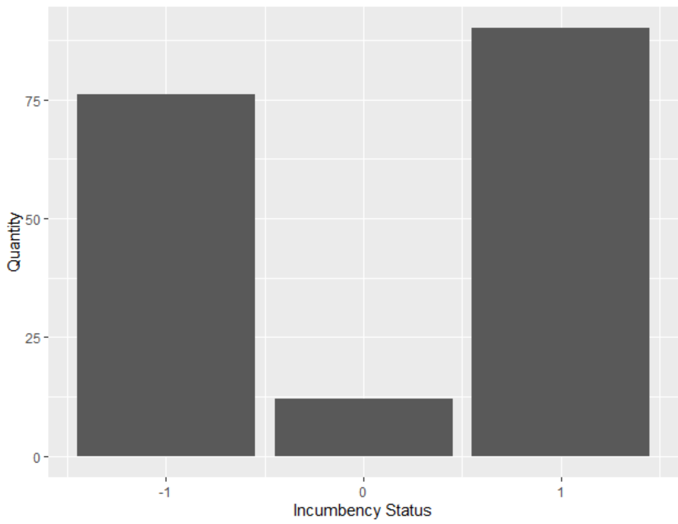
(b) 2005



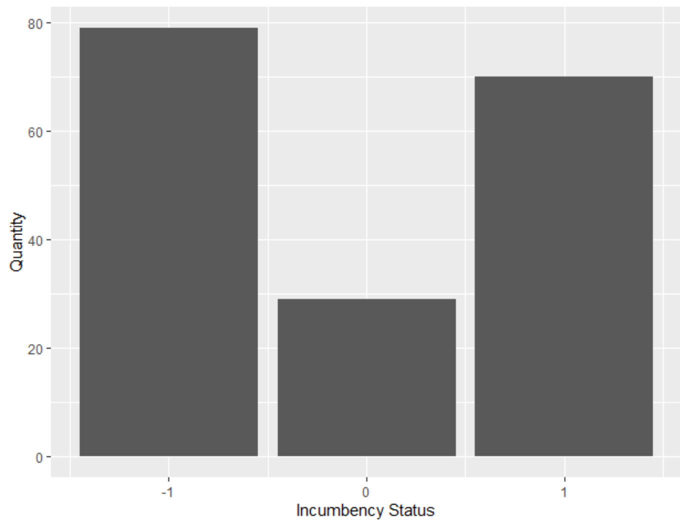
(c) 2010



(d) 2015



(e) 2017



(f) 2019

Figure 11: Distribution of incumbency statuses for all elections since 2001

4 Estimating Incumbency Advantage by Political Party

To incorporate as many constituencies as possible, we will now consider an alternative method. This time, we will investigate the incumbency effect due to a party seeking re-election in a constituency as opposed to an individual. For now, we will consider only party incumbency and omit information regarding individuals seeking re-election and discuss the potential drawbacks of this later on.

The set-up of the model will be such that the Conservative vote share will depend upon its margin of victory (which is negative if the seat was won by another party), whether it won (binary variable 0,1) and an interaction term to allow for regression discontinuity design (RDD). RDD is a model design that elicits the casual effects of interventions by assigning a cutoff above or below which an intervention is assigned. By comparing observations that lie closely on either side of the threshold, it is possible to estimate the average treatment effect in environments in which randomisation is unfeasible.

4.1 Weighted Least Squares Model

The model is as follows and was proposed by Eggers and Spirling[3]:

$$Y_{it} = \beta_0 + \beta_1 ConMargin_{i,t-1} + \beta_2 ConWon_{i,t-1} + \beta_3 (ConWon_{i,t-1} \times ConMargin_{i,t-1}) + \epsilon_{it} \tag{10}$$

where the interaction term represents the discontinuity depending on whether the Conservative party is incumbent or not. In this context, the incumbency effect is determined by β_2 (i.e. the jump in the discontinuity) since it represents the effect of incumbency in the event of a tied election at time $t-1$. The Conservative party's opponent also experiences a party based incumbency advantage if it won the previous election. Therefore, β_2 strictly measures the sum of the party based incumbency for both the Conservative party and the Conservative party's opponent; we accept this loss of information and unidentifiability in exchange for being able to use all constituencies bar two, allowing us to draw more accurate conclusions. This metric still carries relevant information as it measures the relative effect of a seat that is Conservative going into an election versus one that is not.

Suppose we wanted to find the specific party based incumbency for the Conservative party and all its opponents. In a field where there are only three parties (Conservative, Labour, and Liberal Democrat), we could compute the three β_2 estimates associated with Conservative-Labour, Labour-Liberal Democrat, and Conservative-Liberal Democrat races. Denoting I_C, I_L and I_{LD} as the specific party based incumbencies, we would solve the following system of equations:

$$\hat{\beta}_{2,C,L} = I_C + I_L$$
$$\hat{\beta}_{2,L,LD} = I_L + I_{LD}$$
$$\hat{\beta}_{2,C,LD} = I_C + I_{LD}$$

where for $\hat{\beta}_{2,X,Y}$, we would only use constituencies in which parties X and Y were in the top two at time $t-1$. With our data, the confidence intervals of our β_2 estimates would be so wide as to be meaningless due to the scarcity of seats in which the Liberal Democrats came first or second in recent elections.

Again, we assume that the swing between elections is uniformly applied, however we do not have to adjust our vote share for the swing, since the jump in the discontinuity would be of the same size, irrespective of a uniform shift of all votes at time t.

We will propose that the effects of incumbency are more pronounced the more marginal the seat. The rationale behind this choice is that in more marginal seats, more resources are poured into the constituency in order to improve the chances of holding the seat come the next election. This often means that the incumbent MP and party have to work harder in order to hold on, making their local presence more strongly felt.

To account for this, we implement a locally weighted regression using a kernel of our choosing and solve for the target point $x_0 = 0$ which represents the margin of victory[8]:

$$\min_{\beta} \sum_{i=1}^{N} K_h(x_i)[y_i - \beta_0 - \beta_1 x_i - \beta_2 z_i - \beta_3 x_i z_i]^2$$

with x_i and z_i being shorthand for the ConMargin and ConWon variables in (10) respectively.

The kernel, K_h, is a weighting function which assigns a weight to the vote margin, x_i depending on its distance to a tied election.

This is a problem of weighted least squares. The method of weighted least squares is one where the model under consideration is of the form:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon^*$$

where $\epsilon^* \sim N(\mathbf{0}, \Sigma)$ and the covariance matrix, Σ is given by:

$$\begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

The relationship between the weights of each observation, w_i, and the variance of each observation, σ_i^2, is given by:

$$w_i = \frac{1}{\sigma_i^2}$$

and hence heteroscedasticity is assumed. This is quite a strong assumption and presents us with a potential flaw in our model; the implication here is that the safer the seat, the greater the standard deviation of the 1st place vote share. This assumption is not entirely far fetched however as in safe seats, turnout is often suppressed and people are generally less concerned about voting for smaller parties, as it is unlikely to have an effect on the overall outcome; this assumption is also made in Eggers and Spirling[3] and so has precedent in pre-existing literature.

The kernels are indexed by a parameter (or bandwidth) that dictates the width of the neighbourhood being considered in the regression. This leaves us with finding a positive hyper-parameter, h, that provides optimal bandwidth choice for the regression discontinuity estimator.

Before beginning to find an optimal h for this specific problem, we will fix the type of kernel we are using. The kernel under consideration here is the triangular kernel which is defined as follows:

$$K_h(x) = \mathbf{1}_{|x| \leq h}\Big(1 - \frac{|x|}{h}\Big)$$

with the kernel effectively ensuring that all points lying outside of the bandwidth are not counted at all.

4.2 Finding the Optimal Kernel Parameter

The parameter h determines the width of the local neighbourhood and is conventionally chosen based on the entire support of the data. It is typically chosen such that the mean integrated square error criterion is minimised (MISE):

$$\text{MISE}(h) = \text{E}\left[\int_{x:f(x)>0} (\hat{m}_h(x) - m(x))^2 f(x)\,dx\right].$$

where $f(x)$ is the probability density function of the forcing variable which in this case is the margin of victory, $m(x)$ is the conditional expectation of the outcome given the forcing variable (which is defined globally), and $\hat{m}_h(x)$ which is defined as:

$$\hat{m}_h(x) = \begin{cases} \hat{a}_-(x) \text{ if } x < 0, \\ \hat{a}_+(x) \text{ if } x \geq 0, \end{cases}$$

with $\hat{a}_-(x)$ representing the intercept of the kernel-weighted linear regression model fitted to all data points below 0 centred at x and $\hat{a}_+(x)$ representing the intercept of the kernel-weighted linear regression model fitted to all data points above 0 centred at x.

In the context of this specific problem (i.e. finding the discontinuity jump between positive and negative margins of victory), the coefficient estimate $\hat{\beta}_2 = \hat{a}_+(0) - \lim_{x\uparrow 0}\hat{a}_-(x)$. This makes the MISE method problematic as we only seek a bandwidth that is optimal for estimating the jump in the discontinuity; the MISE uses all values of x in the support of $f(x)$ whereas the jump in the discontinuity depends on $m(x)$ only through two values with both values being boundary values. We therefore consider an alternative mean squared error specifically with respect to β_2 as follows[9]:

$$\text{MSE}(h) = E[(\hat{\beta}_2 - \beta_2)^2] \tag{11}$$

and denote h^* as the optimal bandwidth which minimises (11).

Working with (11) directly gives rise to a few problems, as the optimal bandwidth h^* will not converge to zero as the sample size becomes infinite. This is due to biases in different parts of the regression function offsetting one another. The following example illustrates this particular phenomenon.

Suppose we seek to estimate a regression function, $g(x)$, at the point $x = 0$ and the covariates X_i are independent and uniformly distributed on [0,1]. Let us additionally suppose that the true regression function is defined as $g(x) = (x - \frac{1}{4})^2 - \frac{1}{16}$. Using a uniform function as the kernel in the kernel density estimation[10], we obtain the following estimator for $g(0)$ for a fixed bandwidth $h > 0$:

$$\hat{g}(0) = \frac{1}{h} \cdot \frac{\sum_i g(X_i) \cdot \mathbf{1}_{X_i < h}}{\sum_{i:X_i<h} 1} = \frac{1}{h} \cdot \frac{\sum_{i:X_i<h} g(X_i)}{\sum_{i:X_i<h} 1}$$

For a fixed h, let us let $N = \sum_{i:X_i<h} 1$ (i.e. the number of covariates in our sample that is used in the estimator). We will now compute $E(\hat{g}(0))$ to test the bias of this estimator:

$$E(\hat{g}(0)) = \frac{1}{hN} \cdot E(\sum_{i:X_i<h} g(X_i)) = \frac{1}{hN} \cdot \sum_{i:X_i<h} E(g(X_i)) = \frac{N}{hN} \cdot E(g(X_i)) = \frac{1}{h} \cdot E(g(X_i))$$

This now gives us an expectation which we can analytically compute. We will assume $h < 1$. With the imposed constraint on the covariates being less than h, we only consider points in the interval $[0, h]$ in the expectation. The pdf of the uniform distribution we are considering is simply 1 across [0,1] and 0 everywhere else. This makes the expectation as follows:

$$E(g(X_i)) = \int_0^h g(x)\ dx = \int_0^h x^2 - \frac{1}{2}x\ dx = \frac{h^3}{3} - \frac{h^2}{4}$$

Combining with what we know about $E(\hat{g}(0))$ we find that the bias of this estimator is:

$$\text{bias}(\hat{g}(0)) = E(\hat{g}(0)) - g(0) = \frac{1}{h}(\frac{h^3}{3} - \frac{h^2}{4}) - g(0) = \frac{h^2}{3} - \frac{h}{4}$$

The bias of this estimator is therefore zero when $h = \frac{3}{4}$ showing us that when the sample size gets large, the optimal bandwidth will converge to $\frac{3}{4}$ as opposed to 0 rendering a method based on minimising (11) of little use to this specific problem. In addition to this, it makes little sense to base the estimation of h on global criteria, when identification is local.

A bandwidth choice that we will follow will involve finding the bandwidth that minimises a first order approximation to MSE(h) and we will call this the asymptotic mean squared error (AMSE). This method is proposed in Imbens and Kalyanaraman [9]. Prior to stating the AMSE as a function of the bandwidth h, we will set out the key assumptions:

- **Assumption 1** : The marginal distribution of the forcing variable x_i, denoted by $f(x)$, is continuous and bounded away from zero at the thresold 0.
- **Assumption 2** : The conditional mean $m(x) = E[y_i|x_i = x]$ has at least three continuous derivatives in an open neighbourhood of $x_i = 0$. We denote the right and left limits of the kth derivative of $m(x)$ at the threshold 0 as $m_+^{(k)}(c)$ and $m_-^{(k)}(c)$ respectively. Here, $c = 0$.
- **Assumption 3** : The kernel, K, is non-negative, bounded, differs from zero on a compact interval $[0, a]$, and is continuous on $(0, a)$.
- **Assumption 4** : The conditional variance function $\sigma^2(x) = \text{Var}[y_i|x_i = x]$ is bounded in an open neighbourhood of $x_i = 0$ and right and left continuous at 0. We denote the right and left limits at the threshold 0 as $\sigma_+^2(c)$ and $\sigma_-^2(c)$respectively. Both of these values must be strictly positive.
- **Assumption 5** : The second derivatives from the right and left differ at the threshold, i.e.: $m_+^{(2)}(c) \neq m_-^{(2)}(c)$.

With the assumptions having been stated, we can now define AMSE as follows:

$$\text{AMSE}(h) = C_1 \cdot h^4 \cdot \left(m_+^{(2)}(0) - m_-^{(2)}(0)\right)^2 + \frac{C_2}{N \cdot h} \cdot \left(\frac{\sigma_+^2(0)}{f(0)} + \frac{\sigma_-^2(0)}{f(0)}\right) \tag{12}$$

where N is the total number of data points in use and the constants C_1 and C_2 are functions of the kernel:

$$C_1 = \frac{1}{4}(\frac{v_2^2 - v_1 v_3}{v_2 v_0 - v_1^2})^2,$$
$$C_2 = \frac{v_2^2 \pi_0 - 2v_1 v_2 \pi_1 + v_1^2 \pi_2}{(v_2 v_0 - v_1^2)^2}$$

where

$$v_j = \int_0^\infty u^j K(u) du,$$
$$\pi_j = \int_0^\infty u^j K^2(u) du.$$

Now that we have an explicit form of AMSE with respect to h we can differentiate (12) wrt h to obtain a minimum by setting its derivative to zero:

$$\frac{d\text{AMSE}(h)}{dh} = 4C_1 \cdot h^3 \cdot \Big(m_+^{(2)}(0) - m_-^{(2)}(0)\Big)^2 - \frac{C_2}{N \cdot h^2} \cdot \Big(\frac{\sigma_+^2(0)}{f(0)} + \frac{\sigma_-^2(0)}{f(0)}\Big) = 0$$
$$\Rightarrow h^5 = \frac{C_2}{4C_1} \cdot \frac{\sigma_+^2(0) + \sigma_-^2(0)}{N \cdot f(0) \cdot (m_+^{(2)}(0) - m_-^{(2)}(0))^2}$$
$$\Rightarrow h = \Big(\frac{C_2}{4C_1}\Big)^{\frac{1}{5}} \cdot \Big(\frac{\sigma_+^2(0) + \sigma_-^2(0)}{f(0) \cdot (m_+^{(2)}(0) - m_-^{(2)}(0))^2}\Big)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}}$$

This is therefore a stationary point of AMSE. To verify that this is indeed a global minimum, we will consider the sign of the second derivative at this stationary point:

$$\frac{d^2\text{AMSE}(h)}{dh^2} = 12C_1 \cdot h^2 \cdot \Big(m_+^{(2)}(0) - m_-^{(2)}(0)\Big)^2 + \frac{2C_2}{N \cdot h^3} \cdot \Big(\frac{\sigma_+^2(0)}{f(0)} + \frac{\sigma_-^2(0)}{f(0)}\Big)$$

The second derivative is positive if h is positive and the h that results in the first derivative being equal to zero is positive and unique. Therefore, this value of h is a minimum on the interval $(0, \infty)$ as AMSE(h) goes to ∞ as h goes to ∞ and as h goes to zero from the right (due to the $\frac{C_2}{N \cdot h}$ term in (12)).

This tells us that the optimal h, h_{opt}, is as follows:

$$h_{opt} = C_K \cdot \Big(\frac{\sigma_+^2(0) + \sigma_-^2(0)}{f(0) \cdot (m_+^{(2)}(0) - m_-^{(2)}(0))^2}\Big)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}} \tag{13}$$

where $C_K = (\frac{C_2}{4C_1})^{\frac{1}{5}}$.

It would make sense for us to replace each of the unknown quantities in (13) with consistent estimators, culminating in:

$$\hat{h}_{opt} = C_K \cdot \Big(\frac{\hat{\sigma}_+^2(0) + \hat{\sigma}_-^2(0)}{\hat{f}(0) \cdot (\hat{m}_+^{(2)}(0) - \hat{m}_-^{(2)}(0))^2}\Big)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}} \tag{14}$$

One potential issue with this estimator of h_{opt} is that we have relied on the treatment effect (which in this case is whether the party won the previous election) not being constant additive (i.e. $m_+^{(2)}(0) \neq m_-^{(2)}(0)$) since were this to be the case, the bandwidth estimator would be infinite, resulting in all the data points being included. Furthermore, the estimates of the second derivatives, $m_+^{(2)}(0)$ and $m_-^{(2)}(0)$, are not likely to be obtained with a high degree of precision, meaning that the estimated optimal bandwidth has the potential to be very large.

Imbens and Kalyanaraman[9] propose the following alternative estimate for h_{opt} which introduces a regularisation term in order to deal with the aforementioned problems that may arise:

$$\hat{h}_{opt} = C_K \cdot \Big(\frac{\hat{\sigma}_+^2(0) + \hat{\sigma}_-^2(0)}{(\hat{f}(0) \cdot ((\hat{m}_+^{(2)}(0) - \hat{m}_-^{(2)}(0))^2 + \hat{r}_- + \hat{r}_+))}\Big)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}} \tag{15}$$

where

$$\hat{r}_- = 3 \cdot \mathrm{Var}\big(\hat{m}_-^{(2)}(0)\big) \text{ and } \hat{r}_+ = 3 \cdot \mathrm{Var}\big(\hat{m}_+^{(2)}(0)\big).$$

This modification ensures that the bandwidth will not become infinite even in cases where the difference in curvatures at the threshold is zero. The regularisation is derived by considering the bias in the reciprocal of the squared difference in second derivatives:

$$E\Big[\frac{1}{(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c))^2}\Big] - \frac{1}{(m_+^{(2)}(c) - m_-^{(2)}(c))^2}$$

To compute the expectation, we will use the Taylor expansions for the moments of functions of random variables. Let $X = (\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c))$. We seek to compute $E(\frac{1}{X^2})$. To obtain this quantity, let us also make use of the following relationship:

$$E(\frac{1}{X^2}) = E(\frac{1}{X})^2 + \mathrm{Var}(\frac{1}{X})$$

The first moment of a function, $f(X)$, of a random variable with mean μ_X is given as:

$$E(f(X)) = E(f(\mu_X + (X - \mu_X))) = E(f(\mu_X) + f'(\mu_X)\,(X - \mu_X) + \frac{1}{2}f''(\mu_X)\,(X - \mu_X)^2 + \textbf{h.o.t.})$$
$$\approx f(\mu_X) + \frac{f''(\mu_X)}{2}\sigma_X^2$$

The terms of higher order are in terms of $(X - \mu_X)^n$, where $n \geq 3$.

The first order approximation for the second moment is given as:

$$\mathrm{Var}(f(X)) \approx (f'(\mu_X))^2 \mathrm{Var}(X)$$

The terms omitted are moments $E(X - \mu_X)^n$ for $n \geq 4$.

Thus, taking $f(X) = \frac{1}{X}$, we obtain the following:

$$E(\frac{1}{X^2}) \approx (\frac{1}{\mu_X} + \frac{\mathrm{Var}(X)}{\mu_X^3})^2 + \frac{\mathrm{Var}(X)}{\mu_X^4} \approx \frac{1}{\mu_X^2} + 2\frac{\mathrm{Var}(X)}{\mu_X^4} + \frac{\mathrm{Var}(X)}{\mu_X^4}$$

We require that our estimate of $(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c))$ is unbiased and consistent, so we can take $\mu_X = (m_+^{(2)}(c) - m_-^{(2)}(c))$. In addition to this, we will assume that $\hat{m}_+^{(2)}(c)$ and $\hat{m}_-^{(2)}(c)$ are independent, so that $\mathrm{Var}(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c)) = \mathrm{Var}(\hat{m}_+^{(2)}(c)) + \mathrm{Var}(\hat{m}_-^{(2)}(c))$.

Taking this all into account, we arrive at the conclusion that the bias is equal to:

$$\frac{1}{\mu_X^2} + 2\frac{\mathrm{Var}(X)}{\mu_X^4} + \frac{\mathrm{Var}(X)}{\mu_X^4} - \frac{1}{(m_+^{(2)}(c) - m_-^{(2)}(c))^2} + O(N^{-2\alpha})$$

$$= \frac{1}{\mu_X^2} + 3\frac{\mathrm{Var}(X)}{\mu_X^4} - \frac{1}{\mu_X^2} + O(N^{-2\alpha})$$

$$= 3\frac{\mathrm{Var}(\hat{m}_+^{(2)}(c)) + \mathrm{Var}(\hat{m}_-^{(2)}(c))}{(m_+^{(2)}(c) - m_-^{(2)}(c))^4} + O(N^{-2\alpha}) \tag{16}$$

where $\alpha > 1$ and the errors arise from the omission of the moments above and including the third; as the sample sizes increases, the errors decrease at a rate faster than the error associated with the variance.

The regularisation given in (15) eliminates the $\text{Var}(\hat{m}_+^{(2)}(c)) + \text{Var}(\hat{m}_-^{(2)}(c))$ term in the numerator of (16), and hence under this regularisation, the bias tends to zero at a faster rate as N increases, while also preventing the denominator of (15) from ever equalling zero; the regularisation thus improves the situation two-fold.

In order to implement the regularisation, we require knowing three times the variance of the estimated curvatures on the left and the right of the threshold. We estimate the second derivative by fitting a quadratic function to all of the observations that fit inside a chosen neighbourhood ([0,h] or [-h,0] depending on right/left) whose choice will be explained and determined later on. Letting $N_{h,+}$ be the number of samples in the right hand interval and letting:

$$\hat{\mu}_{j,h,+} = \frac{1}{N_{h,+}} \sum_{0 \leq X_i \leq h} (X_i - \bar{X})^j$$

be the jth moment of X_i in the interval [0,h] we can then use the following explicit formula for three times the conditional variance of the curvature to the right of the threshold by:

$$r_+ = \frac{12}{N_{h,+}} \cdot \Big(\frac{\sigma_+^2(0)}{\hat{\mu}_{4,h,+} - (\hat{\mu}_{2,h,+})^2 - \frac{(\hat{\mu}_{3,h,+})^2}{\hat{\mu}_{2,h,+}}} \Big). \tag{17}$$

Fourth moments are very difficult to estimate precisely since larger samples are required in order to obtain estimates of similar quality. This is due to the excess degrees of freedom consumed by the higher orders. Thus, we approximate the forcing variable over a small h as being uniformly distributed on the domain [0,h]. This results in the following moment approximations being:

$$\begin{aligned} \hat{\mu}_{2,h,+} &= \frac{h^2}{12}, \\ \hat{\mu}_{3,h,+} &= 0, \\ \hat{\mu}_{4,h,+} &= \frac{h^4}{60}. \end{aligned}$$

Plugging these approximations into (17), we obtain:

$$\hat{r}_+ = \frac{2160 \cdot \hat{\sigma}_+^2(0)}{N_{h,+} \cdot h^4} \text{ and similarly } \hat{r}_- = \frac{2160 \cdot \hat{\sigma}_-^2(0)}{N_{h,-} \cdot h^4}.$$

With the theory behind the regularisation covered, we can now lay out an algorithm in order to obtain a completely data-driven bandwidth choice from (15).

- **Step 1** : Estimation of density $f(0)$ and conditional variances $\sigma_-^2(0)$ and $\sigma_+^2(0)$

  We start by computing the sample variance of the forcing variable (margin of victory):

$$S_X^2 = \sum \frac{(X_i - \bar{X})^2}{N-1}$$

  We can now turn to the Silverman rule to get a pilot bandwidth for calculating the density and variance at the threshold.

  The standard Silverman rule of $h = 1.84 \cdot S_X \cdot N^{-\frac{1}{5}}$ which uses a uniform kernel and normal reference density is used. The uniform kernel is chosen because we are interested in an estimate of the density which is the proportion of observations near the threshold. A normal reference density is used, as we assume the vote margin to be approximately normally distributed; this is not an unreasonable assumption as the modal vote margin hovers around zero and the rarest vote margins are the highest vote margins (either way).

  We then calculate the number of observations on either side of the threshold within the pilot bandwidth:

$$N_{h_1,-} = \sum_{i=1}^{N} \mathbf{1}_{-h<X_i<0}, \quad N_{h_1,+} = \sum_{i=1}^{N} \mathbf{1}_{0<X_i<h}$$
$$\bar{Y}_{h_1,-} = \frac{1}{N_{h_1,-}} \sum_{i:-h\leq X_i<0} Y_i, \quad \bar{Y}_{h_1,+} = \frac{1}{N_{h_1,+}} \sum_{i:0\leq X_i<h} Y_i$$

  Given this information, we estimate the density of X_i at 0 as:

$$\hat{f}(0) = \frac{N_{h_1,-} + N_{h_1,+}}{2 \cdot N \cdot h_1}$$

  and the conditional variances of Y_i at both sides of the threshold as:

$$\hat{\sigma}_-^2(0) = \frac{1}{N_{h_1,-} - 1} \sum_{i:-h\leq X_i<0} (Y_i - \bar{Y}_{h_1,-})^2$$
$$\hat{\sigma}_+^2(0) = \frac{1}{N_{h_1,+} - 1} \sum_{i:0\leq X_i<h} (Y_i - \bar{Y}_{h_1,+})^2$$

  All the above estimators are consistent as the bandwidth goes to zero at a rate of $N^{-\frac{1}{5}}$ and assumptions 1 and 4 hold.

- **Step 2** : Estimation of second derivatives $\hat{m}_+^{(2)}(0)$ and $\hat{m}_-^{(2)}(0)$

  Before proceeding, we need pilot bandwidths $h_{2,-}$ and $h_{2,+}$ which will be based on a rudimentary estimator of the third derivative of m at 0; we will fit a third-order polynomial to the data using least squares regression, including an indicator for the forcing variable:

$$Y_i = \gamma_0 + \gamma_1 \cdot 1_{X_i>0} + \gamma_2 \cdot X_i + \gamma_3 \cdot X_i^2 + \gamma_4 \cdot X_i^3 + \epsilon_i$$

and then compute $\hat{m}^{(3)}(0) = 6 \cdot \gamma_4$ as our estimate of the third derivative of the regression function. It is worth noting that this in general is not a consistent estimator, however it will converge to some constant as the sample size is increased. To compute the pilot bandwidths we use a result from Fan and Gijbels[11] which provides us with an asymptotic approximation to the mean squared error for the mth derivative of a regression function at the boundary point using a n order local polynomial. Here, $m = n = 2$ and a one-sided uniform kernel is used for each of the two bandwidths. The AMSE that is to be minimised with respect to h for the pilot bandwidth pertaining to the second derivative on the right hand side of the boundary is as follows:

$$\text{AMSE}(h) = \left(\frac{1}{4}(m^{(3)}(0))^2 h^2 + \frac{720}{Nh^5} \cdot \frac{\sigma_+^2(0)}{f(0)} \right)$$

Similar to (12), we can now differentiate this with respect to h and set its derivative to zero:

$$\begin{aligned} \frac{d\text{AMSE}(h)}{dh} &= \left(\frac{1}{2}(m^{(3)}(0))^2 h - \frac{3600}{Nh^6} \cdot \frac{\sigma_+^2(0)}{f(0)} \right) = 0 \\ \Rightarrow h^7 &= 7200 \cdot \frac{\sigma_+^2(0)}{f(0) \cdot (m^{(3)}(0))^2} \cdot N^{-1} \\ \Rightarrow h &= 7200^{\frac{1}{7}} \cdot \left(\frac{\sigma_+^2(0)}{f(0) \cdot (m^{(3)}(0))^2} \right)^{\frac{1}{7}} \cdot N^{-\frac{1}{7}} \end{aligned}$$

Similar to before, the second derivative evaluated at this minimum is positive and so this value yields a minimum over the positive reals. This culminates in the following bandwidths being used for estimation of the second derivatives:

$$\begin{aligned} h_{2,+} &= 3.56 \cdot \left(\frac{\hat{\sigma}_+^2(0)}{\hat{f}(0) \cdot (\hat{m}^{(3)}(0))^2} \right)^{\frac{1}{7}} \cdot N_+^{-\frac{1}{7}} \\ h_{2,-} &= 3.56 \cdot \left(\frac{\hat{\sigma}_-^2(0)}{\hat{f}(0) \cdot (\hat{m}^{(3)}(0))^2} \right)^{\frac{1}{7}} \cdot N_-^{-\frac{1}{7}} \end{aligned}$$

with N_- and N_+ representing the number of observations to the left and right of the threshold, respectively.

Given these pilot bandwidths, we can now estimate the curvatures $\hat{m}_+^{(2)}(0)$ and $\hat{m}_-^{(2)}(0)$ by fitting a local quadratic to all the data points to the right of the threshold and to the left of the threshold respectively by using least squares regression:

$$Y_i = \lambda_0 + \lambda_1 \cdot X_i + \lambda_2 \cdot X_i^2 + \epsilon_i$$

with the curvature being estimated as $2 \cdot \lambda_2$.

- **Step 3** : Calculation of the regularisation terms $\hat{r}_-$ and $\hat{r}_+$ and of $\hat{h}_{opt}$

  With the previous steps taken into account, the regularisation terms can be calculated as follows:

$$\hat{r}_+ = \frac{2160 \cdot \hat{\sigma}_+^2(0)}{N_{2,+} \cdot h_{2,+}^4} \text{ and } \hat{r}_- = \frac{2160 \cdot \hat{\sigma}_-^2(0)}{N_{2,-} \cdot h_{2,-}^4}.$$

  allowing us to calculate the proposed bandwidth:

$$\hat{h}_{opt} = C_K \cdot \Big(\frac{\hat{\sigma}_+^2(0) + \hat{\sigma}_-^2(0)}{(\hat{f}(0) \cdot ((\hat{m}_+^{(2)}(0) - \hat{m}_-^{(2)}(0))^2 + \hat{r}_- + \hat{r}_+))} \Big)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}} \tag{18}$$

4.3 Simulated Example

We will demonstrate this algorithm on simulated data.

Figure 12 shows the simulated data with a discontinuity/jump at the change in sign of the margin (see appendix for how data was specifically simulated). There are 600 data points with 300 on either side of the discontinuity. To compute the optimal bandwidth (which can also be seen in figure 12) we implement the aforementioned algorithm:

- **Step 1** : Estimation of density $f(0)$ and conditional variances $\sigma_-^2(0)$ and $\sigma_+^2(0)$

$$h_1 = 1.84 \cdot S_X \cdot N^{-\frac{1}{5}} = 1.84 \cdot 0.174 \cdot 600^{-\frac{1}{5}} = 0.0889$$

  There are 88 units either side and so $N_{h_1,+} = N_{h_1,-} = 88$ and sample variances $S^2_{Y,h_1,-} = 0.0249^2$ and $S_{Y,h_1,+} = 0.0256^2$. This results in:

$$\hat{f}(0) = \frac{N_{h_1,-} + N_{h_1,+}}{2 \cdot N \cdot h_1} = \frac{88 + 88}{2 \cdot 600 \cdot 0.0889} = 1.65,$$

$$\hat{\sigma}_-^2(0) = 0.025^2 \text{ and } \hat{\sigma}_+^2(0) = 0.026^2$$

- **Step 2** : Estimation of second derivatives $\hat{m}_+^{(2)}(0)$ and $\hat{m}_-^{(2)}(0)$

Fitting a global cubic with a jump at the threshold, we obtain that $\gamma_4 = -0.249$ leading to our estimate of the third derivative at the threshold as $\hat{m}^{(3)}(0) = 6 \cdot \gamma_4 = -1.49$ which produces the following bandwidths:

$$h_{2,+} = 3.56 \cdot \Big(\frac{\hat{\sigma}_+(0)}{\hat{f}(0) \cdot (\hat{m}^{(3)}(0))^2}\Big)^{\frac{1}{7}} \cdot N_+^{-\frac{1}{7}} = 0.459$$

$$h_{2,-} = 3.56 \cdot \Big(\frac{\hat{\sigma}_-(0)}{\hat{f}(0) \cdot (\hat{m}^{(3)}(0))^2}\Big)^{\frac{1}{7}} \cdot N_-^{-\frac{1}{7}} = 0.455$$

Fitting two quadratics (one to the right with the first threshold and the other to the left with the second threshold) we obtain the following estimates:

$$\hat{m}_+^{(2)}(0) = -0.666 \text{ and } \hat{m}_-^{(2)}(0) = -0.0601$$

- **Step 3** : Calculation of the regularisation terms $\hat{r}_-$ and $\hat{r}_+$ and of $\hat{h}_{opt}$

  Calculating the regularisation terms, we obtain:

$$\hat{r}_+ = 0.106 \text{ and } \hat{r}_- = 0.104$$

Using the triangular kernel with $C_K = 3.44$ we obtain the optimal bandwidth estimate:

$$\hat{h}_{opt} = 3.44 \cdot \Big(\frac{\hat{\sigma}_+^2(0) + \hat{\sigma}_-^2(0)}{(\hat{f}(0) \cdot ((\hat{m}_+^{(2)}(0) - \hat{m}_-^{(2)}(0))^2 + \hat{r}_- + \hat{r}_+))}\Big)^{\frac{1}{5}} \cdot 600^{-\frac{1}{5}} = 0.255 \tag{19}$$

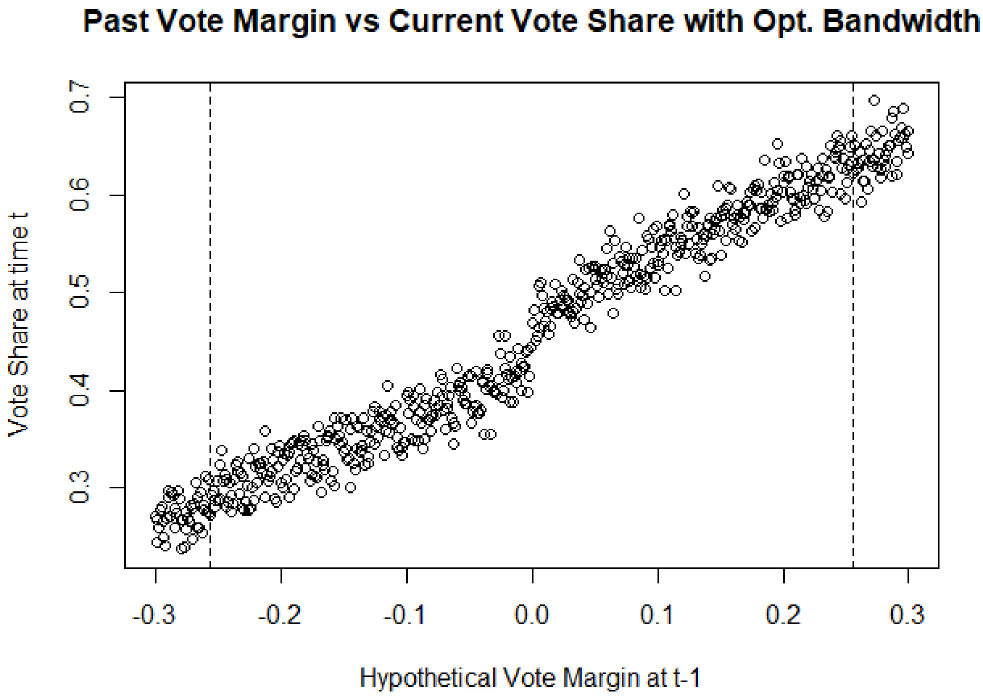


Figure 12: Simulated election data demonstrating the need for residual discontinuity design. The intercept here clearly differs depending on whether or not the Conservative won at time $t-1$, with the change in intercept representing the incumbency effect.

Figure 13 shows the relative weighting of the data points when linearly regressed upon; this represents the kernel which we can see is triangular and the output of this fitted model can be seen in table 3.

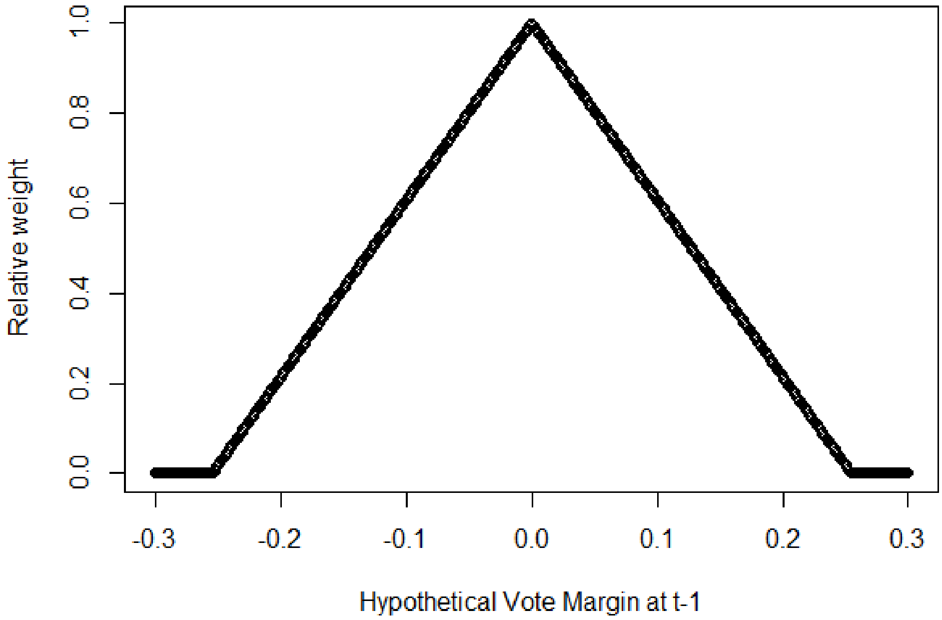


Figure 13: The weights of each observation used to predict the incumbency effect. This is the triangular kernel (commonly referred to as the edge kernel)

Table 3: Results of model (10) on simulated election data

	Dependent variable:
	Simulated Con Vote Share at time t
$\hat{\beta}_0$	0.421***
	(0.002)
$\hat{\beta}_1$	0.519***
	(0.020)
$\hat{\beta}_2$	0.054***
	(0.003)
$\hat{\beta}_3$	0.127***
	(0.028)
Observations	600
R^2	0.952
Adjusted R^2	0.951
Residual Std. Error	0.013 (df = 504)
F Statistic	3,302.849*** (df = 3; 504)
Note:	*p<0.1; **p<0.05; ***p<0.01

As is expected, there is clearly an incumbency bias within this simulated data which is given by 5.4% and is significant at the 99.9% level.

4.4 Results

To include as many data points as possible we consider all constituencies in England save for Chorley and Buckingham as they are both constituencies that have been filled by the Speaker of the House and as such, neither Labour nor the Conservatives run candidates in these constituencies. This makes $N = 531$. Furthermore, we only consider election results from 2015 since it allows us to work with all constituencies incumbency-wise; the boundaries were last changed in 2010 and so data from 2015 till now are complete.

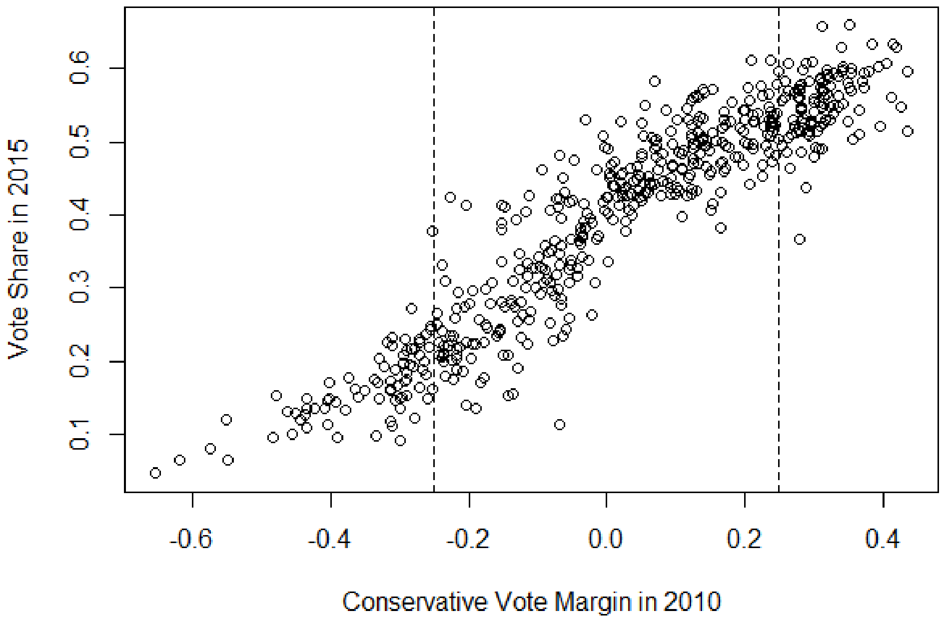


Figure 14: 2015 Conservative results in England with optimal bandwidth

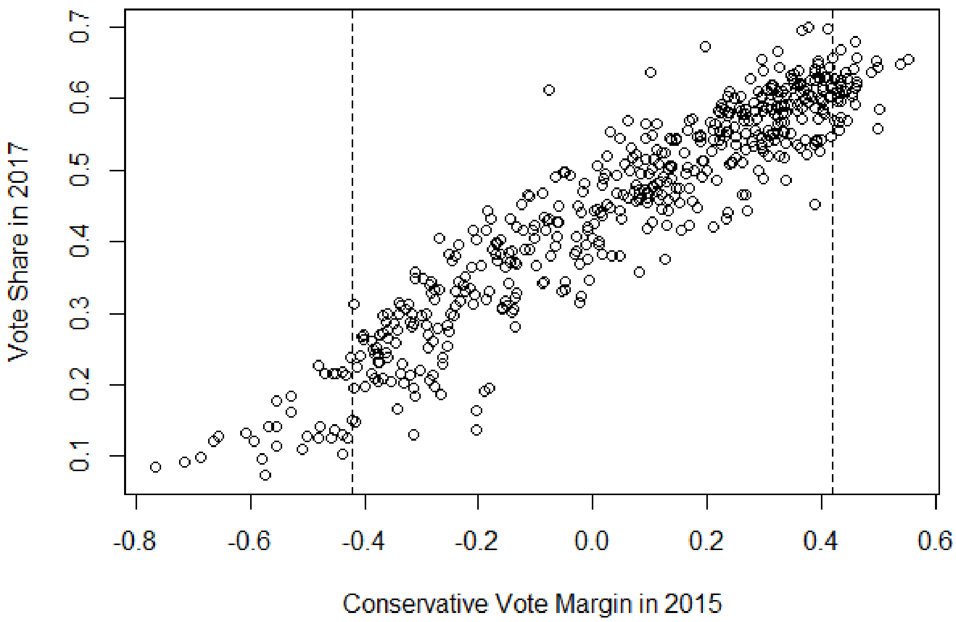


Figure 15: 2017 Conservative results in England with optimal bandwidth

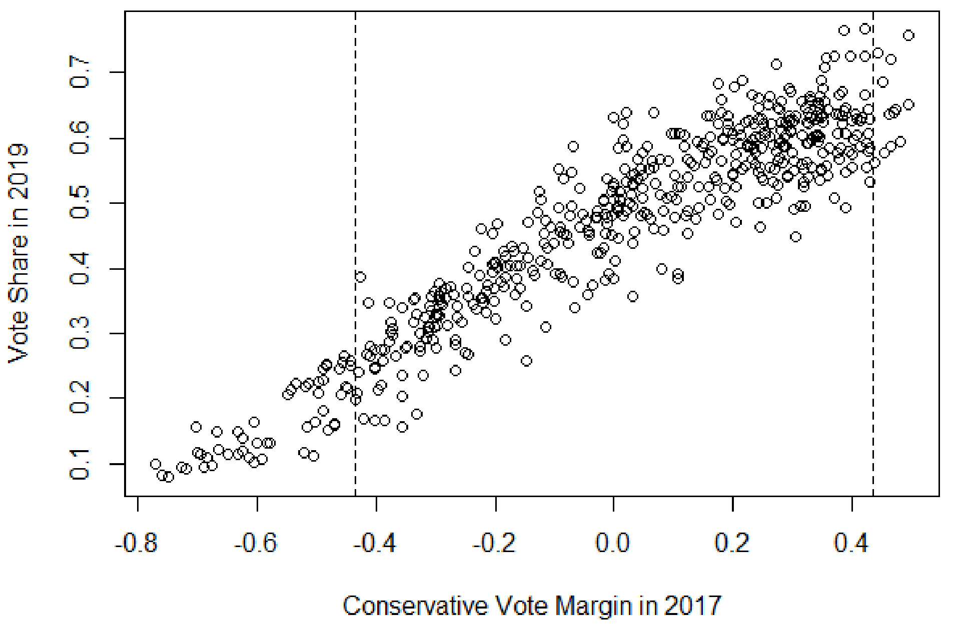


Figure 16: 2019 Conservative results in England with optimal bandwidth

Implementing the algorithm on the data (see appendix for algorithms), we can see that incumbency advantage given to Conservative party candidates where the previous constituency winner was a Conservative manifests itself in 2015 and 2019 with a predicted advantage of $\sim 2.5\%$ and $\sim 1.8\%$ respectively. In 2017, at the 95% significance level, there was no party incumbency advantage for Conservative candidates as can be seen from figure 17.

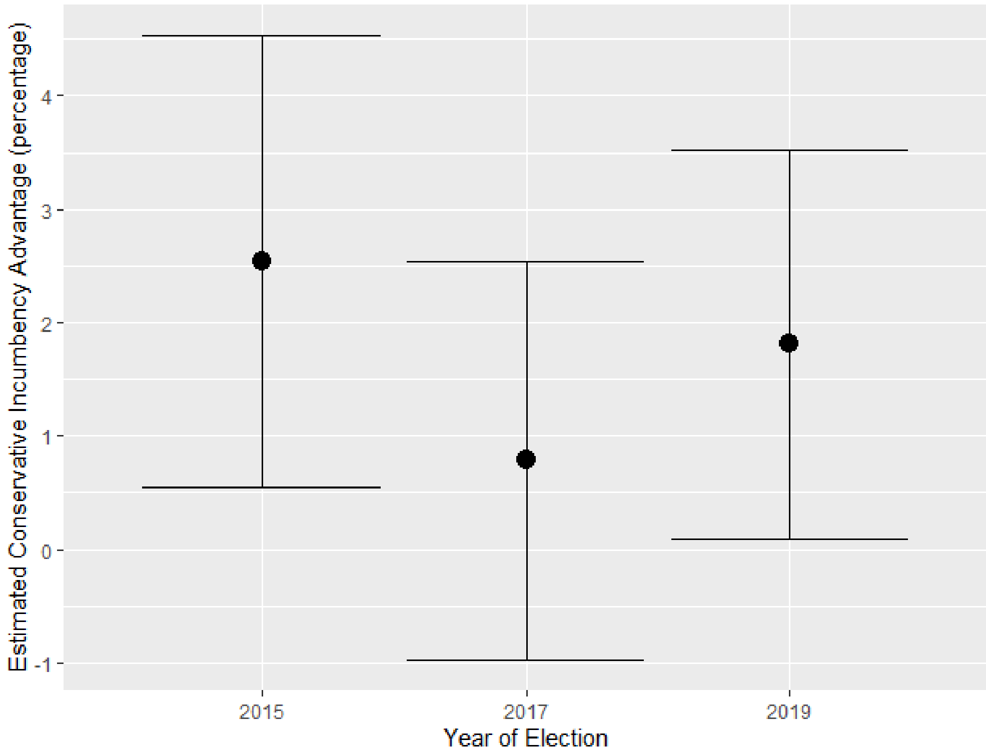


Figure 17: Conservative incumbency advantage estimates with 95% confidence intervals

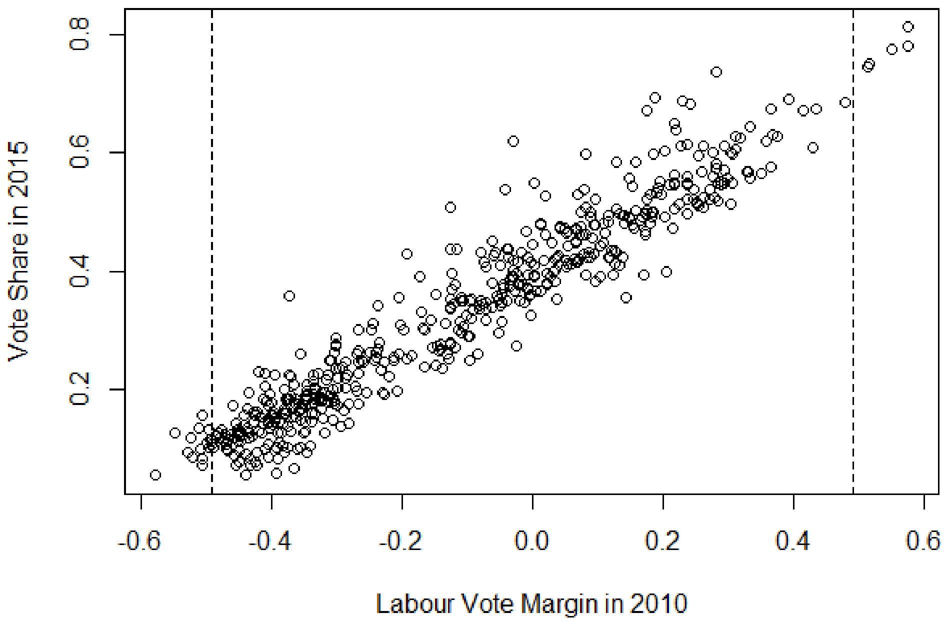


Figure 18: 2015 Labour results in England with optimal bandwidth

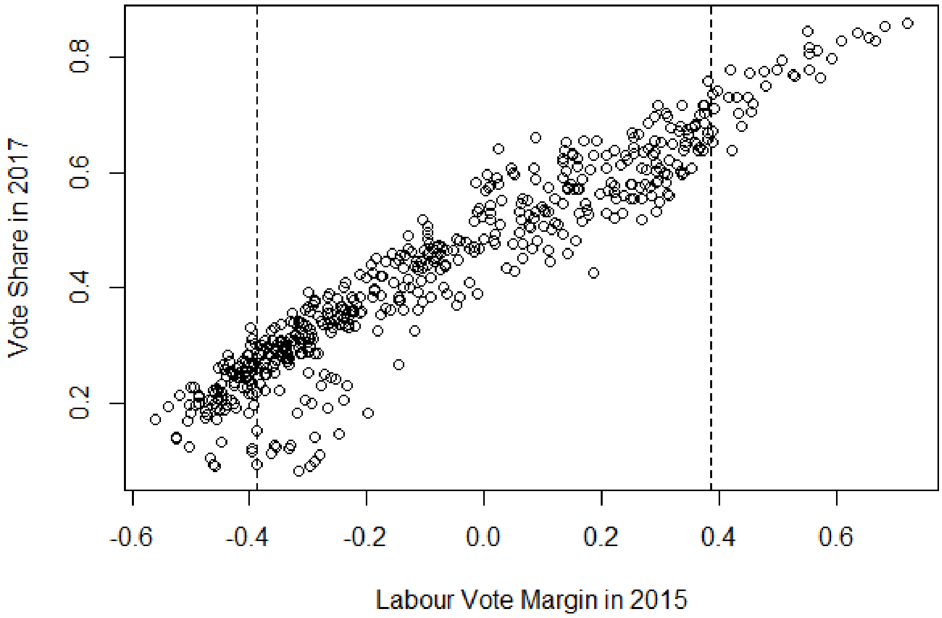


Figure 19: 2017 Labour results in England with optimal bandwidth

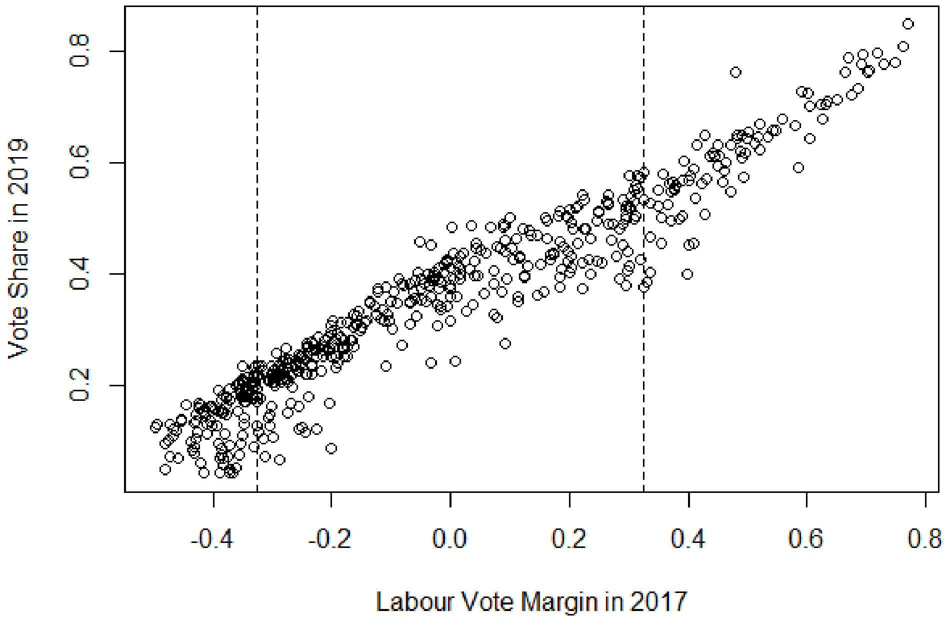


Figure 20: 2019 Labour results in England with optimal bandwidth

Implementing the algorithm on the data for the Labour party, we can see that incumbency advantage given to Labour party candidates where the previous constituency winner was a Labour manifests itself in 2017 with a predicted advantage of $\sim 2\%$. In 2015 and 2019, at the 95% significance level, there was no party incumbency advantage for Labour candidates as can be seen from figure 21.

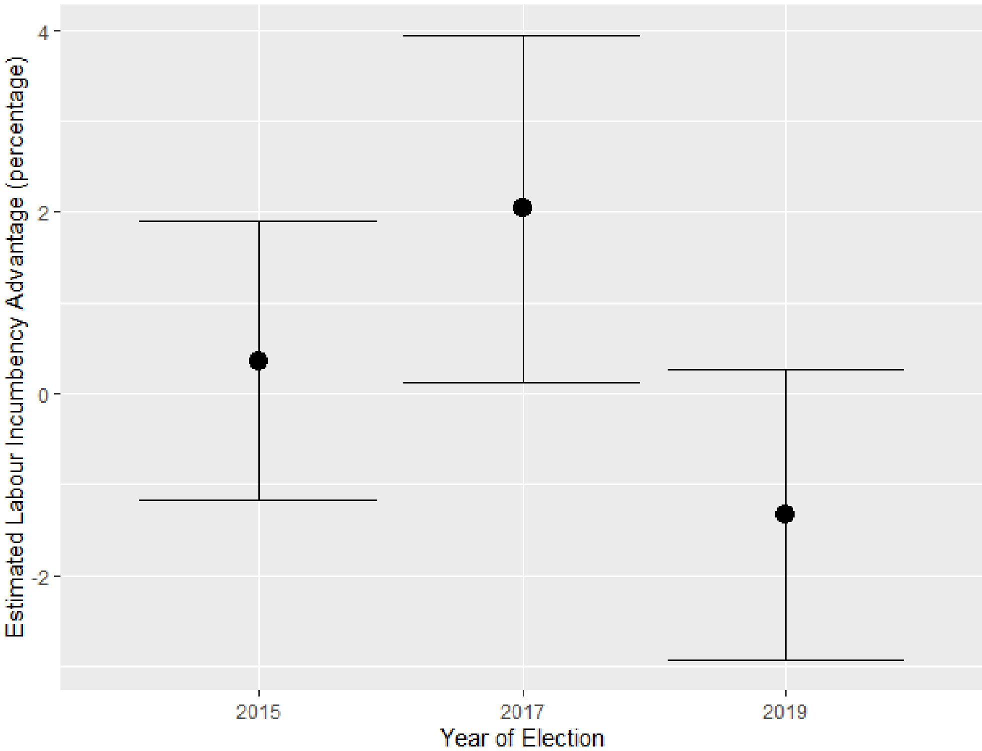


Figure 21: Labour incumbency advantage estimates with 95% confidence intervals

5 Estimating Incumbency Advantage by Individual and by Party

As was touched upon in Section 2, omission of personal incumbency advantage or incumbency attributed to incumbency of the political party may incur an omitted variable bias. The theory derived in Section 2 is applicable in the case of weighted least squares, as the only fundamental difference in the set-up is the behaviour of the variance for different response variables (heteroscedastic instead of homoscedastic); bias deals only with expectation and hence the direction of the bias and underlying theory are unchanged.

Our model from Section 4 leaves out the information on personal incumbency and if this correlates with one of the covariates in (10) and the vote share at time t, then our estimates for the coefficients in the model will be biased and hence inaccurate. Indeed, the cause for concern of OVB is justified as shown in figure 22. We see that the individual incumbency of the candidate correlates positively both with the vote share obtained in 2015 and all other covariates in model (20).

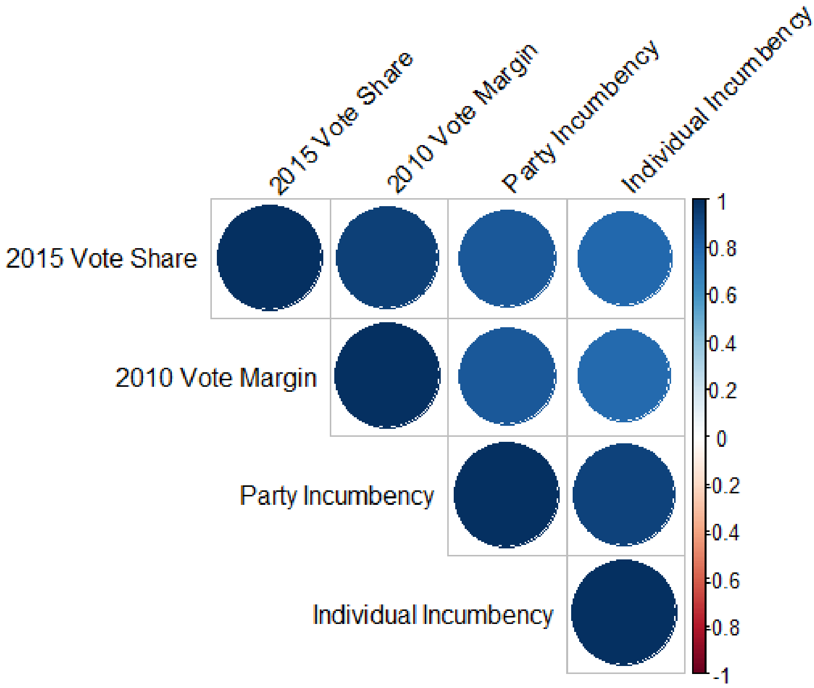


Figure 22: Correlations between covariates in (20) to show cause of OVB. The data used is from the 2015 election and the party examined is the Conservative party.

5.1 Proposed Model

We therefore propose the following model of our own which takes into account both aforementioned types of bias:

$$Y_{it} = \beta_0 + \beta_1 ConMargin_{i,t-1} + \beta_2 ConWon_{i,t-1} + \beta_3(ConWon_{i,t-1} \times ConMargin_{i,t-1}) + \beta_4 I_{i,t} + \epsilon_{it} \tag{20}$$

where I_{it} represents the status of the incumbent candidate (which is set to zero in an open race, 1 if the Conservative is the incumbent and -1 if the incumbent is not a Conservative). The same model is also carried out for Labour and the regression is locally weighted, using the same weights and bandwidths as in Section 4 for the respective elections. β_2 represents the party based incumbency and β_4 represents the personal incumbency. We use the same weights since we continue to assume incumbency bias is more pronounced, the more marginal the seat. We use the same bandwidths since the data used for the response variable and control variable, which are used in computing $\hat{h}_{opt}$, are unchanged. The behaviour of swing from one election to the next is also assumed to be uniform in the same way as in Section 4.

5.2 Results

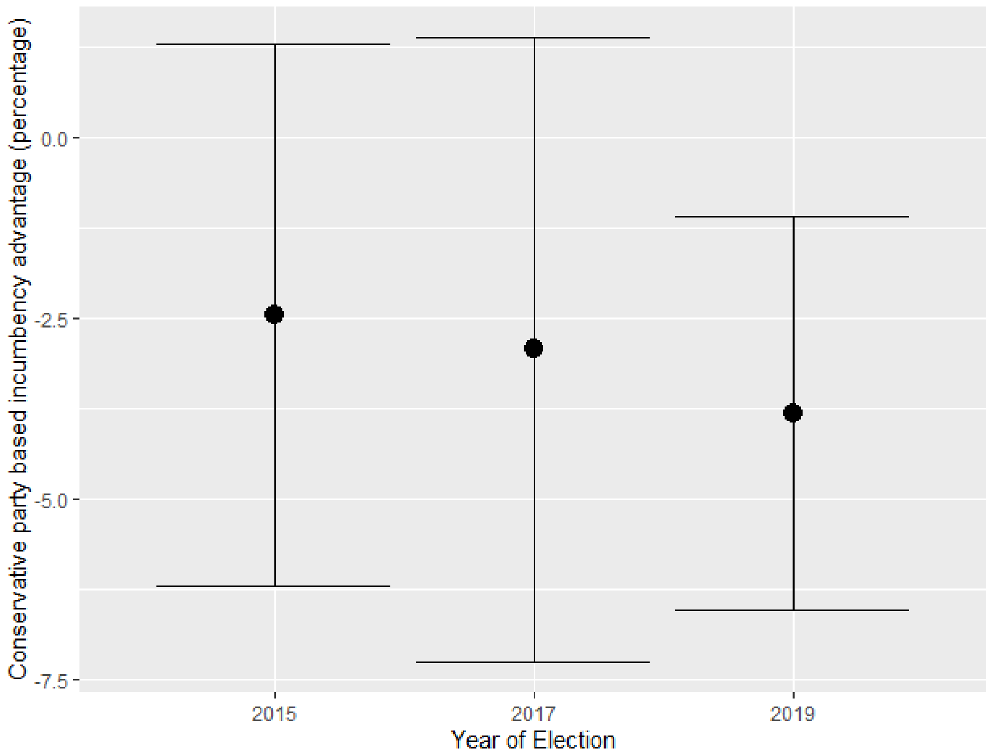


Figure 23: 95% confidence intervals for constituency incumbency of the Conservative party

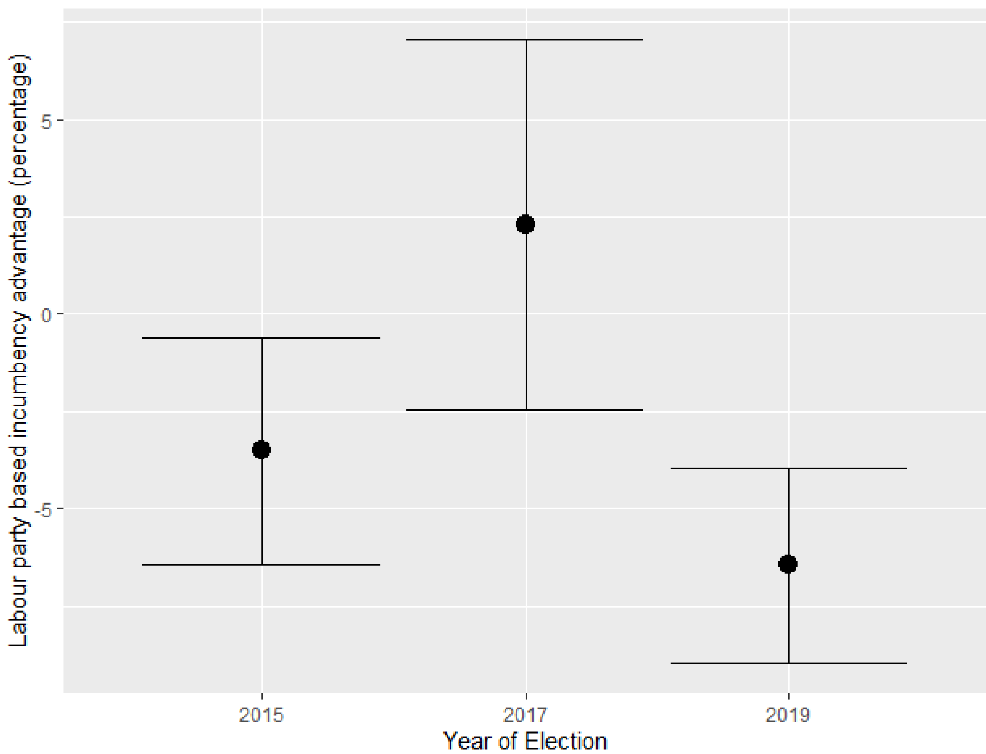


Figure 24: 95% confidence intervals for constituency incumbency of the Labour party

The estimate for constituency incumbency of the Conservative party in the 2015 general election using model (20) is ∼-2.5% as compared with ∼2.5% when using model (10). This demonstrates that the estimate in model (10) is positively biased and this is supported by the positive correlations as per figure 22. We therefore conclude that model (20) is superior as it addresses the OVB incurred by the omission of the different types of incumbency.

An extreme result worthy of note is the Labour party's negative incumbency effect in 2019. This effect is most probably embellished by the fact that the swing against Labour in 2019 was most focussed in its Northern heartlands and hence the uniform swing assumption is unreasonably applied. Not withstanding the swing assumption, the negative bias in 2019 would appear to agree with the post-mortem analyses that were performed on the Labour party's poor showing which suggested that the Labour party fought too much of an offensive campaign; Labour's lack of defence in its heartlands opened it up to haemorrhaging more votes than was needed, discrediting the party's nation-wide strategy and causing all seats held by Labour in 2017 to be negatively predisposed to the ballot box in 2019.

Figure 25 shows us that an incumbent MP in a general election will experience a personal percentage point boost of approximately ∼3% on average.

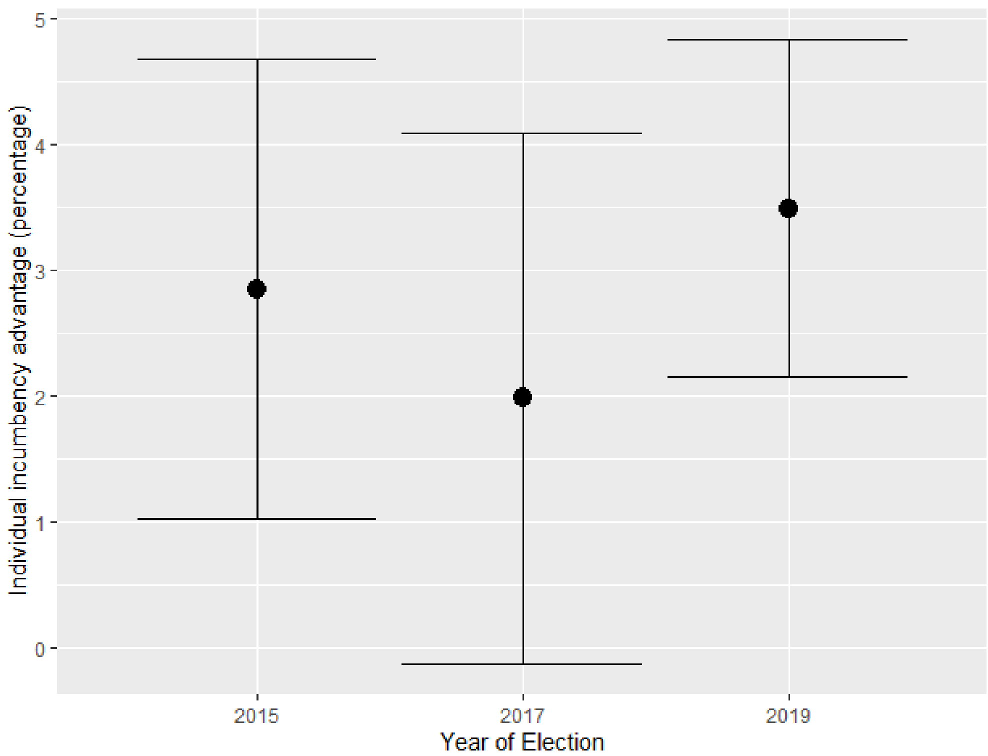


Figure 25: 95% confidence intervals for constituency incumbency of the individual candidate

6 Concluding Remarks

Across all methods in this paper, we see that personal incumbency allows the candidate to experience a positive boost in their re-election prospects. This boost is around 3% (95% C.I. [1.01%,4.53%]) using (20) and the most recent three elections.

The picture surrounding incumbency based upon political party is a lot more mixed. Figures 23 and 24 would suggest that when the effect is not significant, the incumbency bias of the party is in fact negative (for both major parties).

The precision of these estimates of incumbency bias could be improved by working with more granular data so as to accommodate the non-uniformity of swing. If the models used in this paper were to be applied on a regional level, then the number of data points in each region would be so low as to make the incumbency estimates meaningless due to the wide confidence intervals. Local elections however would present us with a way around this as there are $\sim$ 9,000 local wards in England compared to only 533 constituencies. This would allow us also to work out to specific biases associated with the three main parties as there are far more wards than constituencies that the Liberal Democrats either hold or have come second in.

In model (20), it was decided that the Brexit vote in the constituency would be omitted, as this is only a recent phenomenon and the new dividing lines arising from the Brexit debate were only drawn since the 2016 referendum. For the purposes of comparability across the three bias measurements, this was omitted. However, going forwards in all subsequent elections, it may be important to include the Brexit vote as a factor as omitting it is likely to cause OVB.

A point worthy of note and a likely limitation of the approach used in Sections 4 and 5 is the choice of kernel in the RDD method. It was assumed that the variance of the vote share of a specific party increases the greater the margin of victory in either direction. An extension to this report could be to study the volatility of vote shares in constituencies of varying margins and then subsequently fit a weighted regression with a kernel that mirrors this volatility best; finding such a kernel would be a project in and of itself. Again, this could be investigated using local election results as there is far more data with which to work than national constituency data.

Finally, incumbency bias could be investigated further in the other constituent nations of the United Kingdom to see if nationalist parties experience an incumbency boost on account of their party.

7 Appendix

7.1 Code

All code and data used in the fitting of models in this project can be found at the following link:

https://github.com/MRiach/M4R

Data that was not readily available in a database was web scraped using python's package Selenium.

7.2 Statistical Tables

Here we include the output of the fitted models in Section 5. These models formed the basis of our concluding remarks and their full output was not included in the main body of the report.

Table 4: Results of model (20) for 2015, 2017 and 2019 (left to right)

	Dependent variable:		
	Con Vote Share at time t		
	(1)	(2)	(3)
$\hat{\beta}_0$	0.429***	0.456***	0.515***
	(0.011)	(0.012)	(0.008)
$\hat{\beta}_1$	0.847***	0.510***	0.537***
	(0.077)	(0.038)	(0.036)
$\hat{\beta}_2$	-0.024	-0.029	-0.038***
	(0.019)	(0.022)	(0.014)
$\hat{\beta}_3$	-0.407***	-0.100**	-0.221***
	(0.102)	(0.049)	(0.047)
$\hat{\beta}_4$	0.028***	0.020*	0.035***
	(0.009)	(0.011)	(0.007)
Observations	531	531	531
R^2	0.673	0.725	0.718
Adjusted R^2	0.670	0.722	0.716
Residual Std. Error	0.038 (df = 341)	0.036 (df = 460)	0.038 (df = 462)
F Statistic	175.831*** (df = 4; 341)	302.850*** (df = 4; 460)	294.375*** (df = 4; 462)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: Results of model (20) from the Labour party's perspective for 2015, 2017 and 2019 (left to right)

	Dependent variable:		
	Lab Vote Share at time t		
	(1)	(2)	(3)
$\hat{\beta}_0$	0.426***	0.497***	0.430***
	(0.008)	(0.013)	(0.008)
$\hat{\beta}_1$	0.659***	0.668***	0.676***
	(0.023)	(0.037)	(0.037)
$\hat{\beta}_2$	-0.035**	0.023	-0.064***
	(0.015)	(0.024)	(0.013)
$\hat{\beta}_3$	-0.081*	-0.356***	-0.364***
	(0.043)	(0.057)	(0.057)
$\hat{\beta}_4$	0.021***	-0.001	0.033***
	(0.007)	(0.012)	(0.006)
Observations	531	531	531
R^2	0.849	0.759	0.724
Adjusted R^2	0.848	0.757	0.720
Residual Std. Error	0.037 (df = 507)	0.036 (df = 400)	0.030 (df = 328)
F Statistic	712.528*** (df = 4; 507)	315.702*** (df = 4; 400)	214.759*** (df = 4; 328)
Note:			*p<0.1; **p<0.05; ***p<0.01

8 References

[1] Robert S. Erikson, (1971).

`The Advantage of Incumbency in Congressional Elections`

[2] Andrew Gelman, Gary King, (1990).

`Estimating Incumbency Advantage without Bias`

[3] Andrew Eggers, Arthur Spirling, (2017).

`Incumbency Effects and the Strength of Party Preferences: Evidence from Multiparty Elections in the United Kingdom`

[4] Hal Singer, Kevin Caves, (2017).

`When Can an Omitted Variable Invalidate a Regression?`

[5] Principal Fish Website, (2021).

`https://principalfish.co.uk/`

[6] Jason T. Newsom, (2020).

`Nested Models, Model Modifications, and Correlated Errors`

[7] Martin Baxter - data source, (2021).

`https://www.electoralcalculus.co.uk/flatfile.html`

[8] Trevor Hastie, Robert Tibshirani, Jerome Friedman, (2008).

`The Elements of Statistical Learning, Second Edition - Chapter 6`

[9] Guido Imbens , Karthik Kalyanaraman, (2011).

`Optimal Bandwidth Choice for the Regression Discontinuity Estimator`

[10] Yen-Chi Chen, (2017).

`Introduction to Resampling Methods Course at Washington University`

[11] Jianqing Fan, Irene Gijbels, (1996).

`A Study of Variable Bandwidth Selection for Local Polynomial Regression`

[12] Jennifer Hill, Andrew Gelman, (2007).

`Data Analysis Using Regression and Multilevel/Hierarchical Models, p.169-170`