A further investigation of the relative effects of tactical voting in the UK 2017 general election

Patrick Bettington

Abstract

A previous paper had suggested a significant impact of "anti-Brexit" voting on the 2017 general election. However, the study only in-fact showed a correlation with the referendum results. This paper will look at heterogeniety, variation in effects between "two-party" and "open" constituency sub-groups, using interaction terms to see how the correlation with referendum results differs between sub-groups, as well as testing it against a third model based on demographic characteristics.

1 Background

In a report published by Best for Britain (Lebrecht 2017) a few months after the UK 2017 general election, the author hypothesises that tactical voting based on a desire for the UK to leave or remain in the European Union, had a significant impact on the results of the election. She continues to show a statistically significant correlation between the 2017 election results and the leave / remain vote in each constituency. However, throughout the paper the null hypothesis used is simply that there is no correlation between these two things. Lebrecht does make efforts to consider the conditions in which tactical voting is likely to be more prominent by running the model in narrowed sets of constituencies where tactical voting would logically seem more effective, but again only against the null hypothesis that there is no correlation within these constituencies rather than testing against constituencies where tactical voting is logically less effective. As such the paper does not do well at answering the question it set out to.

Furthermore, this hypothesis goes against the narrative largely reported at the time that there was no meaningful difference between the Labour and Conservative policy on leaving the EU and the fact that the Liberal Democrats had positioned themselves as the only party trying to stop the process. Whilst it does not contradict, it does not recognise the results of a widely reported IPSOS Mori report (MORI 2017) (Travis and Barr 2017) which shows the biggest factors in the swing towards Labour as the youth vote, ABC1 voters and BAME voters.

By studying effect heterogeneity between constituency types, rather than the split sample regressions used in the original paper, I will test where the correlation between voting behaviour in the referendum and the election is most significant and try to draw better conclusions as to the effect of tactical voting.

In this paper I will attempt to go beyond the question of whether or not there is a correlation between Leave/Remain preference and Party preference, but to ask if tactical voting on these lines has had a demonstrable effect on the election outcome, and whether or not that model is the most useful analysis or can voting behaviour be equally or better explained using models based on more deep-rooted factors such as age distribution, class and ethnicity.

I will first analyse the process used in the original paper, noting some of the concessions made from a generalised model in order to achieve the results, challenging some of the conclusions. I will then propose a different model and compare the results to that of the original. This model will be based on the age and class of the constituency.

2 Original Model

2.1 Method

The original model that we are considering hypothesises that the number of votes gained by Labour or the Conservatives during the 2017 General elections can modelled as linear relationship between their 2015 vote count, the number of remain votes and the number of leave votes in that constituency.

- Con_17 = a1*Con_15 + b1*Remain + c1 *Leave
- Lab_17 = a2*Lab_15 + b2*Remain + c2 *Leave

Lebrecht made the decision to keep all values as actual vote numbers rather than percentage vote shares as voter turnout differed greatly between the two elections and the referendum. This decision highlights the fact that the models are actually being affected by two separate factors 1) the proportion of Leave / Remain voters 2) the change in turnout.

Lebrecht chooses not to include Scottish constituencies as "Scotland had a completely different tactical voting program" (Lebrecht 2017), then narrows the data set in two ways in order to successfully link the model to the notion of tactical voting.

The first step is to only run the model in 'marginal'[1] seats "since tactical voting would have been more widespread where a voter felt they could potentially make a difference" (Lebrecht 2017, p.2). Whilst the model is then ran and a correlation shown to be statistically significant, no attempt is made to test if splitting the sample in this way made the model more or less significant.

The second step is to run the Labour model only in seats deemed to be a two-party race[2] but to run the Conservative model in all marginal seats as "tactical voting meant that voters were encouraged to vote for the party most likely to beat the Conservatives" (Lebrecht 2017, p.2). Firstly, this is not entirely true, a number of tactical voting groups were set up in response to this election, More United supported pro-EU Conservative[3] candidates in some seats whilst Open Britain actively campaigned to unseat one Labour MP[4]. Secondly, again no attempt is made to show that this narrowing improves the model.

2.2 Data

Whilst election results at a constituency level are easily obtained from parliamentary websites, (Hawkins et al. 2015) (Baker et al. 2017) only some local authorities chose to release the results of the referendum, and these did not always align with parliamentary constituencies. The referendum results used in the model are estimates based on a regression model that uses a number of demographic indicators, linked to the number of signatures of a petition calling for a second referendum for which data is available at a constituency level. This model is shown to be reliable by checking it against the constituencies for which referendum results are available. (Hanretty 2017)

2.3 Results

Recreating these regression models gives the results shown in tables 1 & 2.

The high R^2 values suggest that this is an effective model. The predictor co-efficients show that in 2017 the overall Conservative vote was equal to 78% of their 2015 vote, lowered by 4% of the remain vote count in the constituency, but increased by 30% of the leave vote count. The appeal of this model is that the results seem to be easily interpretable as what percentage of certain groups voted for Conservatives, but actually Lebrecht alludes to the underlying confusion of this interpretation when she says that of the 30% attributed to the leave votes "many of these will have been 2015 Conservative voters too" (Lebrecht 2017, p.2). Saying that the occurrence of one leave vote correlated to an extra 0.304 Conservative votes

[1] Defined as any seat with a majority of less than 10,000 votes in 2015.
[2] Defined as any seat where in 2015 Labour and/or Conservative were the only party to receive more than 10,000 votes.
[3] http://www.moreunited.uk/mps
[4] https://www.open-britain.co.uk/key_seat_list_2017

is not the same as saying that 30% of those who voted to leave went on to vote Conservative.

This problem in interpretation is highlighted by looking at the results for the Labour model. It would be nonsensical to state that 109% of those voting for Labour in the 2015 election voted for them in the 2017 election. However, before considering the effects of the referendum the 2017 Labour vote count was a 109% of the 2015 vote, it was then increased by a further 24% of the remain vote, but there was no statistically significant correlation to the leave vote.

Table 1: Summary of linear regression for Conservative voting behaviour with in 'marginal' constituencies

	Estimate	Std. Error	T value	Pr(>|t|)
con_15	0.783	0.020	39.61	<0.0001
Remain	-0.045	0.012	-3.764	0.0002
Leave	0.304	0.010	31.879	<0.0001
	$R^2 = 0.996$		$R^2_{adj} = 0.996$	

Table 2: Summary of linear regression for Labour voting behaviour in constituencies that are both 'marginal' and 'two-party'.

	Estimate	Std. Error	T value	Pr(>|t|)
lab_15	1.089	0.050	21.757	<0.0001
Remain	0.238	0.027	8.857	<0.0001
Leave	-0.025	0.020	-1.278	0.203
	$R^2 = 0.993$		$R^2_{adj} = 0.993$	

3 Single Analytic Model

3.1 Method

In order to study the effect of tactical voting and not just to show a correlation between leave/remain and Conservative/Labour voting behaviours, I will look at the effect of heterogeneity between two different types of constituencies by adding a binary variable indicating whether or not a constituency is 'two-party'. It has been shown that single analytical models, moderated multiple regressions (MMR) are more powerful than split-sample regressions for detecting effect heterogeneity (Stone-romero 1994) so the model will be ran in both types of constituencies at once. As well as the interaction term the main effect will be left in the model and the continuous variables will be centred (Kontopantelis 2018). As already highlighted the original model does not in fact produce perfectly interpretable results and as we are more interested in the effect between constituencies the data will be scaled by dividing by two standard deviations, such that the effect of the binary variable can be directly compared to those of the continuous variables (Gelman 2008). Meta-analysis showed that there was in some cases a significant interaction

with a 'was marginal' term, but there is no logical assumption to be made about how being a marginal seat or not would affect the justification for tactical voting. The only relevant question is whether or not being seemingly closed out of a race is preventing voters from voting for the third party.

3.2 Results

Initial diagnostics led to the removal of two constituencies from the following analysis. Buckingham is represented by the speaker of the house and as such, dictated by tradition, is uncontested during general elections by major opposition parties. Bristol West was removed because it had such a high Labour vote, with 65% of the total vote, it visibly skewed all models.

3.2.1 Conservative Model

These results confirm that the most significant predictors were the 2015 vote count and the leave vote count. Whilst the main effect of being a two-party constituency is significant, indicating a modest reduction in Conservative votes in seats considered open to more than the two main parties, the fact that the interaction terms are not significant suggest that the effect of remain or leave votes was not greater or lesser in two party seats than open seats. This in turns suggests that there was no statistically significant effect of tactical voting on the Conservative vote count.

3.2.2 Labour Model

In the Labour model, once again the primary factors are the 2015 vote and the remain vote, as before the leave vote correlation is statistically insignificant.

However, in the Labour model there is a large and significant interaction between the remain and two-party variables. This indicates that in two-party seats where the only viable option was Labour or Conservative, the remain vote had a much larger effect on the Labour vote. The associated conclusion is that in seats where a third party was seen as a viable option, the remain vote had a lesser impact.

3.3 Discussion

The results of the Labour model are consistent with Lebrecht's narrative that in seats where tactical voting would logically lead to voting for Labour, the effect of remain vote is stronger, and in seats where the same logic would lead to voting for a 3rd party there was a lesser effect.

However, what this is really showing is that seats considered 'open' in the 2015 election show a weaker correlation between a high remain count and a swing to Labour in the 2017 election. Another explanatory narrative for this trend could be that open seats, with more Liberal Democrat voters saw a higher remain vote count, and a lesser swing towards labour than others.

It is worth noting at this point, as is shown in more detail in section 5, and Figure 3, that whilst on first glance the residuals for this model seem normally distributed, a Shapiro-Wilk test shows some evidence they are not from a normally distributed population and a QQ plot shows 'heavy tails' for the Labour model. This suggests that that some results of the model are more extreme that than a linear regression assumes, that the model does not hold completely true to the central limit theorem, and that there may be an important predictor missing from the model.

Table 3: Summary of the regression model of Conservative voting behaviour including two-party interaction terms

	Estimate	Std. Error	t value	Pr(>|t|)
z.con_15	0.826	0.024	34.972	<0.0001
z.remain	-0.045	0.025	-1.789	0.0749
z.leave	0.354	0.023	15.485	<0.0001
c.was_two_party_15	-0.08	0.02	-4.08	<0.0001
z.remain:c.was_two_party_15	-0.054	0.037	-1.446	0.1494
z.leave:c.was_two_party_15	-0.024	0.036	-0.677	0.4992
	$R^2 = 0.946$	$R^2_{adj} = 0.944$		

Table 4: Summary of the regression model of Labour voting behaviour including two-party interaction terms

	Estimate	Std. Error	t value	Pr(>|t|)
z.lab_15	0.842	0.028	29.691	<0.0001
z.remain	0.115	0.026	4.482	<0.0001
z.leave	-0.081	0.025	-3.253	0.0013
c.was_two_party_15	0.083	0.032	2.596	0.01
z.remain:c.was_two_party_15	0.23	0.054	4.254	<0.0001
z.leave:c.was_two_party_15	-0.068	0.05	-1.36	0.1751
	$R^2 = 0.894$	$R^2_{adj} = 0.891$		

4 Alternative Model

4.1 Method

An alternative hypothesis to explain the drastic change in voting patterns between 2015 & 2017 is simply that the combined policies of the Conservative government led to a significant number of people voting for Labour. As reported by IPSOSmori (MORI 2017), the swing towards Labour was particularly prevalent in those voters under the age of 45 and of A, B and C1 socio-economic classification.

From here we arrive at the following linear relationships:

- con_17 = a1*con_15 + b1* X0_44_pc+ c1*ABC1_pc
- Lab_17 = a2*Lab_15 + b2* X0_44_pc+ c2*ABC1_pc

I followed a similar methodology used with the previous model, but this time running the model in 'safe' seats as well as marginal seats as there is no logical way that age or class could be determined by whether or not a seat is marginal.

4.2 Data

Current age profile (2013) and socio-economic classifications (SEC) (2011) were both estimated by taking the most recent census data from 2011. These values are presented as a percentage of the total population of a constituency.

4.3 Results

4.3.1 Conservative Model

The Conservative model has an R^2 value that is higher than the previous model and shows a statistically significant correlation with both predictors. This model implies that the 2015 vote is the strongest predictor but that higher numbers of under 45 year-olds and of ABC1 SEC members both correlate with a statistically and practically significant reduction in the number of Conservative votes.

4.3.2 Labour Model

The Labour model also has an R^2 value that is higher than the previous model and shows a statistically significant correlation with both predictors. This model implies that the 2015 vote is again the strongest predictor but, conversely, that higher numbers of under 45 year-olds and of ABC1 SEC members both correlate with a statistically and practically significant increase in the number of Labour votes.

4.4 Discussion

This alternative model shows that an as strong and even greater correlation can be shown between voting behaviour and demographic characteristics of a constituency as with leave/remain preference.

It is interesting to note that in both cases the negative impact of the predictors on the Conservative count was greater in magnitude that the corresponding positive effect on the Labour count, and that once again the Labour model is not as effective as the Conservative one.

Table 5: Summary of the regression model of Conservative behaviour using census data predictors

	Estimate	Std. Error	t value	Pr(>|t|)
con_15	0.996	0.014	72.11	<0.0001
X0_44_pc	-0.180	0.009	-19.71	<0.0001
ABC1_pc	-0.144	0.013	-11.04	<0.0001
	$R^2 = 0.992$	$R^2_{adj} = 0.992$		

Table 6: Summary of the regression model of Labour behaviour using census data predictors

	Estimate	Std. Error	t value	Pr(>|t|)
lab_15	0.931	0.016	58.235	<0.0001
X0_44_pc	0.103	0.014	7.131	<0.0001
ABC1_pc	0.071	0.013	5.379	<0.0001
	$R^2 = 0.929$	$R^2_{adj} = 0.929$		

5 Diagnostics

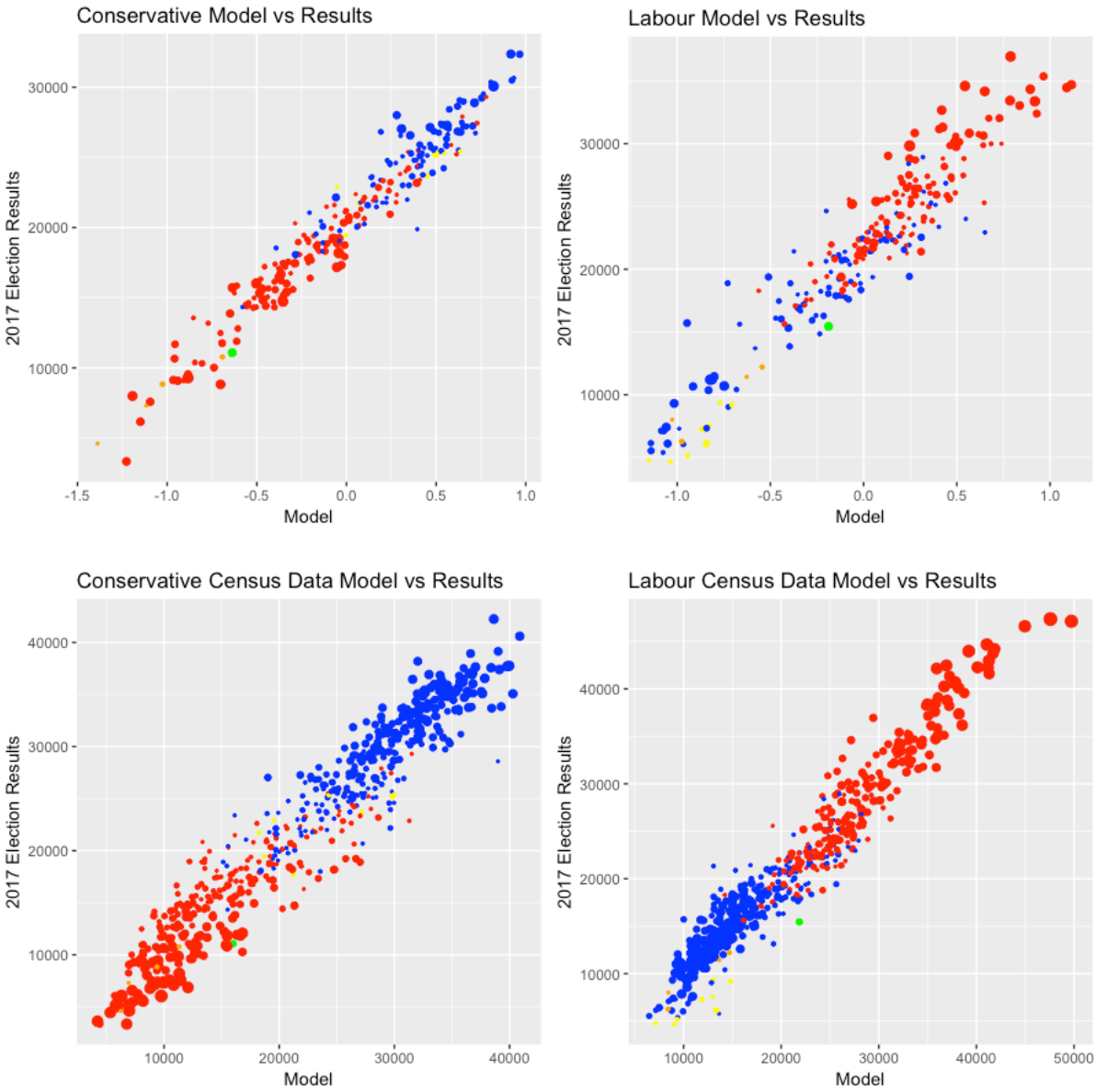
The use of linear regression models is based on 5 assumptions (Gelman & Hill 2007):

- Validity
- Additivity & Linearity
- Independence of errors
- Equal Variance of Errors
- Normal distribution of errors

5.1 Additivity & Linearity

If we plot the models' predicted outcome against the actual results of the general election, as we have seen earlier the Conservative model is a better predictor, but both show a high degree of linearity, see Figure 1.

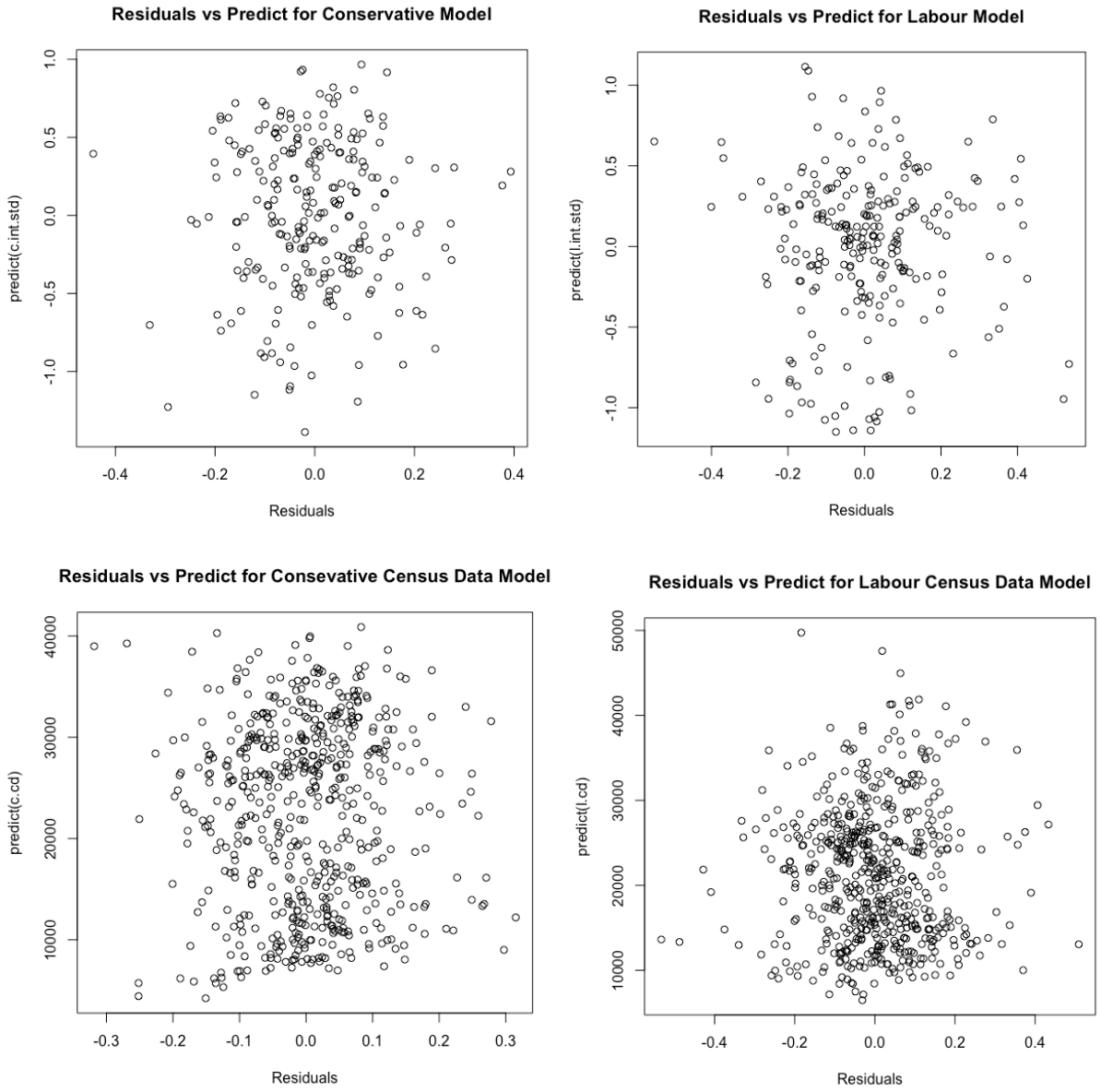
Figure 1:Predicted Results vs Observed results



5.2 Independence & Variance of Errors

Looking at plots of the residuals against the predicted outcome, figure 2, shows both independence and homescedasticity. The errors are spread evenly along the Y-axis, showing the are independent of the number of votes predicted and they are distributed evenly across the X-axis, showing that they are of equal variance for different predicted number of votes.

Figure 2: Residuals vs Predicted results



5.3 Normal Distribution of Errors

A Shapiro-Wilk test for each model, Table 7, shows that for the conservative models the null hypothesis is rejected and so the errors are normally distributed, but in the Labour model, depending on a chosen alpha level, they are not. However, figure 3, QQ plots of the residuals and Figure 4, histograms of the residuals indicate that in the central regions at least the residuals approximate normality. I interpret this as an indication that in the Labour models there is possibly another predictor missing, that the model is not as accurate as it could be, although with large sample sizes it is expected that residuals will stray from a normal distribution.

Table 7: Shapiro-Wilk normality test results

	W	Pr(>|t|)
c_resids	0.989	0.0523
l_resids	0.976	0.0004
c.cd_resids	0.995	0.0485
l.cd_resids	0.987	0.0006

Figure 3: QQ Plots of Residuals

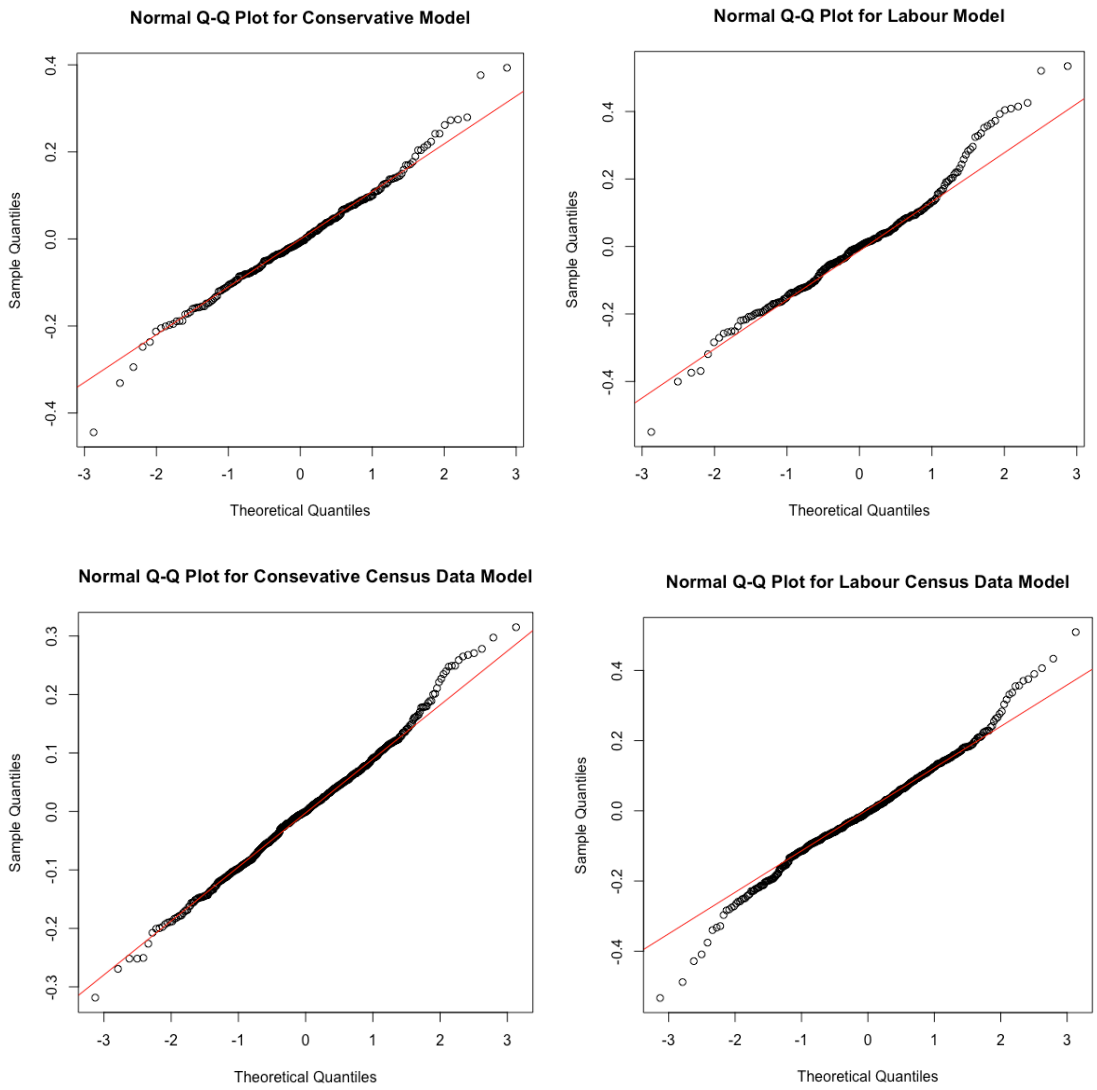
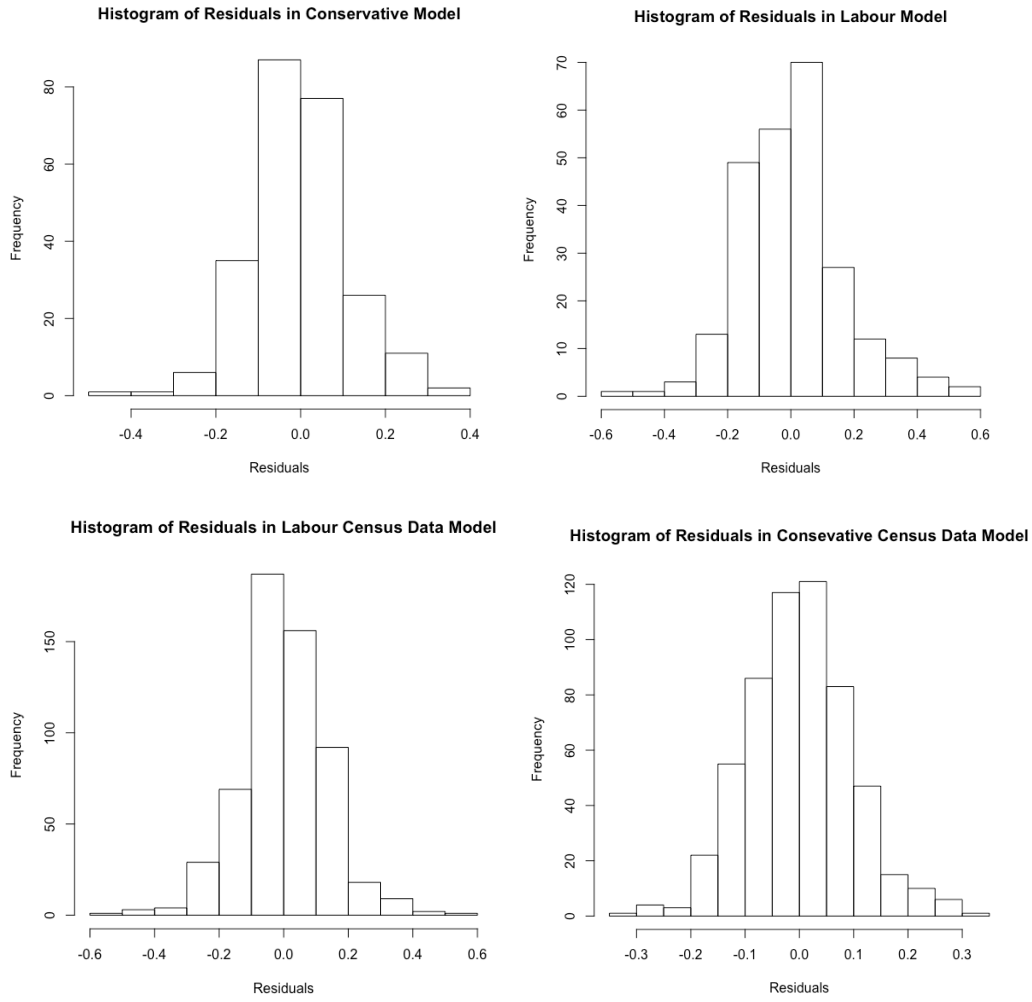


Figure 4: Histograms of residuals



6 Discussion

I have added to the findings of Lebrecht by testing the correlation against a more pertinent null hypothesis, and in doing so further supported the original narrative that in the Labour model there could well be a significant impact of tactical voting, but shown that this impact cannot be demonstrated within Conservative voting figures. However, I have also displayed that the same level of correlation can equally be found with demographic characteristics.

Especially in a such a complex field as voting behaviour, when a huge number of variables feed into what appears to be a very simplistic outcome, it is not sufficient to simply show correlation as opposed to a lack of correlation. It is necessary to test for other correlations which could equally explain the outcome. In this case it is clear that at least one other model can be as effective as the original model proposed, and in fact further analysis not detailed in this report due to word count constraints, and because it would not add anything new to the argument, suggests that a whole range of predictors, race, gender, increase in turnout, could equally be used to inform the model in finding the change between 2015 and 2017 general election results, with a high degree of success. Being unable to confidently progress from correlation to causation is a clear limitation of this model and with this in mind I believe it to be potential dangerous to make the claim that tactical voting had a significant impact on the election results.

7 References

Baker, Carl, Lukas Audickas, Alex Bate, Richard Cracknell, Vyara Apostolova, Noel Dempsey, Roderick McInnes, Tom Rutherford, and Elise Uberoi. 2017. “General Election 2017: Full Results and Analysis,” September. Commons Briefing papers. http://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7979.

Gelman, Andrew. 2008. “Scaling Regression Inputs by Dividing by Two Standard Deviations.” Statistics in Medicine 27 (15). John Wiley & Sons, Ltd.: 2865–73. doi:10.1002/sim.3107.

Hanretty, Chris. 2017. “Areal Interpolation and the UK's Referendum on EU Membership.” Journal of Elections, Public Opinion and Parties 27 (4). Taylor & Francis: 466–83. doi:10.1080/17457289.2017.1287081.

Hawkins, Oliver, Richard Keen, Nambassa Nakatudde, Steven Ayres, Carl Baker, Rachael Harker, Paul Bolton, Neil Johnston, and Richard Cracknell. 2015. “General Election 2015,” July. Commons Briefing papers. http://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7186.

Kontopantelis, E. 2018. “Investigating Heterogeneity of Effects and Associations Using Interaction Terms.” Edited by M Sperrin. Journal of Clinical Epidemiology 93: 79.

Lebrecht, Abigail. 2017. “The Impact of the Ant-Brexit Tactical Vote in 2017.” https://bestforbritain.org/static/best/pdf/Tactical_Voting_2017.pdf.

MORI, Ipsos. 2017. “How Britain Voted in the 2017 Election,” June. https://www.ipsos.com/ipsos-mori/en-uk/how-britain-voted-2017-election.

Stone-romero, Eugene F. 1994. “Relative Power of Moderated Multiple Regression and the Comparison of Subgroup Correlation Coefficients for Detecting Moderating Effects..” Edited by Lance E Anderson. Journal of Applied Psychology 79 (3): 354.

Travis, Alan, and Caelainn Barr. 2017. “Youthquake Behind Labour Election Surge Divides Generations,” June. https://www.theguardian.com/politics/2017/jun/20/youthquake-behind-labour-election-surge-divides-generations.

2011. “Census 2011 - NS-Sec.” Office for National Statistics. http://www.nomisweb.co.uk/census/2011/ks611ew.

2013. “Census 2011 - Age Structure.” Office for National Statistics. http://www.nomisweb.co.uk/census/2011/ks102ew.