Forecasting the 2010 general election using aggregate local election data

Colin Rallings, Michael Thrasher and Galina Borisyuk

(LGC Elections Centre, University of Plymouth)

International Symposium on *Methods and Models for Election Forecasting: The UK General Election of 2010,* organised by the Democracy, Citizens and Elections Research Network (DCERN) in association with Institute for Social Change and Politics and the Elections, Public Opinion and Parties (EPOP) Specialist Group of the UK Political Studies Association. University of Manchester, March 19-20th 2010.

Abstract

This paper presents a revised method for estimating national vote shares using aggregate data from local government by-elections. The original method, originally developed prior to Labour's landslide victory in 1997, provided an accurate forecast for that election but subsequent general election forecasts proved less satisfactory, although the method continues to be effective when forecasting annual local election outcomes. Over the past decade the changing pattern of party competition meant that some elections were excluded from the modelling because they failed to meet the criteria that all three major parties, Conservative, Labour and Liberal Democrats, had contested both the by-election and the previous main election, normally held in May. Relaxing these criteria, although increasing the number of available cases would adversely affect the forecast, over- or under-estimating party votes. Instead, the revised method overcomes the problem of differential competition by estimating vote shares for parties that contest one but not both elections. A further innovation is the calculation of a weighted moving quarterly average which takes account of the number of days elapsed between the by-election date and the date of forecast. Using the new method we provide estimates for likely party shares for the forthcoming 2010 general election.

Introduction

Unlike most other election forecasting models, the example described here is primarily designed for another purpose - to forecast national equivalent votes at annual local elections. Judged by this criterion it has proved successful and more successful than opinion pollsters whose forecasts from respondents' answers to local voting intentions proved less than satisfactory. The model's raw material are aggregate level data obtained from local council by-election results that take place in virtually every week of the year. The theoretical basis behind our approach is that, unlike their parliamentary equivalents that generate a media circus and become a vehicle for voters' protests, council by-elections are relatively mild affairs where people behave in a normal manner; and when they don't behave normally the "errors" have a way of cancelling themselves out.

From 1995, when we began using this model for forecasting the May local elections, it immediately demonstrated its value. So much so that we used it to forecast the 1997 general election and were pleased to see that it out-performed the national polling companies (Rallings and Thrasher 1999). Perhaps sensing some serious commercial competition, but mostly we feel out of a sense of scientific curiosity, Nick Moon from National Opinion Polls (NOP) commissioned a retrospective application of the model to the 1992 general election. Crucially, (disappointing to us but money well spent for Nick Moon) the by-election model too would have forecast a narrow Labour victory instead of a Conservative lead of eight percentage points (Rallings and Thrasher 1999). Undaunted, and making sure that the 1992 result was only published in an academic journal where no-one of substance would read it, we continued to believe that the model could be used to fit the circumstances of a parliamentary rather than local election.

In 2001 we encountered what might be termed 'a snag'. In February 2001 the first cases of foot and mouth disease were reported, leading to limits on the movement of livestock and widespread disruption elsewhere. Local council by-elections were postponed where possible and from the end of February to the synchronous local/general election in June there were just 51 cases, many fewer than normal and none of which were held in the month prior to the election. Nevertheless, our forecast for The Sunday Times published on June 2 used our analysis of split-ticket voting at the synchronous 1997 elections to generate figures from the model data (Rallings and Thrasher 1998; 2001; 2003). The forecast read Labour 41% (42.0% actual), Conservative 32% (32.7%) and Liberal Democrats 20% (18.8%). Interestingly, three of the five companies conducting national polls over-estimated Labour by 3-5 points in their eve of poll surveys, continuing the pattern from the early 1990s. Following the 2001 election, where recorded turnout fell below 60%, voter apathy appeared to spread to the local parties. It was not that there were many fewer by-elections than before (encouragingly, local councillors were still resigning/retiring/dying in healthy numbers) but that the pattern of party competition started to change. Where straight three-party contests had once been

commonplace they now became less so. In some areas more candidates from minor parties began to participate, albeit in a rather haphazard fashion. More and more cases were being excluded from the by-election modelling because of the pattern of party competition at both the by-election and its May predecessor were incompatible with the task of estimating national vote shares. On May 1, 2005 our Sunday Times forecast was again adjusted on the basis of split-ticket voting (now 2001 as well as 1997 aggregate level data were available). The forecast was Labour to win a majority of 96 seats having polled 37% (36.1%), the Conservatives 34% (33.2%), and Liberal Democrats 21% (22.6%). The eve of poll findings from the polling companies were as accurate.

And so to 2010. In the following section we outline the initial by-election model before reporting on our efforts to address the increasing problems affecting the admission/exclusion of cases. Next, we introduce the revised model, applying forecasts retrospectively both to the May local electoral cycle from the 1990s onwards. Finally, we use the most recent findings to discuss the likely outcome for the 2010 general election. This remains a work in progress. We are testing whether the number and location of by-elections, to an extent affected by the wider electoral cycle, is a factor in skewing general election forecasts. We are also working with new data that records the cause of the by-election vacancy since the circumstances may affect subsequently the distribution of party support. We remain committed to the value of these data for forecasting purpose since these are, when all is said and done, "real votes in real ballot boxes" and constitute in Austin Ranney's terms the 'hardest' data political scientists can get (Ranney 1962).

The by-election model mark 1

We have been collecting local council by-election results from across Britain since the mid 1980s. Each year there is an average of 290 vacancies although this fluctuates with the broader electoral cycle. The forecast model requires information about both the by-election result and the outcome at the preceding May election for each ward (the local electoral division). . Clearly, assuming that the pattern of party competition is identical across elections it is straightforward to calculate change in vote share and swing for a given ward but forecasts, generalising from the particular, require a set of benchmark figures that are common across a range of wards. This comes in the guise of the 'national equivalent vote', (NEV) an estimate of how the country as a whole might have voted extrapolated from actual local election voting in any given year (Curtice and Payne 1991; Rallings and Thrasher 1993).

For any given ward election in May, therefore, we now know both the distribution of party support in the ward and how that compares with the country as a whole. The initial method used only by-elections which featured candidates from Conservative, Labour and Liberal Democrat parties at both the May election and the by-election. An additional caveat was that cases where votes for other parties and Independents at either the May or by-election were greater than 10% of the total vote

would be excluded from consideration. The exception to this related to cases where the intervention and support for other parties/independents was consistent across the two elections. Because of a non-uniform local electoral cycle it was important to note the particular year when a by-election ward had held its May election since this became an important part of the calculation. A worked example is provided in Table 1 while the method is described formally in the Appendix.

Table 1: Calculating the current national equivalent vote using the

November, 1996 by-election result in Ixworth ward, St Edmundsbury Council

Steps in the method	Conservative	Labour	Lib Dem
a) By-election share in ward	43.4	26.2	30.4
b) 1995 May vote in the ward	36.1	29.6	34.3
c) Change in vote share (a-b)	+7.3	-3.4	-3.9
d) 1995 NEV	25	47	23
e) Estimate of current NEV (d+c)	32.3	43.6	19.1

Of course, in a given ward it is possible that the change in a party's vote share may be greater than its NEV for a particular year, leading to a current NEV estimate that is nonsensical. However, by averaging the estimates across all by-elections over a month/three month period, any extreme results are smoothed out. In essence, therefore, between the May and the subsequent by-election the model is calculating change in each party's vote share, adding/subtracting that change to the NEV value for the relevant year and averaging across cases to estimate a current NEV for a given point in time.

The by-election model: mark 2

In 2001 there was a rather dramatic change to the pattern of party competition at by-elections that had implications for modelling. In the previous year some 68% of by-elections featured candidates from all three main parties; this dropped to 62% in 2001 with a further fall to 55% in 2002. Although the proportion of three-party contests recovered from this low point it did not match the consistent levels seen throughout the 1990s. This, combined with a retreat from three-party contests in the main May elections, especially in the English shires, led to model estimates being sourced from a declining base of data.

Our response was first to investigate thoroughly the precise structure of party competition and second to devise methods that might then compensate for missing values, permitting more cases to be used to estimate national support. A third aim was to determine the optimal time frame for averaging – it should be a trade-off between averages being responsive to new information and not being too highly sensitive to random variations. Estimates based on a relatively short run of by-elections would be up to date and reflective of the current electoral mood but might suffer from having too few cases. By contrast, using a longer time span would increase the absolute number of cases but might unduly weaken the influence of recent cases.

A starting point was to examine for every case the structure of party competition at both the by-election and the previous May election. In more than 7,000 by-elections there was three-party competition at both the May and by-election in 3,425 cases. In a further 544 by-elections a Liberal Democrat candidate, present for the May contest was missing from the subsequent by-election. In another 250 and 77 cases it was the Labour and Conservative candidate respectively that missed the by-election. Of course, the process worked in the opposite direction with by-election vacancies attracting greater party competition than had the May equivalent. For example, in 648 cases where three main parties contested a by-election the Liberal Democrats had not challenged when the main May election was fought. In a further 255 and 155 cases it was Labour and the Conservative candidates respectively that are missing from the May but present in the by-election contest. In other examples the structure of party competition was partial but stable in the sense that perhaps only two of the three parties competed at both elections. In 457 cases, for example, only Conservative and Labour challenge one another with the Liberal Democrats absent on both occasions. In a further 331 examples the two protagonists are Conservative and Liberal Democrats while 104 cases are Labour versus Liberal Democrats only. The structure of party competition, therefore, dictated differing responses in devising new procedures designed to include more cases for devising model estimates.

The first examples consider cases where the pattern of party competition is more extensive at the main elections in May than it is for the subsequent by-election. Local voters in May could select from Conservative, Labour and Liberal Democrat candidates but the Liberal Democrats (in the example shown) decide to stand aside from the by-election contest (Example 1, Appendix). Previously, the model ignored such cases but a different method is now employed, providing a notional by-election share for the Liberal Democrats which is equal to the minimum share value the party achieved across the relevant local authority at the May election. The theoretical basis for this assumption is that the likely explanation for the Liberal Democrats to withdraw from the by-election is the expectation of receiving a low level of votes. The votes for the parties that did contest the by-election are then adjusted to take account of the estimated vote for the absent party. It is important that the process of normalising votes in this way does not distort the actual result, for example, transforming a positive change in vote share for one of the parties that did contest into a negative one. In order to prevent this the share change for a party that both contests and increases its share

is never allowed to fall below zero. Following these adjustments the by-election result is then treated in the usual way for the purposes of estimating national vote shares.

There are other occasions when the extent of party competition is further fragmented; three main parties contest the May election but now **two** fail to contest the by-election. For example, both Labour and Liberal Democrats do not present by-election candidates to challenge the Conservatives (Example 2 in Appendix). The problem here is that the change in Conservative vote share from May is likely to be inflated because of the absence of two of its competitors. In such cases we proceed with the method described above, now estimating by-election shares for both Labour and Liberal Democrats based on the minimum values across the local authority at the previous May election. Following this procedure the Conservative by-election share is recalculated by subtracting from its actual share the estimated shares for both Labour and Liberal Democrats. Again, if the actual Conservative by-election share is an increase from its share in May but the process of estimating shares for the two missing parties transforms that to a decrease then change is limited to zero. Where the Conservative by-election share does actually decrease, despite the two missing parties but presumably because of the support for independents or other smaller parties, then no estimates are made for change in share for Labour and Liberal Democrats while the Conservatives are given the actual May to by-election change.

Another scenario is that now just two parties contest the May election but only one of these parties has a by-election candidate. An example is that Conservative and Labour challenge one another in May but that Labour does not contest the by-election (Example 3, Appendix). Since there is no Liberal Democrat standing at either election its change is regarded as missing data in the modeling procedure. An estimate for Labour's by-election vote is made in the usual manner by assuming that its by-election vote would be equivalent to its share in its worst performing ward across the local authority at the previous May election. The adjusted Conservative by-election vote share then becomes its actual share minus the estimated Labour share with the caveat that the direction of change cannot be counter-intuitive: it is set to zero if an actual positive change becomes a negative one after adjustments and estimated vote shares are made.

Thus far the examples shown have focused on the procedures necessary when the pattern of party competition at the by-election is less than at the previous May election. There are cases where the opposite situation applies and more parties challenge for the by-election vacancy. The first situation involves two main parties with candidates in May but these are subsequently joined by the missing party when the by-election is fought. We assume that Conservative and Labour are rivals in May but the Liberal Democrats put in a by-election appearance (Example 4, Appendix). In such a case an estimated May vote for the Liberal Democrats is based on the party's worst performing ward across the local authority. A restriction is imposed and this assumes that its by-election vote share is a positive change in vote from the May election; if not then change in share is limited to zero. For the

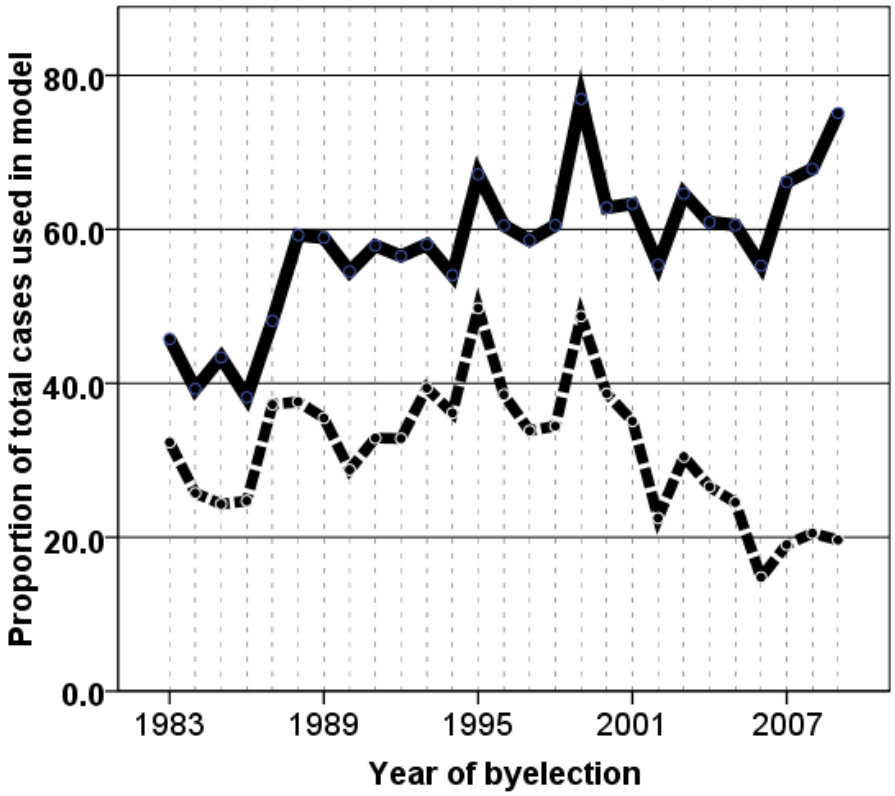
two parties that contested both elections change is set to zero if after estimating a May vote for the Liberal Democrats a negative change in vote becomes a positive one.

A minor adjustment to the example given above is that two parties contest the by-election but only one of these two was present at the previous May election. In May the Conservatives compete against independents and/or minor parties only but are then subsequently challenged by a Labour by-election candidate (Example 5, Appendix). Since there is no Liberal Democrat at either election its change is treated as missing. Labour's estimated May vote share is that in its worst performing ward across the local authority. The Conservative share in May is recalculated to be its actual vote minus the estimated Labour share. The restriction is that Labour's by-election share cannot be less than its estimated May share and is set to zero if that would happen. Similarly, Conservative change is set to zero if the consequences of these adjustments are that the direction of change is altered.

A more extensive adjustment is required when only one of the main parties contests the seat in May but the by-election sees all three parties contest. Although the Conservatives stood a candidate on each occasion its two rivals did not fight the May election (Example 6, Appendix). Estimates are calculated in the normal manner for each of Labour and Liberal Democrats that are equal to vote share in their weakest wards across the authority.

Finally, a more complex pattern of party competition occurs when two participate in the May election but a different pair of parties contest the by-election. For example, Conservative and Labour compete for the seat in May but then the by-election has no Labour candidates but local Liberal Democrats decide to contest (Example 7, Appendix). Step one estimates a May vote share for the Liberal Democrats calculated in the usual manner. Step two sees Labour's May vote reduced by subtracting its worst performing ward result while step three recalculates the share for the Conservatives. The normal restrictions are then applied.

The result of making such compensations is that a greater proportion of by-election cases may be included in modeling national equivalent vote estimates. The extent of that increase is shown in Figure 1. The two curves show the proportion of by-elections used in estimating national vote shares. The broken line is the proportion that were usable under the strict criteria of three-party competition in both the May and by-elections and a large fraction of total votes cast for the main parties. The solid line is the proportion after compensating for incomplete three-party competition and allowing for missing values when one or other of the three main parties fails to compete at either election. The most dramatic difference occurs in the period after 2001 when, as we reported earlier, there was a significant move away from the pattern of party competition that had evolved over an almost twenty-year period before then.



The final element in the revision process was to consider how best to create estimates for a given point in time. Providing estimates of national support based on by-elections from a single week would be ill advised since these can vary considerably, influenced by a range of local and national factors. A more robust approach is to use a broader time period. This has the effect of smoothing large fluctuations that sometimes occur using weekly data. Closer examination of results and trends over a twenty-five year period suggested that a more reliable procedure is a weighted quarterly moving average. Thus, each forecast is based not simply on the figures for a single month but also some information from the preceding two months. The weighting procedure takes into consideration the time elapsed from when each by-election occurs and the date of forecast, usually the last day of the month of interest. Thus, a by-election that happens on the last day of, let's say June, has a bigger impact on June's averages compared to a by-election that was held on April 1. There is a linear decrease in weights that reflects the days elapsed from the forecast date. In turn, the above June by-election will also have an impact on July and August 'averages' but its impact on the model estimate is reduced as time moves on. In short, if the number of by-elections stays more or less the same across all months then the influence of this June by-election decreases linearly (it is highest for June, smaller for July, and smallest for August) before it is completely removed from the quarterly calculation.

Estimating national equivalent vote share using the revised model

The first test of the revised model lay in its ability to forecast correctly the May elections using the April model estimate. The first iteration of the model was a reliable method for forecasting the national equivalent vote for local elections and clearly it would be a retrograde step if the revised method performed less well. Figures 2a-c show for each election year since 1993 both the April by-election model vote share (comprising data from the April, March and February results and the eventual May/June NEV for the three main parties. The solid line represents the by-election model share and the dotted line is the actual vote share. Overall, the methods for estimating missing data do not appear to impact negatively on the model forecasts for NEV.

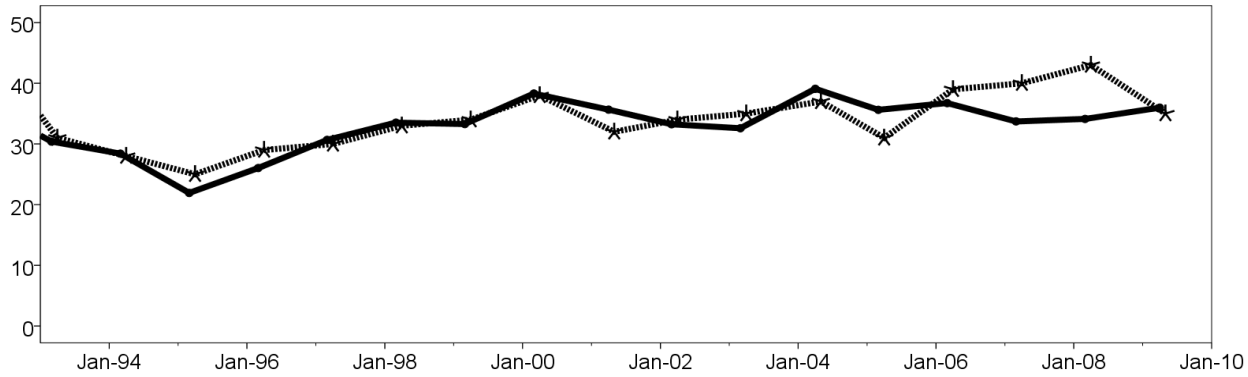
For the Conservatives the two curves are close together but there is no consistent pattern in terms of the forecast accuracy. The largest gap over the period is in 2005 when the model over-estimated the Conservative NEV (note: this is the party's local election performance rather than its general election vote share) and both 2007 and 2008 when it did somewhat better at the actual elections than had been forecast from the by-election data.

A close examination of the pattern of support for Labour again shows that the two curves are close together with two clear exceptions, both of which are general election years (2001 and 2005). The 1997 general election does not repeat this gap but it should be noted that this period (1995-1997) represented the peak of Labour's local electoral performance and it is unlikely that the party could have improved further on its by-election results. The 2001 and 2005 data are suggesting that Labour appears to raise its game (or its supporters take notice) where a general election contest arises. It is very rare for Labour to do better in by-elections than in the main May local elections; the one exception over this period is in 1997 but the gap is rather small.

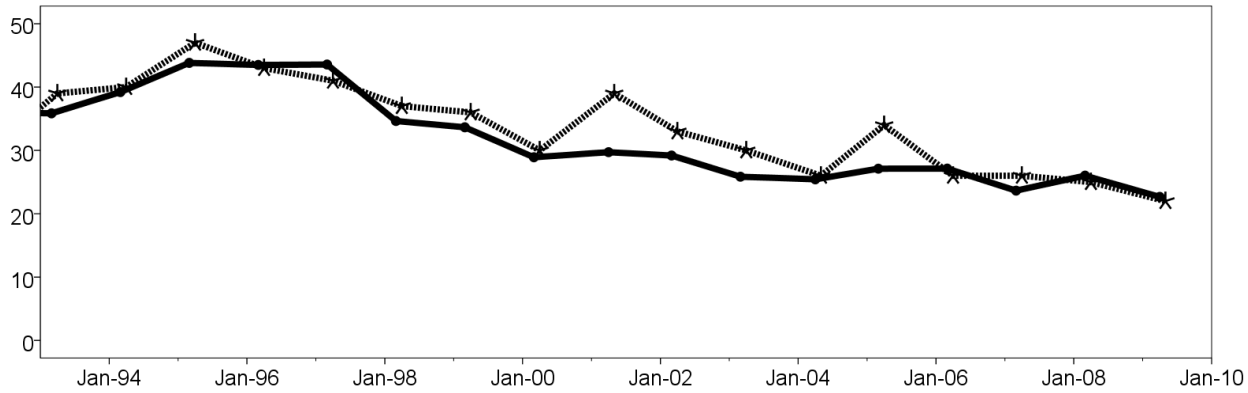
For the Liberal Democrats (Figure 2c) the pattern is the reverse of Labour's performance. The tendency is for the by-election model to exaggerate the eventual NEV. This feature is particularly noticeable in both 2003 and four years later in 2007. It does seem that the Liberal Democrats perform better in the by-election situation than at the national level when much of the country is voting in the main May elections. Of course, compared with the general election performance this gap (between by-election model and general election) is greater still.

Figure 2: The by-election model and NEV: 1993-2009

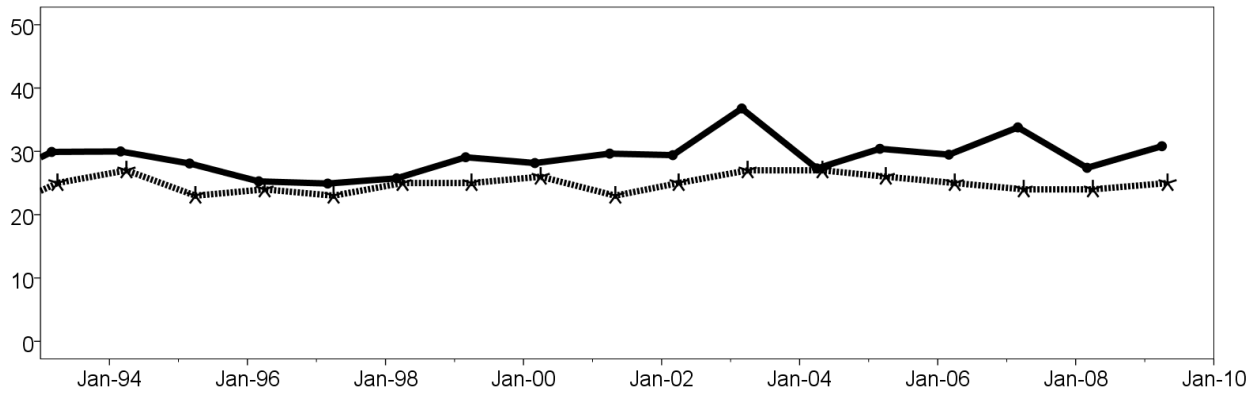
(a) Conservative



(b) Labour



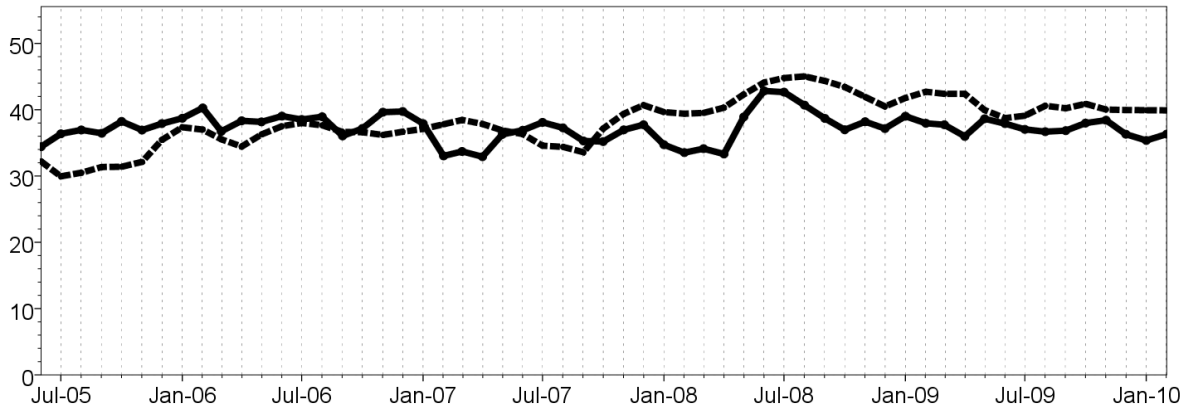
(c) Liberal Democrats



The trends in the by-election model and opinion polls since 2005

In this section we consider how the model has estimated support since the 2005 general election and compare that with the opinion polls (viz. weighted moving quarterly average). The intention, using Figures 3a-d, is to show more clearly the relationship between model estimates and the polls. Figure 3a tracks support for the Conservatives and shows that although both sets of figures follow one another quite closely, seldom being more than a few percentage points apart, there is no consistent pattern in the sense that one estimate is always above/below the other. In the immediate aftermath of the May 2005 election Conservative support rises gradually but then, following Brown's hesitation in calling a snap election in autumn 2007 accelerates towards a peak in midsummer 2008 according to the by-election model and two months later in terms of poll rating. Since then the trend has been gradually downwards.

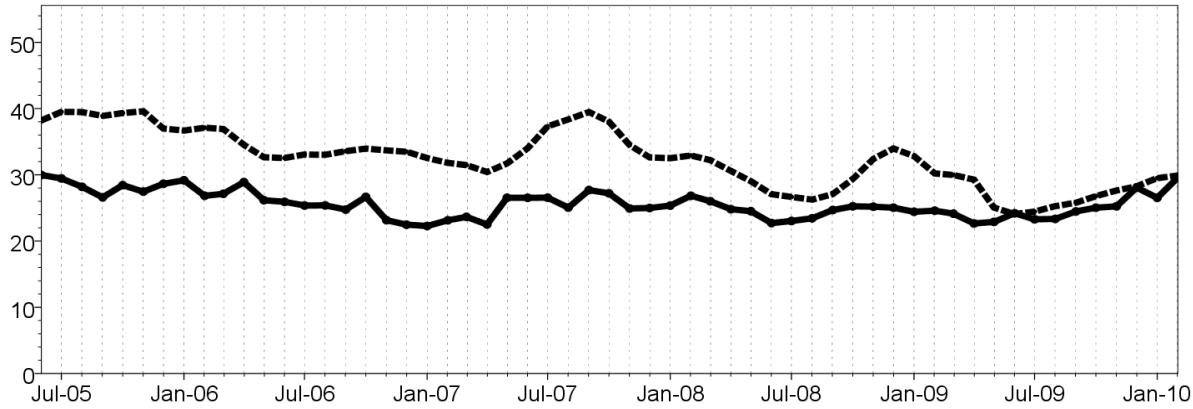
Figure 3a: the by-election model and opinion polls, Conservative



In the run-up to the 2005 general election Labour's estimated national vote using the by-election model was around 30% while the polls were somewhat higher, averaging in the high 30s (Figure 3b). Following the general election the party receives a small boost in the polls but its performance in by-elections notably declines towards the mid 20s. Thereafter the two lines track one another whilst remaining between 6-12 percentage points apart – however people are responding to surveys it appears that Labour is unable to translate such expressions of support into votes. Blair's announcement to leave office appears to be the catalyst that turns the party's fortunes around on both measures, noticeably so in terms of the poll ratings. Brown's hesitation pops the ratings bubble until the economic crisis and the Prime Minister's role in brokering international agreements contributes to another bounce in the polls for his party. In the meantime, Labour's by-election performance has remained lackluster with the improvement barely registering on this measure. In June 2009 the two measures intersect for the first time this parliament when the weighted poll

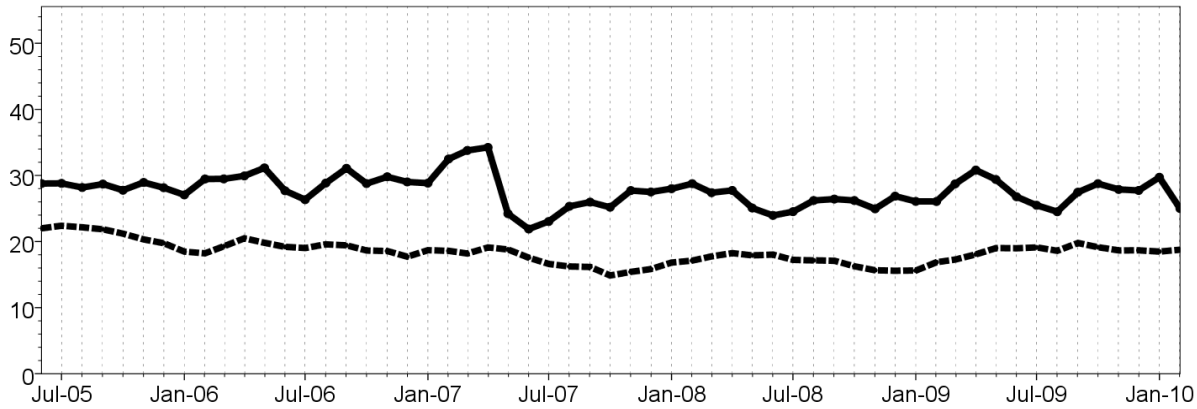
average reaches 24.1% and presumably hits the bedrock support for Labour; the by-election model had been estimating support of around that level for the previous year.

Figure 3b: the by-election model and opinion polls, Labour



The post 2005 data for the Liberal Democrats confirm the pattern of the previous twenty-two years. The party consistently performs better at local elections than its national poll rating suggests (Figure 3c). For the eighteen months following the general election the party hovers on or around the 30% mark but January 2007 marks a significant up-turn in its fortunes, reaching a four-year high of 34% in April that year. But then there is a rather dramatic 10-point collapse in May and June followed by a recovery over the autumn and winter months. Polling suggests that party support has ranged over a few percentage points throughout the entire parliament, entering the run-in to the 2010 general election a point or two lower than for its predecessor.

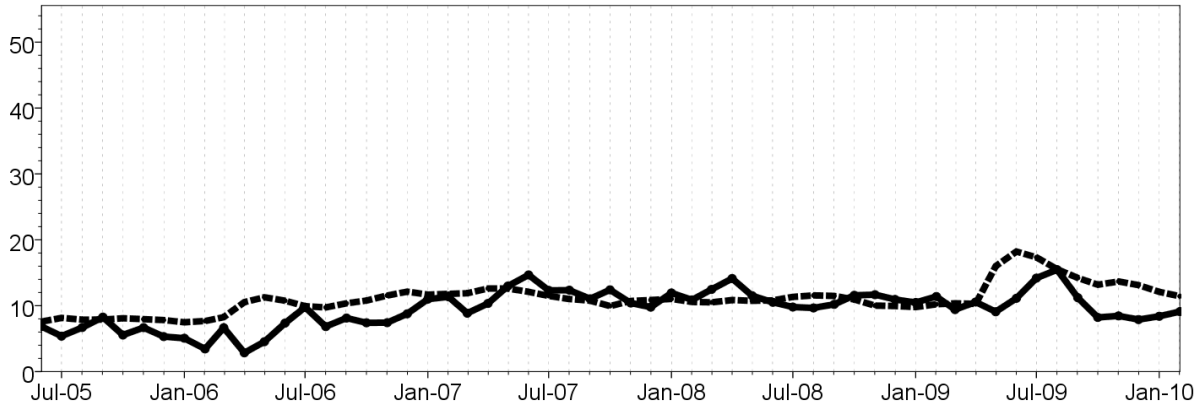
Figure 3c: the by-election model and opinion polls, Liberal Democrats



Finally, Figure 3d shows the trend in support for other parties. The two lines are close together for most of the time although the influence of the approaching European elections in June 2009, which triggers a growth in support for such parties in the opinion polls is delayed slightly for the by-election

model (a case of polls driving votes?) and is short-lived although as the general election nears the two lines may be intersecting.

Figure 3d: the by-election model and opinion polls, Others



Any student of parliamentary by-elections and their outcomes would know that they are not reliable guides to how people might vote at a general election but it does appear that aggregate local electoral data are picking up similar movements in electoral opinion to the individual level data acquired in national surveys. The votes cast there are real votes and voters have incurred some costs in performing these actions. There is no need to adjust them in terms of weighted past vote or the likelihood of actually voting but there is a need to take account of variance in the pattern of party competition. On some occasions the by-election model takes time to respond to issues that are immediately apparent in the polls but on other occasions they may be a better guide to the underlying trends. They provide a more reliable indicator of how local voters may behave than surveys but what do they portend for the 2010 general election.

Estimating the 2010 vote shares

As we stated at the outset the model is principally designed for another purpose than the one it is being employed here. Judged solely on the basis of its ability to forecast national equivalent vote shares for the annual local election cycle it is a success, seldom being more than a couple of percentage points out for any single party if general election years are excluded. There is no doubt, however, that for the model to work to estimate parliamentary voting then certain adjustments should be made. The main adjustment affects support for the third party, the Liberal Democrats, which is clearly in receipt of support from some voters expressing a by-election protest and other voters that over the past decade and a half have become split-ticket voters (either because they genuinely prefer to vote Liberal Democrat at local council elections or because of tactical voting

reasons). There are other, perhaps less prominent factors, that should also be taken into account when adapting this model to generate general election forecasts. One such is Labour's general weakness in apparently getting its vote out. Even allowing for some over-estimation of its support by the pollsters, Labour has under-achieved, with the notable exception of the pre-1997 period.

At the time of writing the current NEV estimate is showing Conservatives 36.3%, Labour 29.6% and Liberal Democrats on 25.0%. In preparing this general election forecast we assume, therefore, that a re-distribution of the current Liberal Democrat share is required. Without any current polling data to work from we assume from our experience of split-ticket voting that eight in ten current Liberal Democrat local by-election voters will remain loyal to the party at the general election; one in ten will switch to Labour while 6%, approximately one in 16, will vote Conservative. The adapted by-election model forecast, therefore, is Conservative 37.8%, Labour 32.1% and Liberal Democrats 20.0%. An assumption of uniform national swing converts these percentages into a House of Commons comprising Conservatives 287 seats (39 seats short of an overall majority), Labour 280 seats (46 seats short), Liberal Democrats 51 seats and other parties 32 seats. Of course, there are still some by-election votes still to be cast and we reserve the right for our final forecast to alter the algorithm for converting Liberal Democrat local support into that likely to be encountered at a general election.

References

Curtice, J., and C. Payne. "Local Elections as National Referendums in Great-Britain." *Electoral Studies* 10, no. 1 (1991): 3-17.

Rallings, C., and M. Thrasher. "Exploring Uniformity and Variability in Local Electoral Outcomes - Some Evidence from English Local Elections 1985-1991." *Electoral Studies* 12 (1993): 366-84.

Rallings, C., and M. Thrasher. "Split-Ticket Voting at the 1997 British General and Local Elections: An Aggregate Analysis." In *British Elections & Parties Review*, edited by D. Denver, J. Fisher, P. Cowley and C. Pattie. London: Frank Cass Publishers, 1998.

Rallings, C., and M. Thrasher. "Local Votes, National Forecasts - Using Local Government by-Elections in Britain to Estimate Party Support." *International Journal of Forecasting* 15 (1999): 153-62.

Rallings, C., and M. Thrasher. "Measuring the Level and Direction of Split-Ticket Voting at the 1979 and 1997 British General and Local Elections: A Survey- Based Analysis." *Political Studies* 49, no. 2 (2001): 323-30.

Rallings, C., and M. Thrasher. "Explaining Split-Ticket Voting at the 1979 and 1997 General and Local Elections in England." *Political Studies* 51, no. 3 (2003): 558-72.

Ranney, A. "The Utility and Limitations of Aggregate Data in the Study of Electoral Behaviour." In *Essays on the Behavioral Study of Politics*, edited by A. Ranney. Urbana: University of Illinois Press, 1962.

Appendix

Estimation of current national equivalent vote shares

The model may be formally expressed as follows:

Let ***CONsh***, ***LABsh***, and ***LDsh*** be the ward vote share at the by-election for Conservative, Labour and Liberal Democrats respectively while the parties' vote share at the previous ward election held in May would be $\boldsymbol{CONsh_{May}}$, $\boldsymbol{LABsh_{May}}$, and $\boldsymbol{LDsh_{May}}$. The national equivalent vote at the date of the relevant previous May local election is designated as ***NEV.CON, NEV.LAB***, and ***NEV.LD***.

The difference in share of vote between May and By-elections for enumerated parties might then be represented as follows:

$\boldsymbol{CONchange = CONsh - CONsh_{May}}$
$\boldsymbol{LABchange = LABsh - LABsh_{May}}$
$\boldsymbol{LDchange = LDsh \quad - LDsh_{May}}$

Finally, to estimate the current national equivalent vote we add to the previous national equivalent vote the difference between a party's by-election vote share and its vote share recorded at the May election.

Estimates of current NEV are calculated as follows,

NEV.CON + CONchange, ***NEV.LAB + LABchange***, ***NEV.LD + LDchange.***

This process is repeated for as many by-elections as fit our specified criteria and each party's current national equivalent vote is arrived at by averaging the results over a stated time period.

Missing data imputation

Example 1: Imputation for missing vote share for Liberal Democrats absent from by-election contest

Let all three main parties have candidates in May election but Liberal Democrats provide no candidate in the by-election:

May election → By-election
con+lab+ld → con+lab

The imputation procedure implies the following notional by-election result:
$\boldsymbol{LDsh^{new} = MINIMUM_{across\ district\ in\ May}(LDsh)}$
$\boldsymbol{CONsh^{new} = CONsh - CONsh/(CONsh+LABsh) * LDsh^{new}}$
$\boldsymbol{LABsh^{new} = LABsh - LABsh/(CONsh+LABsh) * LDsh^{new}}$

In order to preserve the actual *direction* of changes for all parties some restrictions are imposed:
If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set ***CONchange = 0***

If $LABsh > LABsh_{May}$ but $LABsh^{new} < LABsh_{May}$ then ***LABchange = 0***

> *Note:* In above formulae and everywhere else in the paper, superscript '*new*' reflects notional by- or May election when any procedure of imputation is applied. Following the imputation procedure, notional election result (i.e. 'new') is then treated in the usual way for the purposes of estimating NEV.

Example 2: Imputation for missing *vote share for both Labour and Liberal Democrats absent from by-election contest* (con+lab+ld → con)

Notional By-election result:
$\boldsymbol{LDsh^{new} = MINIMUM_{across\ district\ in\ May}(LDsh)}$
$\boldsymbol{LABsh^{new} = MINIMUM_{across\ district\ in\ May}(LABsh)}$
$\boldsymbol{CONsh^{new} = CONsh - LABsh^{new} - LDsh^{new}}$

If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set ***CONchange = 0.***
If $CONsh <= CONsh_{May}$, then
LABchange = MISSING, ***LDchange = MISSING***; C$\boldsymbol{ONchange = CONsh - CONsh_{May}}$

Example 3: Imputation for missing Labour by-election share in case when Liberal Democrats absent from both May and by-election contest **(**con+lab → con)

LDchange = MISSING

Notional By-election result:
$\boldsymbol{LABsh^{new} = MINIMUM_{across\ district\ in\ May}(LABsh)}$
$\boldsymbol{CONsh^{new} = CONsh - LABsh^{new}}$
If $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}$ then we set ***CONchange = 0***

*Example 4: Imputation for May vote share for Liberal Democrats (*con+lab → con+lab+ld)

Notional May-election results:
$\boldsymbol{LDsh_{May}^{new} = MINIMUM_{across\ district\ in\ May}(LDsh)}$
$\boldsymbol{CONsh_{May}^{new} = CONsh_{May} - CONsh_{May}/(CONsh_{May}+LABsh_{May}) * LDsh_{May}^{new}}$
$\boldsymbol{LABsh_{May}^{new} = LABsh_{May} - LABsh_{May}/(CONsh_{May}+LABsh_{May}) * LDsh_{May}^{new}}$

Restrictions:
If $LDsh < LDsh_{May}^{new}$ then ***LDchange =0***.
If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$ then we set ***CONchange = 0.***
If $LABsh < LABsh_{May}$ but $LABsh^{new} > LABsh_{May}^{new}$ then we set ***LABchange = 0.***

Example 5: Imputation for missing Labour May vote share in case when Liberal Democrats absent from both May and by-election contest **(**con → con+lab)

There is no LD candidate in May and by-elections. So, we cannot assess LD changes:
LDchange = MISSING
Notional May-election results:
$\boldsymbol{LABsh_{May}^{new} = MINIMUM_{across\ district\ in\ May}(LABsh)}$
$\boldsymbol{CONsh_{May}^{new} = CONsh_{May} - LABsh_{May}^{new}}$

Restrictions:
We have to get at least a non-negative change for LAB, so
If $LABsh < LABsh_{May}^{new}$ then ***LABchange =0***.

If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$ then we set ***CONchange = 0.***

Example 6: Imputation for missing May vote shares for both Labour and Liberal Democrats (con → con+lab+ld)

Notional May-election results:
$\boldsymbol{LABsh_{May}^{new} = MINIMUM_{across\ district\ in\ May}(LABsh)}$
$\boldsymbol{LDsh_{May}^{new} = MINIMUM_{across\ district\ in\ May}(LDsh)}$
$\boldsymbol{CONsh_{May}^{new} = CONsh_{May} - LABsh_{May}^{new} - LDsh_{May}^{new}}$
Then usual procedure for calculation of changes and estimated shares is applied.

Restrictions:
We have to get at least a non-negative change for *LAB* and *LD*, so
If $LABsh < LABsh_{May}^{new}$ then ***LABchange*** *=0*
If $LDsh < LDsh_{May}^{new}$ then ***LDchange*** *=0*

If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$ then we set ***CONchange = 0***

Example 7: Imputation for missing vote shares for both May and by-elections for different parties (con+lab → con+ld)

Notional May-election results:
$\boldsymbol{LDsh_{May}^{new} = MINIMUM_{across\ district\ in\ May}(LDsh)}$
$\boldsymbol{LABsh_{May}^{new} = LABsh_{May} - MINIMUM_{across\ district\ in\ May}(LABsh)}$
$\boldsymbol{CONsh_{May}^{new} = CONsh_{May} - LDsh_{May}^{new} + LABsh_{May}^{new}}$

Restrictions:
We have to get at least a non-negative change for *LD*:
if $LDsh < LDsh_{May}^{new}$ then ***LDchange =0*** .

If $CONsh < CONsh_{May}$ but $CONsh^{new} > CONsh_{May}^{new}$
or $CONsh > CONsh_{May}$ but $CONsh^{new} < CONsh_{May}^{new}$
then we set ***CONchange = 0 .***