

# Evaluating our 2015 General Election Forecast

*Chris Hanretty, Ben Lauderdale & Nick Vivyan*

*2015-05-21 (v1.0)*

## Introduction

Election night 2015 was one of high drama. The dust has now settled and we have had time to assess our forecasts in light of the results and see where we went particularly wrong. Immediately after the election we posted some quick analysis of some aspects of our forecasting performance ([see here](#), [and here](#)). This is our general evaluation of how the model performed. Below, we discuss the disappointing aggregate performance of our forecasts before turning to look in more detail at where exactly these problems came from in terms of the different components of the forecasting model.

Throughout, we'll focus on seats in England, Wales and Scotland only. Our model for Northern Ireland was both different in nature, and fairly rudimentary, due to the relative lack of polling data available there. We will also assume that you have read our [pre-election description of our model](#).

How well did our forecast do? It is safe to say “not as well as we would have liked”. Immediately after the election we went searching for a single big problem with our model, however as we have further examined the evidence it turns out there is not a single culprit. Instead, there are a couple of smaller problems with our model that interacted with two different ways that the polls were off target. Unfortunately, these errors all pushed in the same direction: underestimating Conservative seats.

The Conservatives emerged from the General Election with a slim majority, an outcome which our forecast estimated was extremely unlikely. The eventual seat totals for the Conservatives (330 excluding the Speaker), Labour (232), and the Liberal Democrats (8) were all outside the 90% prediction intervals of our final pre-election forecast, which predicted 278 (90% prediction interval 252-305) Conservative, 267 (240-293) Labour, and 27 (21-33) Liberal Democrat seats.

The eventual seat totals for the SNP (56), Plaid Cymru (3), the Greens (1), and UKIP (1) were all closer to our central expectations and within our 90% prediction intervals: 53 (47-57) SNP, 4 (2-6) Plaid Cymru, 1 (0-1) Green, and 1 (0-2) UKIP. Nevertheless, a total absolute seat error of 110 (found by summing the absolute difference between our predicted seat totals and actual seat totals across parties) indicates that our forecasts were inaccurate overall in predicting the outcome of the election. This level of error was substantially larger than [that implied by the advertised uncertainty from the model](#).

We were by no means alone in the inaccuracy of our point estimates. Other poll-based forecasting models also suffered from the systematic error in the pre-election national polls and under-predicted the Conservative seat totals, many to a greater extent than did our model. Thus, while we would of course have preferred to have better point predictions, as forecasters we are more concerned with the fact that we underspecified the level of prediction uncertainty.

We can break our forecasting model down into three components: a ‘national’ component, aimed at forecasting GB party vote shares based on the national polls and historical evidence concerning the relationship between polls and election outcomes; a ‘constituency’ component, aimed at estimating variation in support for each party across seats based on constituency polls, YouGov microdata, and other demographic characteristics; and a ‘reconciliation’ component, which takes our estimates of current constituency support and reconciles these to the forecast national vote shares, yielding forecasts for each constituency. We will look at each of these in turn.

## Assessment of the National Model

Some of our problems arose from the national component of our model. Our national vote forecasts were based on the national polls, and these national polls proved to be systematically wrong. Our job as forecasters is to take into account the possibility that the polls are systematically wrong. We *did* attempt to do this in two different ways – by ‘correcting’ the polls-based estimates and by adding uncertainty based on the degree to which polling data has deviated from election results in past elections. Unfortunately, our corrections of the polls helped with some parties and hurt with others, and the amount of uncertainty we added was probably inadequate. In this section we discuss some limitations of how we did this, and why they combined with the national poll miss to leave us little margin for error when it came to translating the national vote shares into seats.

The first two columns in the table below show the actual GB vote share received by each party and the final pre-election estimates of support we calculated by pooling all of the polls. Whereas the final polls put the Conservatives and Labour neck-and-neck on 33.7%, the Conservatives ended up actually winning 37.7% of the vote to Labour’s 31.2%. Crucial here is that the errors in the polling estimates of Conservative and Labour support went in different directions, yielding a large error of 6.5% in the estimated Conservative lead over Labour.

Table 1: Comparing polled and forecast GB vote share to the actual result

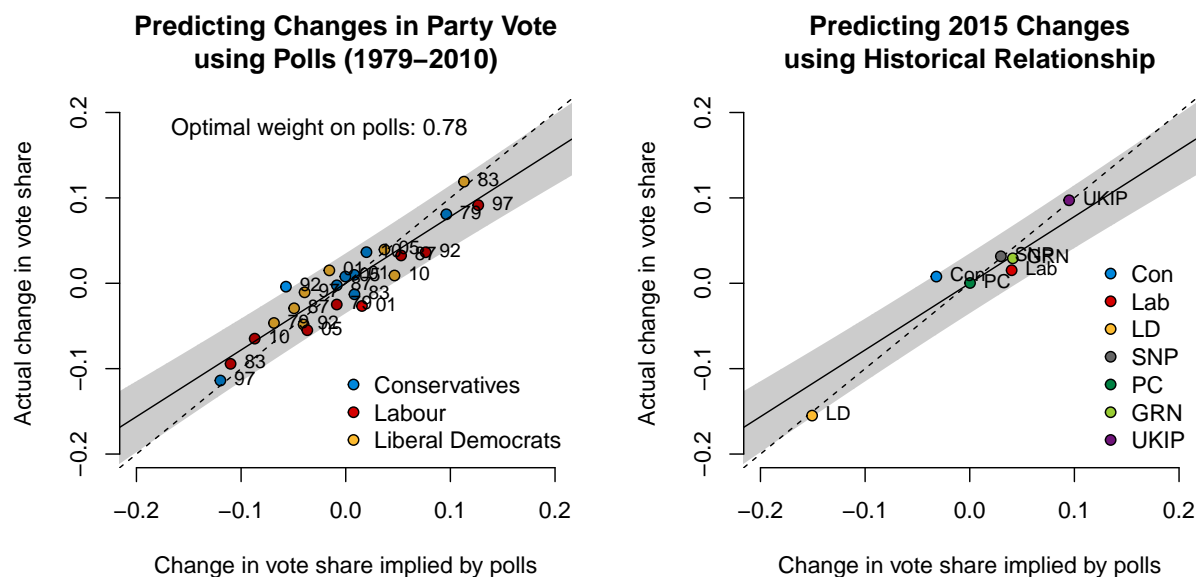
	<b>2010</b>	<b>Polls</b>	<b>Forecast</b>	<b>Result</b>	<b>Poll error</b>	<b>Forecast error</b>
Con	36.9	33.7	34.4	37.7	4.0	3.3
Lab	29.7	33.7	32.8	31.2	-2.5	-1.6
LD	23.6	8.5	11.7	8.1	-0.4	-3.7
SNP	1.7	4.7	4.0	4.9	0.2	0.8
PC	0.6	0.6	0.6	0.6	0.0	0.0
GRN	0.9	5.0	4.1	3.8	-1.2	-0.3
UKIP	3.2	12.7	10.6	12.9	0.2	2.3
Oth	3.6	1.2	1.7	0.9	-0.3	-0.8

### Correction to the GB Polls

First, based on analysis of the relationship between (pooled) polls and eventual election outcomes going back to the 1979 general election, we built a ‘correction’ of the polls into our forecast. Our forecast of the GB vote share of each party was a weighted average of that party’s 2010 GB vote share and the average of the polls, with the weight on the latter getting larger as election day approached. On election day, the weight on the polls was 0.78, so the implied “swingback” from the polls to the 2010 result was 22%. You can see the effect of this by comparing the second and third columns of the above table, or by looking at the difference between the diagonal (dashed line) and regression (solid line) in the following figure. Our forecast had the Conservatives doing slightly better than the final polls suggested, and Labour doing somewhat worse. This correction was nowhere near large enough to account for the national polling miss, with the consequence that we still under-estimated Conservative performance by 3.3 points and over-estimated Labour by 1.6 points. Still, it pushed the Conservative - Labour margin in the right direction.

At the same time though, this correction made our error for the Liberal Democrats worse. We over-estimated the performance of the Liberal Democrats, assuming that their support would be somewhat closer to 2010 than the polls suggested. Processed through the rest of our model, this made us unduly skeptical about Conservative prospects of for winning seats from the Liberal Democrats, when in reality it was these gains that proved critical to securing their majority. This same correction had mixed effects with respect to the other parties as well, improving our GB vote share estimates for the Greens, but making our SNP and UKIP estimates worse. For these parties, the consequences of this correction for the seat forecasts were very limited.

Our assumption that the correction applied to the polls should be the same across parties was a problem. One alternative approach would have been to estimate party-specific poll corrections for the Conservatives, Labour and Liberal Democrats. We were wary of doing this prior to the election because of the rise of the SNP and UKIP, and our consequent need to extrapolate from a historical polling record that was limited to the traditional three major parties. As such, and given that it was not clear that observed historical differences between the three major parties were actually robust patterns rather than flukes of the small number of elections for which we had data, we elected to use a symmetric model. A symmetric model, where the poll corrections were the same for every party, meant that we could apply the same correction to all parties. But of course easier to model does not mean correct. [Steve Fisher at electionsetc](#) applied an asymmetric approach which estimated party-specific corrections, and it helped him achieve more accurate forecasts of 2015 party vote shares than we managed.



## Uncertainty in the GB Estimate

As we have noted above, we did not assume that (even after the aforementioned correction) our GB vote share estimates would be spot on. We tried to estimate the level of systematic error that remained historically when using (corrected) pooled-poll estimates of party support to predict eventual election outcomes. This additional uncertainty did nothing to our central vote share projections, but rather widened the surrounding prediction intervals. The figure above depicts a shaded band for our estimates of the predictive uncertainty in applying the historical relationship between the swings implied by polls and the actual swings in elections. This is the 90% prediction interval, and – as the left panel shows – it does include about 90% (22 of 24) of the historical data points that we used in calibrating the model. The right panel overlays the 2015 results and shows that, while the polls and our forecast underpredicted the Conservatives and while our forecast over-predicted the Liberal Democrats (in contrast to the polls), our 90% prediction intervals for the national vote swings did include the observed values for all parties, if only narrowly for the Liberal Democrats and the Conservatives. This meant that the observed polling miss for the Conservatives was at the edges of what our historical data suggested was possible. The failure of the Liberal Democrats to outperform their polling was also surprising to our model, but then the Liberal Democrat loss of vote share implied by the polls was larger than anything in our calibration data, so we were extrapolating outside the historical support of the evidence.

This plot suggests that we should have been okay in terms of our prediction intervals. Unfortunately, some of this uncertainty was lost due to the mathematical details of the model. The problem was that the way in which we modelled this additional uncertainty did not directly estimate the extent to which polling errors are

negatively correlated across parties – that is, in elections when the polls have tended to under-estimate the Conservatives they have also tended to over-estimate Labour. As made clear when [we described our model before the election](#), we simply pooled the 3 parties x 8 elections of discrepancies (see left column of page 2 of preceding link) and assumed independent errors.

Now, we did capture the basic fact that (because all vote shares must sum to 100%) positive errors for one party imply negative errors for others, because we used a sum to 100% constraint in our model which induced the requisite negative correlation. Moreover, when we have looked at the election results simulated by the model, this constraint succeeded at inducing a level of negative correlation comparable to what is observed in the historical data. But this indirect approach to capturing the correlation in errors nevertheless reduced the effective uncertainty regarding differences in vote share between parties, and especially between the two largest parties (Labour and the Conservatives). This resulted in over-confidence concerning the Conservative *lead* over Labour in the popular vote, which in turn meant that we placed too little probability on scenarios where either party gained many more votes – and therefore many more seats – than the other.

## Improving GB Vote Share Estimates

The good news here is that in the future we can make changes to address these errors. The correlation in polling errors across parties can be modelled more explicitly when estimating uncertainty and the correction model can either be modified or dropped. [Matt Singh has argued](#) that other sources of information such as local election results and leader popularity may provide useful predictions as to how the polls are going to be wrong. Given the rise of parties such as the SNP and UKIP and uncertainty regarding their prospects in future elections, as well as the fact that the pollsters will be trying to fix their past failures, another option going forward may be to focus on *how much* the polls are likely to be wrong and stop trying to guess *exactly how* the polls are going to be wrong. We did not do particularly well at either in this election, but we are a lot clearer on how to fix the former problem than the latter one.

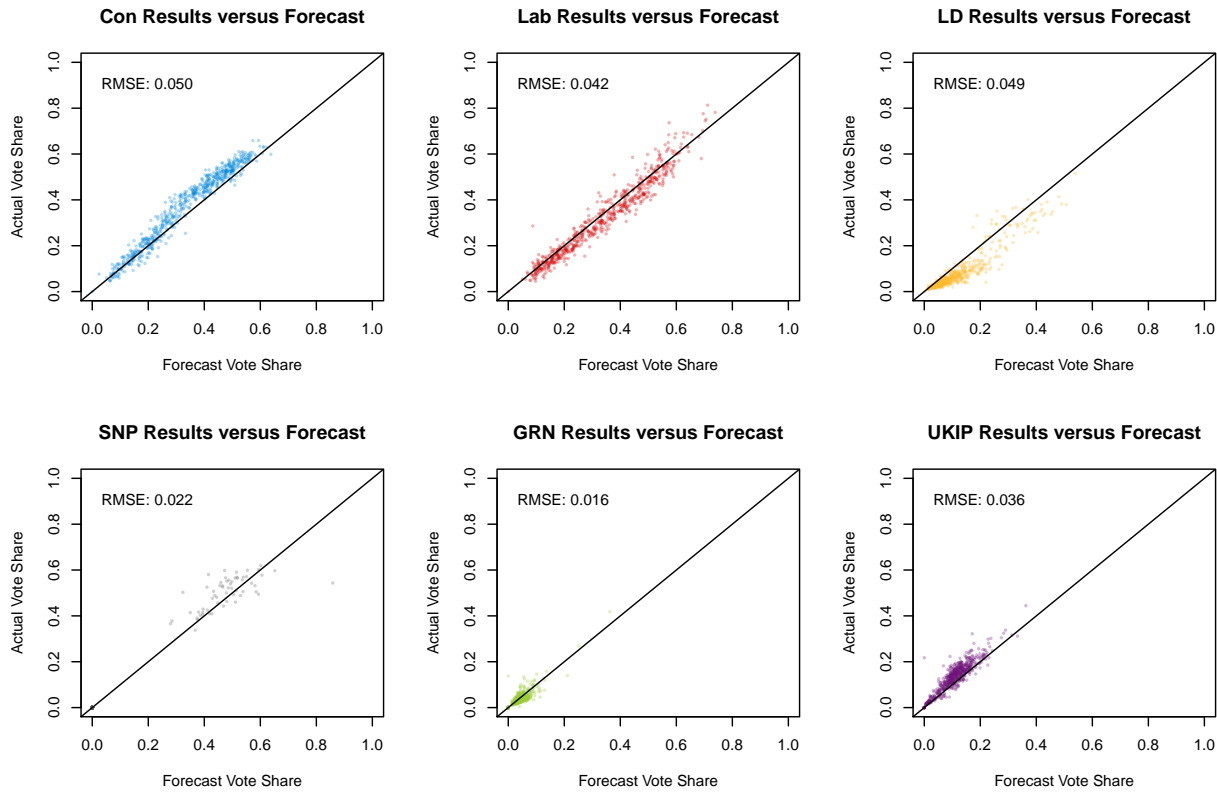
## Assessment of the Constituency Model

The fact that the national result was at the edge of our prediction intervals meant that any further discrepancies between our model and the electoral results that led us to underpredict Conservative seats were going to lead to a seat result that was well outside our prediction intervals, rather than at the edge of those intervals. While we will show below that the constituency component of our model did a decent job of capturing variation in party support across seats, the Conservatives outperformed our forecasts most in the marginal constituencies where that translated most efficiently into additional seats. It was this combination of Conservative overperformance of the national polls as well as in the marginals *relative to the national polls* that led to the failure of our model, because while we had (imperfectly) modelled the possibility of a national polling miss, our model had no way to capture the possibility that the constituency-polls would miss Conservative strength in the marginals in the way that they did. Many of the constituency polls we were using came from well in advance of the election, so this is not necessarily a problem with the polls themselves, but rather the limitations of how we interpreted them.

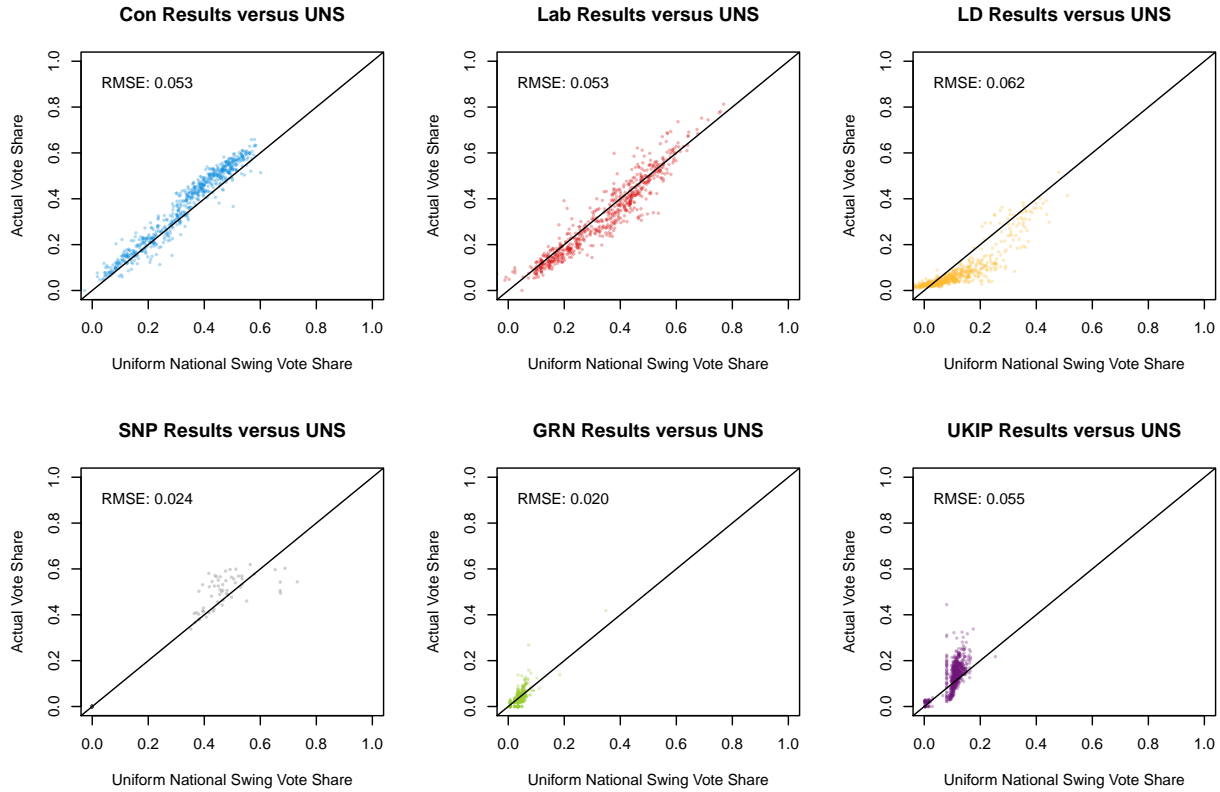
## Constituency Vote Share Predictions

The fact that our national vote share forecasts were off meant that our *constituency* vote share forecast were going to be off in terms of average levels. The graphs below illustrates this. For each of six parties, we plot our forecast vote share for each seat along the x-axis, and the actual vote share along the y-axis. Take the Labour plot as an example. The dots are consistently below the forty-five degree line because Labour under-performed nationally relative to our forecasts. But the dots nevertheless follow a trend which is parallel to that forty-five degree line and, as we will show below, the degree of scatter around the trend is small compared to alternative methods like national swing models. In other words, in places where Labour were forecast to do well relative to their average performance, they did tend to do so.

This is generally the case for all parties. However for the Conservatives there is a small – and very important – kink where they outperform our forecasts more in marginal seats (30-45% vote shares) compared to other types of seat.



Our failure to pick up this Conservative overperformance in the marginals is not evidence that we did worse than ready alternatives for translating votes into seats. The constituency component of our model performs better than the seat predictions one would have generated if one had taken our final national vote share forecasts and applied the standard uniform national swing (UNS) method. So as not to be unfair to uniform swing given the huge divergence of voting patterns across different parts of Britain in 2015, we calculate England-, Wales- and Scotland-specific national swings and add these to 2010 constituency vote shares. (To back out these nation-specific vote shares from our model we assume 2010 constituency turnout levels.)



Using this type of nation-specific UNS to predict party seat totals would have yielded a total absolute seat error of 122, compared to a total absolute seat error of 110 when we model constituency level variation in party support. We can also compare the two approaches by looking at the number of individual seats each would have predicted correctly. Again, our model out-performs UNS by this measure, predicting 568 (89.9%) of 632 seats correctly compared to 561 (88.8%). Finally, as the table below shows, if we use root mean square error (RMSE) to compare the accuracy of constituency vote share predictions – rather than seat-winner predictions – our forecasts consistently outperform UNS across all parties.

Table 2: RMSE of our forecast versus UNS applied to GB and to the nations of England, Scotland and Wales separately

	Forecast	UNS (Nat)	UNS (GB)
Con	0.050	0.053	0.052
Lab	0.042	0.053	0.082
LD	0.049	0.062	0.062
SNP	0.022	0.024	0.090
GRN	0.016	0.019	0.020
UKIP	0.036	0.055	0.062

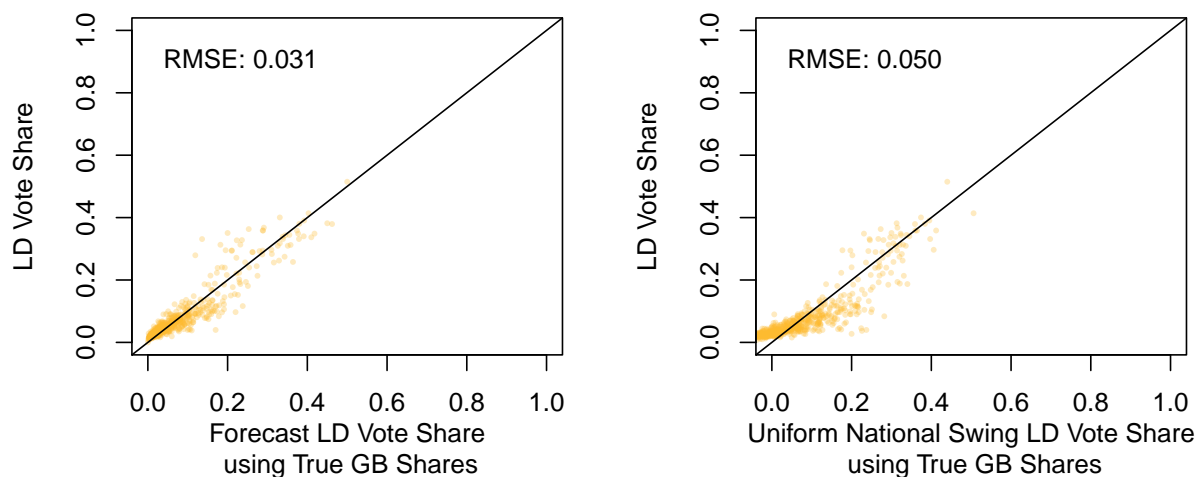
The preceding comparison between our constituency-level forecasts and UNS was based on the inaccurate national vote shares that we forecast before the election, but the patterns remain the same if instead we do the calculations based on the true national-level election results. To do so, we simply reconcile our constituency estimates to the actual national vote shares received by the main parties at the 2015 election. When we do this, the total absolute error in party seat totals falls to 40, and the model predicts a highly respectable 598 (94.6%) of 632 seats correctly. Fed with the true national vote shares, nation-specific UNS does slightly better on absolute seat error (30), but worse on individual seats correctly predicted (592 or 93.7%). Once

again, if we compare vote share predictions rather than seat predictions by RMSE, our forecasts consistently outperform UNS across all parties.

Table 3: RMSE of our forecast versus UNS (applied to GB and to the nations of England, Scotland and Wales separately), using true 2015 national vote shares as inputs

	Forecast	UNS (Nat)	UNS (GB)
Con	0.034	0.038	0.039
Lab	0.037	0.050	0.080
LD	0.031	0.050	0.052
SNP	0.022	0.023	0.090
GRN	0.015	0.018	0.019
UKIP	0.026	0.048	0.056

The two figures below illustrate how our model improved upon uniform national swing. Both figures plot true 2015 Liberal Democrat constituency vote shares on the vertical axis against estimates of vote share on the horizontal axis. In the left plot these estimates on the horizontal axis come from reconciling our model to the true 2015 national vote shares. In the right plot the estimates on the horizontal axis come from applying nation-specific uniform swing based on the true 2015 national vote shares. Comparing the two figures suggests that, while our model may not have captured everything that was going on at the constituency level, it was a substantial improvement on uniform swing, especially for a party losing so much vote share that UNS implies negative share in many individual seats.

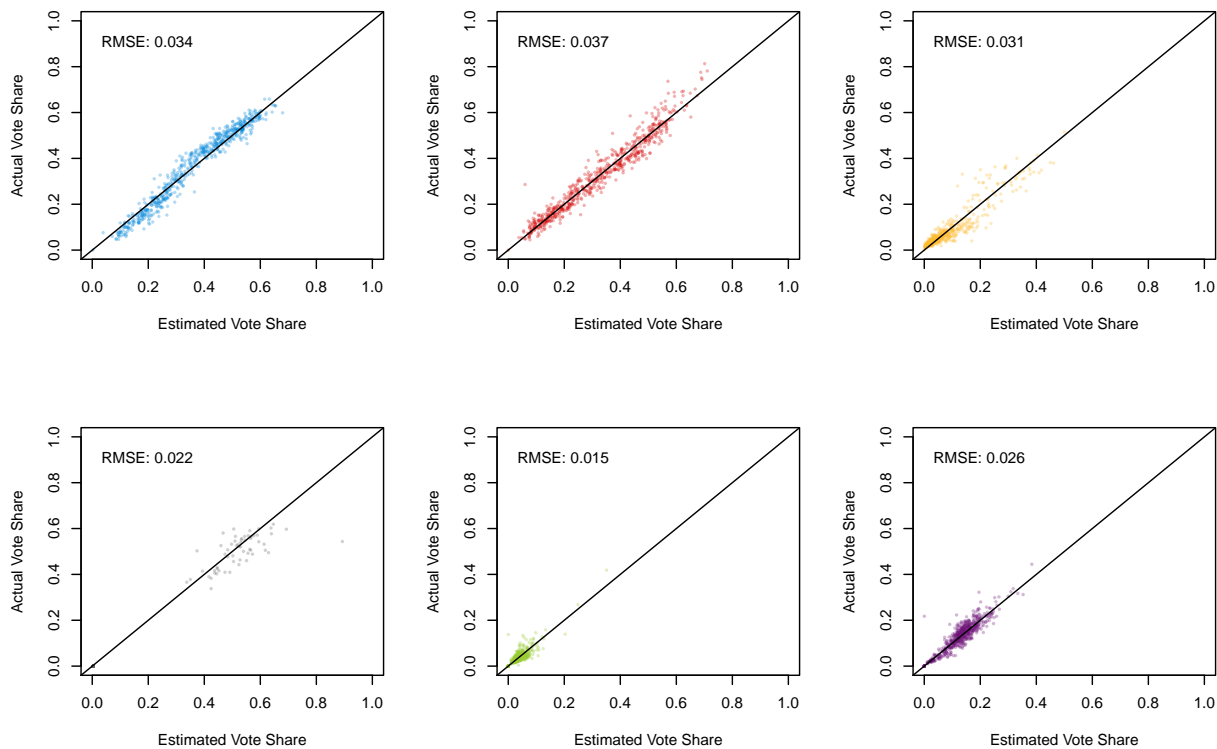


What we take from this is that we consequentially underestimated Conservative performance in the marginals, but we still improved on national swing models. Our model incorporated information from constituency-level polling, which seems to have helped our overall predictive performance.

### Conservative Overperformance in the Marginals

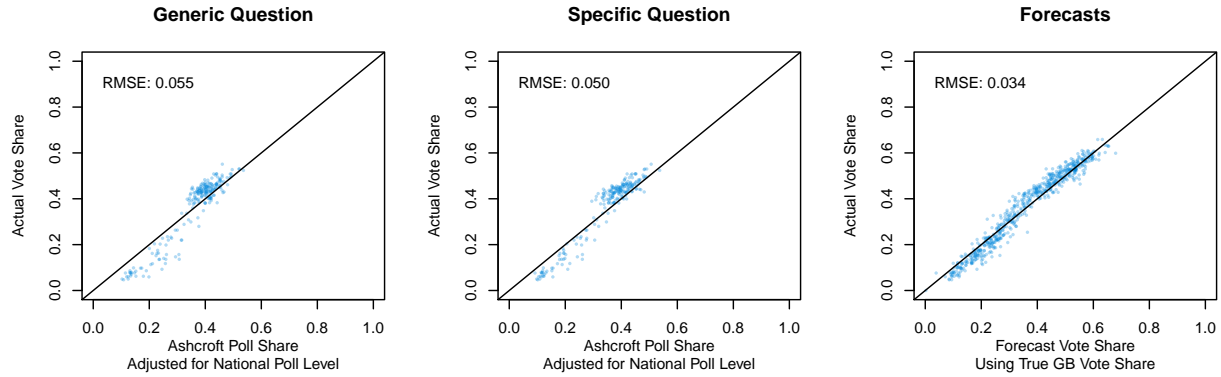
As we have seen, our constituency-level model had one notable failing, which was that the Conservatives outperformed our forecasts in the marginals, precisely where it mattered most for seat totals (as we saw

earlier, the Conservatives also outperformed UNS-based expectations in these seats). We can get a sense of how big a problem this was by looking at the seat totals that we recover in the alternate scenario we described above, where we input the correct GB vote shares and calculate what seat totals our model would then have implied. In this scenario, we estimate 310 Conservative seats, 242 Labour seats, 16 Liberal Democrat seats and 57 SNP seats. These numbers are much closer to the true outcomes, but still understate the Conservatives by 20 while over-stating Labour and the Liberal Democrats by 10 and 8, respectively. Why do we think that much of this residual difference is explained by our constituency model not picking up Conservative overperformance in the marginals? The figure below shows our estimated constituency vote shares in this scenario (where the true GB vote shares are known to the model) along the horizontal axis, versus actual results on the vertical axis. Now the dots for most parties tend to cluster around the forty-five degree line, indicating little systematic under- or over-estimation of performance relative to the result. However, for the Conservatives, the kink in the crucial marginal seats remains – in these seats, the Conservatives still systematically outperformed our model estimates, even though the model was fed with the true GB vote shares.



What led to our model under-estimating Conservative performance in key marginals? It appears that this was largely a consequence of an under-estimation of Conservative performance in the raw constituency polling data that went into the model. The figure below plots actual Conservative vote shares against Conservative support as estimated by individual constituency polls, after these polls are corrected for national trends and national polling error (by taking the poll estimate of party support and adding in the difference between the eventual GB national vote shares outcomes and the national pooled polls on the date of the constituency poll). The left-hand panel shows results based on a generic vote intention prompt, the center panel shows results using the constituency-specific vote intention prompt, the right panel shows our forecasts (reconciled to the true GB vote share). Whichever vote intention measure is used, we see the same pattern as in our forecast: actual Conservative vote shares out-performing survey-based measures of support *mainly* in the marginal seats where they were polling at 30-45%.





Ideally, our model would have been able to correct for the constituency polls understating Conservative support (relative to election results) in marginal seats. But this was not easy to do because this is the first election for which a large number of constituency polls have been carried out, so there was no good historical data allowing us to compare past constituency polling estimates with past constituency results. While we could have tried to see whether parties over- or under-performed in marginal seats relative to UNS in past elections, this is not necessarily that useful in telling us about under- or over-performance relative to constituency-specific polls. Now that the 2015 election is over, we can at least start to look at the relationship between the results of the many constituency polls published in the run-up to the election and the eventual constituency election results. This may help improve the model in the future.

## Constituency Probabilities of Victory

In a [separate evaluation report](#), we have compared our constituency-level probabilities of victory to those from the betting markets on each constituency. We find that our forecasts and the betting markets had identical performance in individual seat predictions, with a slight edge to the markets in terms of avoiding overconfidence in the results (compared by Brier score). Had our national-level model been more uncertain, this would have helped our forecasts with respect to this measure by inducing more uncertainty at the seat-level as well.

## Summary

In general, the constituency-level part of our model was an improvement on existing methods, as we had hoped. Broadly speaking, our forecast vote shares were highly correlated with the election outcomes, we outperformed UNS applied at the level of England, Wales and Scotland, and we did about as well as the betting markets (which had access to the information from our forecasts). We got some important things right, by virtue of our constituency-level model. Without making special exceptions in the model for minor parties, we correctly predicted the single Green seat, the single UKIP seat, and the three Plaid Cymru seats (we had a possible PC pickup elsewhere in expectation, but no specific fourth seat predicted). We picked up the SNP surge in Westminster voting intention after the Independence referendum before [the bookies factored it into their odds](#). Our model picked up that the Labour seats-votes advantage of recent elections [was likely to be eroded](#) in this election due to the reconfiguration of support across constituencies occasioned by the rise of SNP and UKIP, which [indeed proved to be the case](#).

The Conservative over-performance in marginal seats versus uniform swing was not in the constituency polling data. This meant that our ability to predict it was limited by the lack of constituency polling in past elections to indicate that we might have this kind of problem. Since this overperformance was worth something like 20 extra seats for the Conservatives versus what the polls suggested and our model expected, it proved to be the difference between a seat result around the edge of our prediction intervals (due to the national poll miss) and a seat result substantially outside our prediction intervals.

## Performance of the Reconciliation Model

As we described in our [pre-election description of our model](#), we developed a new reconciliation/swing model based on the generalized error distribution to achieve approximately uniform swing over most vote shares, combined with not violating the zero bound on vote share. Previous attempts to respect the zero-bound using proportional swing models have yielded poor performance, because UK constituency-level swings have seldom been proportional to past vote share. While this generalized error swing model was only used for reconciliation of the constituency-level model and the national-model, it appears to have worked well. The plots from the previous section for the Liberal Democrat vote show this, as well as the fact that we correctly predict nearly 95% of seats when we apply our swing model to our constituency-level estimates using the correct national vote shares, even though the constituency-level estimates underestimated Conservative performance in the marginals.

## Conclusion

To our knowledge, there were no other attempts to provide internally consistent forecasts of vote share, uncertainty estimates for vote share, and probabilities of victory, at both the constituency-level and GB-level for the 2015 UK General election. Unfortunately, we have to conclude that our model performed poorly in terms of its central national predictions and under-estimated the level of uncertainty surrounding these predictions. As the authors of the forecast, this is immensely frustrating to us. We got many aspects of the modelling right, much of it worked as intended, but in the end what counts is how all the pieces of the modelling work together.

There are two background facts about the polling that hurt the performance of the model:

1. The national polls underestimated Conservative performance.
2. The constituency polls underestimated Conservative performance in the marginal constituencies, relative to their national performance.

In the context of the national and constituency-level polls having these issues, there were three modelling choices we have identified that hurt the performance of our forecasts:

3. Our national-level assumption that all parties would revert a bit towards their 2010 performance was not very predictive.
4. Our approach to modelling the covariance of errors between the national-level polls and the election results underestimated uncertainty.
5. Our constituency-level model did not incorporate the possibility of systematic biases varying by constituency marginality.

The national poll miss is well known to all. Had we known the correct national vote shares, we would have forecast the Conservatives with 32 more seats than we actually did. However, this is not enough to explain why our forecast failed, even with the two issues we have identified with our national-level model. The issues with the constituency-level polls explain the remaining 20 seats of underprediction with respect to the Conservative totals (our total underestimate for the Conservatives was 52 seats). Our headline seat forecast suffered from these two errors occurring in the same direction.

Noting the mechanical effect of the polls being off is not an excuse for our forecasting model falling short, but these are important numbers for understanding what went wrong. Combined with the two ways that the polls missed systematically, our modelling choices kept our forecast from doing what it aimed to do. Had only some of these issues arisen, the result would have been at the edges of our 90% prediction intervals. We were expecting the possibility of polling misses on the scale of what occurred, so if there had only been the observed national polling miss, the result would have been around the edges of our prediction intervals.

Crucially, it was the additional effect of Conservative overperformance in the marginals relative to uniform swing, which does not appear to have been captured in the constituency polls, that led to our forecasting miss. We did not take the possibility of this kind of error into account in our model. Without the national polling miss, our underprediction of the Conservatives in the marginals would not have been consequential enough to lead to a result outside our prediction intervals.

We could have gotten lucky with the polls in this election and the issues with our model would have failed to matter for the headline forecast. However, even then, the problems with our model would have been nonetheless real. It is the nature of forecasting that not all modelling errors matter every time. This time they did, and that will help us do better next time.