Estimating local authority level distributions of referendum voting

using aggregate and survey-level data

Michael Thrasher*, Galina Borisyuk*, Colin Rallings*, Harry Carr† and Michael Turner††

* The Elections Centre, Plymouth University

† Sky Data, SkyNews; †† BMG Research

Introduction

The paper describes the design and implementation of a forecasting model for the EU referendum on June 23, 2016. The model provided prior estimates of both turnout and percentage vote for Leave and Remain across the almost four hundred counting areas. When some areas declared either their turnout or voters' decision on the referendum question the model used those data to revise its estimates.

The method was first used successfully for the 2014 independence referendum in Scotland. In 2016 it was used as the basis for forecasting the referendum result on Sky News' overnight programming, including on-screen forecasts of the likely outcome. While other forecasting methods were developed we believe that this method is particularly suited to dis-aggregating publically available national polling data and could be easily adapted for use in forecasting future referendums in the UK.

Using polling evidence to forecast seat distributions in the House of Commons requires both knowledge about the previous election and also assumptions about the operation of change in vote or 'swing'. For one-off referendums such comparisons are impractical. Of course, if all that is required is a forecast of national vote then the polling evidence should, in theory, be sufficient. But polling data is not entirely suitable in providing voting intention forecasts at lower levels of aggregation (in this case local authorities) although some attempts were made prior to the EU referendum to pool data and increase the number of cases in that way. The task of making small area forecasts is left to the exit poll at a general election. The idea of running a referendum exit poll was ruled out mainly because there was no basis for selecting sampling areas that would reflect the national vote.

The absence of both an exit-poll based forecast of the national result together with a reasonably detailed picture of how that might play out in hundreds of places presents a major challenge for national broadcasters, including Sky News. Quite apart from depriving them of the 10.00 p.m. moment as the polling stations are closed and the result projection is announced there is the on-going problem over the following hours of knowing beforehand whether place 'x' or place 'y' might opt to remain within the EU. Moreover, broadcasters need some basis for estimating that given result in place 'x' this is what we now expect from place 'y'.

The task, therefore, lay in devising an accurate forecasting tool that could:

a) Provide accurate estimates of the leave/remain vote across the 382 counting areas used to collate and declare the overnight referendum vote. These estimates would need to be sensitive to the changing state of public opinion in the days leading to the vote.

b) Provide accurate estimates of likely turnout in the counting areas. This was especially important given that the referendum was a national vote but the counting areas varied widely in size with the largest area containing 700 times the electorate as the smallest area.

c) Re-calibrate both turnout and the leave/remain vote for each counting area as soon as some areas began first to declare the number of ballot papers issued and second the leave/remain outcome in their areas.

d) Project the overall total number of votes that would be cast and the percentage share for leave/remain vote some considerable time before the Electoral Commission's official declaration expected around breakfast time on June 24.

The paper begins by considering the challenge of forecasting referendum voting at the sub-national level before providing a detailed description of the method devised for the EU referendum. The third section describes the evening of June 23/4 when three separate operations were in place in order to ensure an accurate forecast for the Sky News' broadcast. In the conclusions we address some of the strengths and weaknesses of this approach and assess its suitability as a tool for forecasting the outcome of future referendums.

Forecasting referendum outcomes

Compared with general election/presidential forecasting there is relatively little research effort expended on forecasting models for referendum outcomes. This might be because it is a mug's game – any referendum forecast is either right or wrong and to be on the wrong side of a two-horse race impresses no one (few in the outside world are interested in confidence intervals of +/- 8 percentage points). And if the outcome appears clear-cut beforehand (e.g. the 2011 AV referendum) then making forecasts is largely a redundant exercise.

Those forecasts that do appear are almost exclusively based on the polling evidence, either using the voting intention question itself or some other attitudinal-based questions to form a judgement call. For the EU referendum some went beyond a simple national forecast and published counting level estimates of likely vote outcome. Two similar approaches are those of Ron Johnston, Kelvin Jones and David Manley of the University of Bristol and that produced by Chris Hanretty of the University of East Anglia[1].

The Bristol team were given access to a large dataset provided by YouGov (N=60,000 responses) from which they modelled the likely probability of leave/remain choices among individuals in different age and education groupings. They then link these probabilities to the local authority level demographic characteristics derived from the 2011 census. Mapping local authority areas according to both the probability of voting 'leave' and the UKIP vote share obtained at the 2014 European Parliament elections reveals a correlation of 0.75 and the familiar spatial pattern of anti-EU sentiment that is strong in areas of eastern England but weak in London and Scotland. Because their

---

[1] Commentaries from Johnston et al. are here: http://blogs.lse.ac.uk/politicsandpolicy/can-we-really-not-predict-who-will-vote-for-brexit-and-where; from Hanretty here: https://medium.com/@chrishanretty/the-eu-referendum-what-to-expect-on-the-night-521792dd3eef#.i097nocx7

analysis uses polling data that points towards a remain outcome it identifies only a small number of local authorities where, at the time of writing, April 2016, the leave voters would be a majority.

Hanretty's approach is similar but his estimates are based on data obtained from the post-election wave of the British Election Study undertaken in May 2015. Using these data to construct estimates of opinion at the local authority level he assumes a uniform swing to take account of the movement in public opinion since the polling was undertaken and also ignores any variations in turnout that might follow when electors are asked to vote in a referendum rather than a general election. Wishing to be transparent about his conclusions, Hanretty provides public access to his detailed estimates.[2] Other publically-available local authority-level estimates were produced by Matthew Goodwin of the University of Kent but his calculations excluded Scotland.

Finally, John Curtice and Steve Fisher, of Strathclyde and Oxford universities respectively, published an abbreviated list of local authorities that would likely be strongly in favour of one of the two available options as well as those likely somewhere in the middle of the pro/anti EU spectrum[3]. Doubtless, because of their association with the BBC and its requirement for strict neutrality on the EU referendum coverage, Curtice and Fisher did not publish a full list of local authority estimates but did reveal their method of estimation. In many respects this is similar to Johnston et al. since here too it is based on a large dataset supplied by YouGov. Similarly, the age and education profiles of local authorities from the 2011 census are used to calibrate the data. Curtice and Fisher, like Hanretty, assume that the variance in turnout at the referendum will mirror the pattern of turnout amongst different social groups at the 2015 general election as evidenced by BES data. They too acknowledge the strong correlation between local authorities where UKIP polled well at the 2014 European elections and add the measure of how well left-leaning parties polled at those elections.[4] Regarding turnout, Curtice and Fisher estimate a national figure of 60% and variations from that are based on the pattern encountered at the 2014 European election. Finally, they adjust their point estimates for each local authority based on an assumption that the country might vote 50/50 for leave/remain. This approach meant that as each local authority declared its result on the night the extent to which it diverged from the estimate would indicate which side of the argument would likely prevail.

An alternative approach

The approach outlined in this paper differs from those above. Rather than identifying the demographic characteristics of voters likely to support leave or remain we assess survey respondents' previous voting choices alongside their referendum voting intention. This leads to estimates of how each party's general election supporters might split on the referendum issue. It

---

[2] The data are available at, https://docs.google.com/spreadsheets/d/1tRE59IkgeRREISpM75I8gR0MdkGe1diParW0hVO109Y/edit#gid=881507152

[3] See https://electionsetc.com/2016/06/22/how-the-bbc-will-be-benchmarking-the-results-on-eu-referendum-night/

[4] Curtice and Fisher produce estimates for both Gibraltar and Northern Ireland separately and these two counting areas are exceptions for all of the methods discussed here.

enables the modelling, in theory, to become more sensitive to campaign cues provided (or not in the case of Labour in 2016) by the parties and their principal actors. A second difference lies with the derivation of local authority-level estimates for the referendum vote. Instead of using 2011 census attributes this approach uses the votes cast at the 2015 general election and transposes them from parliamentary constituencies to the local authorities. A further difference relates to likely turnout where we compare variations in local authority-level turnout across a range of elections as well as the 2011 AV referendum. Below, we describe each of these steps in more detail.

The first step involves cross-tabulating referendum voting intention with respondents' recall of past voting. In the case of the Scottish referendum the question related to vote choice at the 2011 Scottish Parliament election and for the EU referendum the 2015 general election vote. This establishes the proportion of each party's supporters as they divide over the referendum question. For each set of polling data some account is taken of respondents' likelihood to vote while those still yet to make up their minds are excluded altogether.

Table 1: Referendum voting intention by 2015 general election vote choice

		General Election 2015							
Referendum choice	*Con*	*Lab*	*LD*	*UKIP*	*Other*	*Didn't Vote*		*Total*	*Total*
Remain	N=	94	128	37	3	48		100	410
	%	**27.4**	**45.1**	**50.7**	**2.6**	**52.2**		**17.9**	27.9
Leave	N=	180	82	21	90	28		118	519
	%	**52.5**	**28.9**	**28.8**	**76.9**	**30.4**		**21.1**	35.4
Won't vote	N=	69	74	15	24	16		341	539
	%	20.1	26.1	20.5	20.5	17.4		61.0	36.7
Total	N=	343	284	73	117	92		559	1,468
	%	100	100	100	100	100		100	100

Both BMG and Sky Data supplied raw data for the modelling and an example taken from a BMG poll is shown in Table 1[5]. For past general election vote only five party categories are used with Scottish and Welsh Nationalists and the Green party (cases in this survey were 41, 3, and 44 respectively) included in the 'Other' category. Respondents that had abstained at the general election, or could not recall their general election vote or preferred not to give an answer but were intending to vote at the referendum were included in a single category 'Didn't vote'. Referendum voting intention was allocated to three categories – leave (this included those would 'definitely vote' and had

---

[5] Demographics and past vote weighting is applied

decided leave or those 'leaning towards' leave, remain and a catch-all group which included those yet to decide as well as those who would not vote.

From the particular percentages contained in Table 1 we calculate the remain and leave vote as follows:

Remain = 0.274 * Con + 0.451 * Lab + 0.507 * LD + 0.026 * UKIP + 0.522 * Other + 0.172 * Abstainers

Leave=0.525 * Con + 0.289 * Lab + 0.288 * LD + 0.769 * UKIP + 0.304 * Other + 0.211 * Abstainers,

where Con, Lab, … are number of votes for the corresponding party at the 2015 general election

and 'Abstainers' are the number of electors who did not vote in 2015 (or did not say how they had voted).

The same procedure may be used with published tables from other polls although more steps may be required. For example, most EU referendum polls provided a table examining referendum voting intention by recall of 2015 vote. Some, though not all, provided a further classification based on likelihood to vote in the referendum.

This procedure provides the proportions of each party's supporters at the national level. From this information it is a simple matter to use the national figures to construct *constituency-level* estimates of the referendum vote but more difficult to re-configure the data for the local authority counting areas. The second step is therefore to use the evidence regarding the likely distributions of party supporters and apply these to aggregate-level data for the relevant counting areas.

For the Scottish referendum the 32 local and islands councils were used to count and declare the results. The most recent council elections were held only two years before the referendum but these were unsuitable because of the large number of votes cast for non-party candidates. The 2011 Scottish Parliament elections would in any case be a better option but the problem lay in transposing these votes to the local authority level. Fortunately, we were able to call upon the expertise of David Denver of Lancaster University to undertake that task on our behalf.

The same problem affected the 2016 referendum. Unlike the United States, for example, where aggregate voting data are available at precinct level, the UK does not release general election votes at smaller areas than the constituency. Therefore, after removing both Gibraltar and Northern Ireland from consideration (the former had not voted at the general election while polling data largely ignored the latter) the task was to transpose the votes cast in 650 parliamentary constituencies into the referendum counting areas.

The south west of England provides some examples that help to illustrate what is required. Figure 1 includes an outline of the boundaries of Taunton Deane borough council. In this case the local

authority and parliamentary constituency are coterminous. Figure 1 also features a more complex example. The constituency of Bridgwater and West Somerset comprises all 16 wards from West Somerset district and a further 17 wards from Sedgemoor council. We regard these two groups of wards as the two 'building blocks' that will be used to disaggregate the 2015 general election vote. The parliamentary constituency of Devon Central (Figure 3) provides one of the more complicated examples to transpose voting data to the local authority level. This constituency comprises wards from four local authorities – East Devon, Mid Devon, Teignbridge and West Devon. West Devon's wards are placed into two constituencies – Devon Central (identified by upper case 'A') as well as Torridge and West Devon (lower case 'a'). Likewise, those Mid Devon wards located in Devon Central are identified by 'B' and Teignbridge 'D'. Only a small area of East Devon district council contributes towards the Devon Central seat ('C') while its remaining wards are scattered across two other parliamentary seats, East Devon and Tiverton 'c1' and Honiton 'c2'. In total across the 650 constituencies there are 900 of these 'building blocks' that are used to assign the 2015 general election vote to the local authorities that administered the count in the EU referendum.
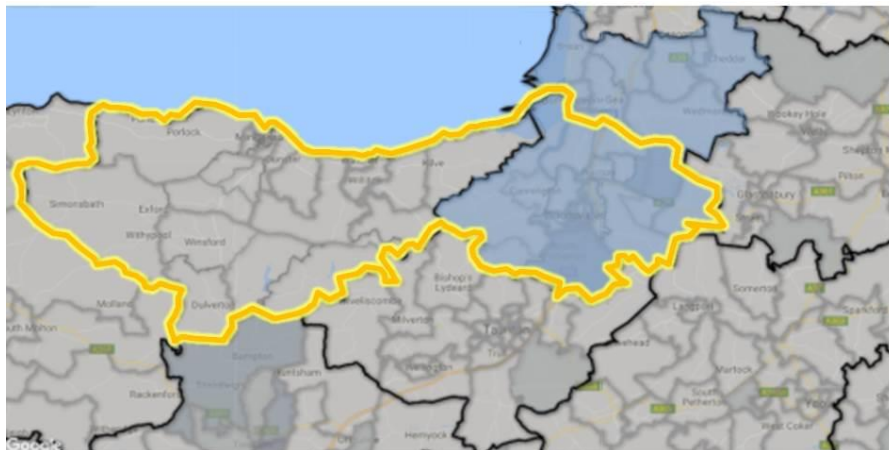

Figure 1: Area covered by Taunton Deane and Bridgwater & West Somerset constituencies
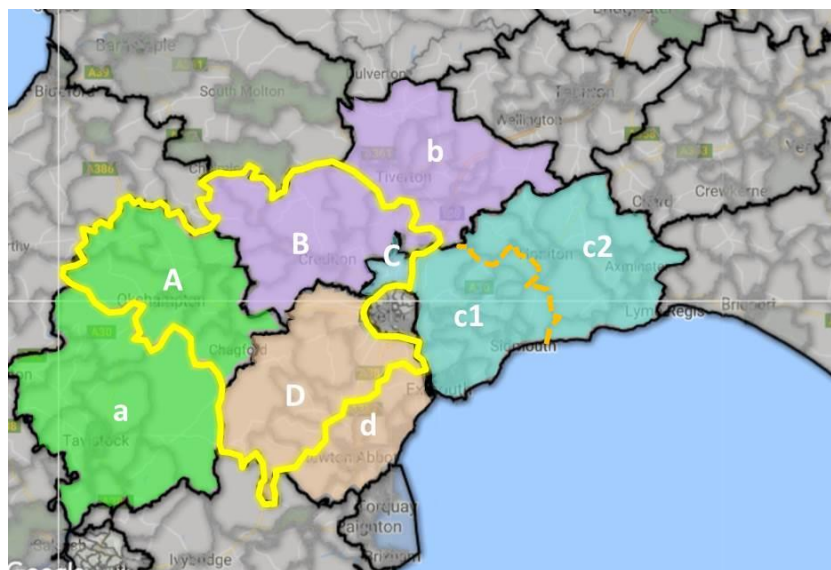

Figure 2: Area covered by Devon Central

Across Britain there are 35 cases that conform to the example of Taunton Deane where the boundaries of a parliamentary constituency are coterminous with a local authority. A majority of these, 23 seats, are in England with a further nine in Scotland and three from Wales. Given that the data obtained from the cross-tabulation provide the coefficients that should be applied to each party's general election constituency vote it is a simple procedure to calculate the expected vote in these particular local authority areas.

In cases where multiple constituencies lie entirely within the boundaries of a single local authority the procedure for establishing the local authority-level general election vote is equally straightforward. Leicester and Nottingham city council areas, for example, each have exactly three parliamentary seats contained within their respective boundaries.

There are 198 parliamentary seats, however, where constituency boundaries cross over into two local authorities as in the case of Bridgwater and West Somerset. For example, although the London borough of Barking and Dagenham contains the whole of the Barking parliamentary seat its neighbour, Dagenham and Rainham has some of its wards that are located in the borough of Havering. In these types of cases, we use that fact that local authority wards are the basic building block for parliamentary seats to guide the process for converting parliamentary voting to the local authority level[6].

Consider a parliamentary seat comprising 20 equally-sized local authority wards, ten of which belong to local authority A, the remainder in local authority B. In those circumstances an equal number of electors come into the constituency from each of the two local authorities. Thereafter, it is simply a matter of dividing the parliamentary votes in half and re-allocating these votes to the respective local authorities. In another example a parliamentary seat derives 80% of its electors from one local authority and 20% from another. In this case the re-allocation of votes would reflect those proportions.

There are 25 constituencies such as Devon Central which draw electorates from three or more adjoining local authorities but precisely the same method as above is applied to these. So, the Arundel and South Downs constituency has approximately 28% of its electorate reside within Arun district council, 10% in Chichester, 47% in Horsham and the remaining 15% in Mid Sussex.

Having identified the correspondence between parliamentary constituency and local authority electorates the next step is to estimate local authority levels of party support from the 2015 general election. The process of re-distribution that we adopt does not currently take account of the partisan characteristics of wards as evidenced by voting at local elections. Conservative voters in a particular constituency may be disproportionately drawn from one particular local authority but this procedure re-distributes according to the proportion of electorate and not their actual partisan preferences. We do not feel that striving for such a level of precision would greatly improve the accuracy of forecasting.

---

[6] There are a small number of wards that divide across constituencies following local authority boundary changes but these are few in number. In these cases we simply divide ward-level electors in equal proportions.

To demonstrate how 2015 votes are reallocated to the local authorities consider the case of Bridgwater and West Somerset (B&WS).

(1) Re-distribute each constituency 2015 party votes to the building blocks associated with the constituency in proportion to the size (defined in terms of electorate) of blocks. B&WS consists of two building blocks (all of the 16 wards from West Somerset district and 17 out of 25 wards from Sedgemoor district council)

$$\text{Electorate}_{\text{B\&WS}} = \text{Electorate}_{\text{West Somerset}} + \text{Electorate}_{\text{17 wards of Sedgemoor}}$$

$$\text{Electorate}_{\text{B\&WS}} = 27{,}868 + 58{,}704$$

Therefore, 2015 votes cast in the Bridgwater and West Somerset constituency are re-allocated to two separate building blocks, West Somerset and part of Sedgemoor, in the ratio of approximately 1:2 (27,868 / 58,704 = 0.475).

The 2015 general election results show $\text{CON}_{\text{B\&WS}}$ = 25,020. This means that for block one, West Somerset, the Conservative 2015 vote estimate is 8,340 with the remaining 16,680 Conservative votes allocated to block two, part of the Sedgemoor district.

The numbers of vote cast for other parties and the number of abstainers are redistributed from B&WS constituency to its two blocks using this procedure.

(2) Using the building blocks' estimates, construct local authority-level estimates of 2015 votes.

West Somerset contains just one block, so all of the West Somerset estimates are already calculated at stage (1).

Sedgemoor votes are calculated as the sum of votes 'received' from B&WS constituency and votes allocated from the area that comes in from the adjacent constituency of Wells. The Conservative vote in Sedgemoor is estimated as:

Con Vote = 16,680 from B&WS + 10,334 from Wells = 27,014

(1) The final step is to bring together the calculations based on evidence from the survey data with those of the 2015 aggregate data. Using the coefficients from polling data (Table 1), we calculate the expected number of leave and remain votes for the Sedgmoor local authority area as:

Remain = 0.274 * Con + 0.451 * Lab + 0.507 * LD + 0.026 * UKIP + 0.522 * Other + 0.172 * Abstainers

= 0.274 * 27,014 + 0.451 * 7,949 + 0.507 * 11,732 + 0.026 * 9,238 + 0.522 * 2,772 + 0.172 * 26,274 = 23,141

Leave=0.525 * Con + 0.289 * Lab + 0.288 * LD + 0.769 * UKIP + 0.304 * Other + 0.211 * Abstainers

= 0.525 * 27,014 + 0.289 * 7,949 + 0.288 * 11,732 + 0.769 * 9,238 + 0.304 * 2,772 + 0.211 * 26,274 = 33,348

% Remain = Remain / (Remain + Leave) *100 = 23,141 / (23,141 + 33,348) * 100 = 41%

% Leave = 59%

The resulting estimates were compared with those produced by Chris Hanretty (Figure 3). Because his published estimate of the overall leave vote is lower than our forecast the data points do not fall on the main diagonal but the most interesting feature is the very high correlation between estimates derived in such different ways. Critically, there is also agreement that a majority of local authorities in both Scotland (where our own estimates are compromised by an absence of Scotland-specific polling data) and London (with some notable exceptions) would be among the areas least enthusiastic for Brexit.
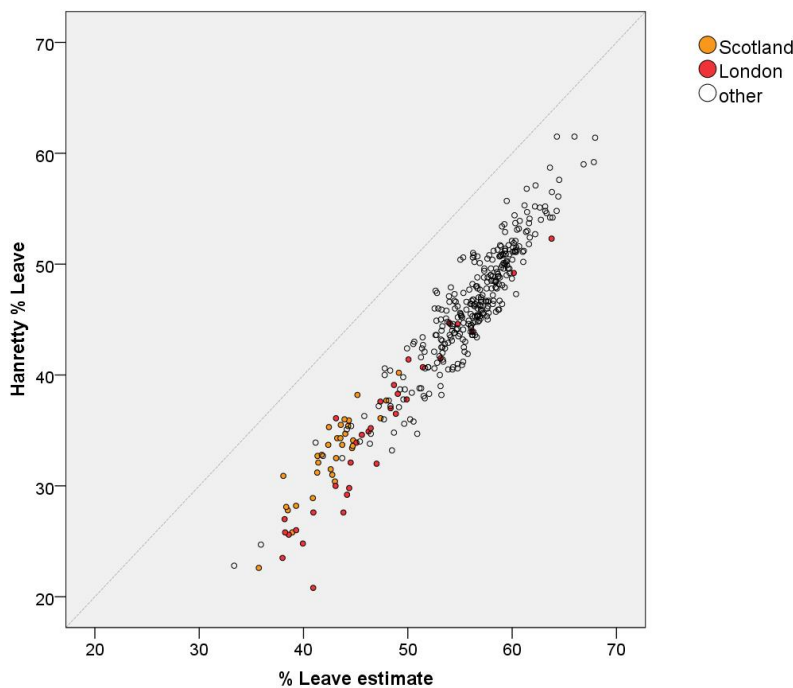

Figure 3: Comparing methods of local authority level estimate of percentage leave vote

It is important to note that the method described here does not add or subtract votes from the overall number in May 2015 – for example, there are still some 3,862,775 votes cast for UKIP across Britain although we accept that not all are allocated exactly to the correct local authority. At worst,

the effect of any mismatching will be slightly to over- or under-estimate the likely leave/remain vote split for a given local authority.

A second potential criticism is that the re-distribution method uses electorates based on those eligible to vote in local elections whereas the referendum vote franchise is the same as that for a general election. However, after comparing the correspondence between the two franchises we are confident that differences do not generally impede the method of re-casting votes since we are concerned with the relative size of parts of constituencies that are required to be moved between local authorities.

Given that the largest counting area by electorate size is seven hundred times the smallest it was important that the modelling produced reasonably accurate forecasts in this regard. The national level of turnout too is derived from the polling data by using information, normally supplied, regarding each respondent's likelihood of voting. Polling methodologies differ, of course, both in terms of the phrasing of the question, the scaling of the variable, and what happens with the information thereafter in terms of excluding all those unlikely to vote, including only those certain to vote or some other method. Moving from a national estimate of turnout to local estimates used previous research that demonstrates that while electoral participation fluctuates according to the type of election the relative position of local authorities/parliamentary constituencies to one another is reasonably stable. Three sets of turnout were used to estimate a rank ordering of local authorities – the 2015 general election, 2014 European parliament elections and 2011 AV referendum (but not in the case of Scotland and Wales where the counting areas were different; instead turnout for the devolved elections in 2012 was used). For each election the deviation from the overall turnout was noted and then the average deviation of the three separate elections was used to establish the authority's rank order. In effect, the estimate of local authority turnout became a combination of the overall turnout estimate plus the local authority deviation (Table 2).

Table 2: Compiling estimates of referendum turnout

Local authority	2011		2014		2015		Average deviation	2016 turnout estimate (d)
	Actual turnout	Deviation (a)	Actual turnout	Deviation (b)	Actual turnout	Deviation (c)		
A	45	+3	37	+1	69	+2	+2	72
B	40	-2	33	-3	63	-4	-3	67

(a) 2011 overall turnout = 42%

(b) 2014 = 36%

(c) 2015 = 67%

(d) 2016 overall turnout estimate = 70%

Having outlined the procedure for obtaining the likely division of each party's supporters into the leave/remain camps and the transposing of the 2015 general election support to the geography of the referendum counting areas we now consider how the forecast model works in practice.

The model uses two estimates derived from separate polling sources although more than two estimates may be used if required. The simple average of these two estimates provided the initial starting point for forecasting the likely referendum vote in each local authority.

When actual results are declared the model adjusts these starting positions once actual declarations are announced. First, in respect of turnout, if the actual aggregate turnout in those particular authorities that have declared is higher/lower than the expected aggregate figure then any difference between the estimate and actual turnout is used to re-calibrate turnout for remaining counting areas, adjusting the turnout estimate higher or lower accordingly. This process is repeated as each counting area announces its turnout.

As each counting authority announces its Leave/Remain vote similar adjustments are made to the estimates for those places yet to declare. Following the first declaration the actual leave vote is compared with the estimated vote for that area in order to assess the relative accuracy of the two polling derived estimates. If the result falls closer to one rather than the other then instead of taking a simple average of the two estimates the procedure now weights them differently, giving more power to the more accurate poll. This process continues as the number of declared results rises with weights being adjusted to take account of the closeness of actual votes and those estimated by the polling data. It is important to note that the comparisons between the actual and estimated figures are at the aggregate rather than individual counting area. Thus, for example, if fifty areas declare their result then the model compares the aggregate outcome in those cases with the estimated aggregate for the same places.

Some exceptions to this general rule were introduced to take account of different geographies. Firstly, the model makes no adjustments following declarations from either Gibraltar or Northern Ireland. This is because we were unable to make initial estimates from the polling data for these two counting areas. Secondly, it was anticipated from prior polling data that both Scotland and London would vote in different ways to the rest of the country. In each of these two exceptions only results from the same region were used to make adjustments to the estimates. By the same token results from both London and Scotland were not used to make adjustments to counting area estimates elsewhere across the rest of England and in Wales.

On the night

An important feature of the on the night broadcasting was to run different procedures in parallel as a sort of 'belt and braces' exercise. The involved data collected as part of the series of polls being conducted among Sky's customer base as well as survey evidence assembled by BMG research.

Sky Data is obtained from polling more than 10 million Sky customers as a form of research panel to conduct nationally representative surveys. We select samples with specified interlocking targets for gender, age and Mosaic group, conduct interviews online. Data are weighted by gender, age, mosaic, region, education, housing tenure, ethnicity, work sector and past vote.

Between May 2015 and June 2016 a total of 27,245 respondents responded across nineteen separate polls. We pooled these data and analysed the propensity to vote Leave or Remain in the EU referendum and reported likelihood to vote for members of each Mosaic group within each of five defined areas - North, South, Midlands, Wales, Scotland and London. Both Northern Ireland and Gibraltar were excluded. As others had done we assumed that people within each Mosaic group would vote in the same proportions for Remain and Leave across each area.

For each local authority, we multiplied the proportion within each mosaic group within each area saying they were certain to vote by the proportion of the population made up by each Mosaic group to forecast a likely turnout figure. We then multiplied this by the size of the electorate to forecast the overall number of voters in each local authority.

We then multiplied the proportion within each mosaic group within each area voting for Remain and Leave (excluding undecided voters) by the proportion of the population of each local authority made up by each Mosaic group to forecast likely Remain and Leave figures. We then multiplied each by the size of the electorate to forecast the overall number of votes for Remain and Leave in each local authority. As a cross check we compared these local authority estimates with those obtained by the method outlined earlier and found a high correlation.

Summing these figures, our aggregates predicted a turnout of 71% and overall results of Remain 48%, Leave 52%. We chose not to adjust these, as the predicted result reflected our final EU referendum polls, and the predicted turnout reflected higher levels of voting certainty in our polls than we found prior to the 2015 general election.

Between October 2015 and June 2016, BMG interviewed around 16,000 UK adults online and some 12,000 by telephone. For consistency the approach for both online and phone modes remained the same throughout the campaign and on Referendum Day. For both approaches BMG used interlocking weighting targets for gender, age, Government Office Region (GOR), Indices of Multiple Deprivation (IMD) quartile and the 2015 General Election result.

BMG's on-the-day poll was conducted between 1000 and 2100 hours and sampled 5,394 UK adults eligible to vote. The poll included voters, those who had voted already either by post or in-person, "pledges", those who said they would vote but had not yet done so at the time of interview, and

non-voters, for turnout estimation. There was approximately one fifth of those who voted or said they would vote but refused to say how. For these respondents BMG imputed how they may have voted based on key indicators of sentiment towards the EU. Regarding those who had voted and adding including those where the choice was imputed from other data the split was narrowly in favour of leave – 52% to 48%. Among those that had not yet voted and including imputation for those who refused to say the outcome was reversed – a narrow win for remain – 52% to 48%. The ratio of pledges to actual voters was 55% and 45% of the final sample respectively. When combined together in these proportions this produced a final rounded figure of Remain on 51% and Leave on 49%, with a Remain lead of 1.3%.

Overnight on June 23 there were, therefore, three separate modelling exercises operating, two of which would use the same procedure of producing estimates based on 2015 general election vote by European referendum vote after controlling for turnout. The third method was based on the Sky data and would adjust local authority estimates according to average error on the forecast – de facto a uniform swing model. As stated earlier the Sky data estimates were based on pooled survey data but the remaining two methods were based on recent polling and in the case of BMG an on the day telephone poll [details please]. For the Sky broadcast forecast we used estimates from two recent polls that indicated contrary outcomes – one pointing to a victory for leave, the other for remain.

Figure 4 shows the initial setting of the studio forecast model before any places had declared. The cases are sorted in rank order of leave vote per cent according to the first poll. The equivalent values for each authority according to the second set of polling data are also shown. The blue lines represent the simple average value for each local authority of the two estimates. The reason for the apparent 'noise' in the curves stems from the fact that for any two given polls there is likely to be differences in the estimates of the proportion of each party's general election vote and its distribution towards leave or remain.
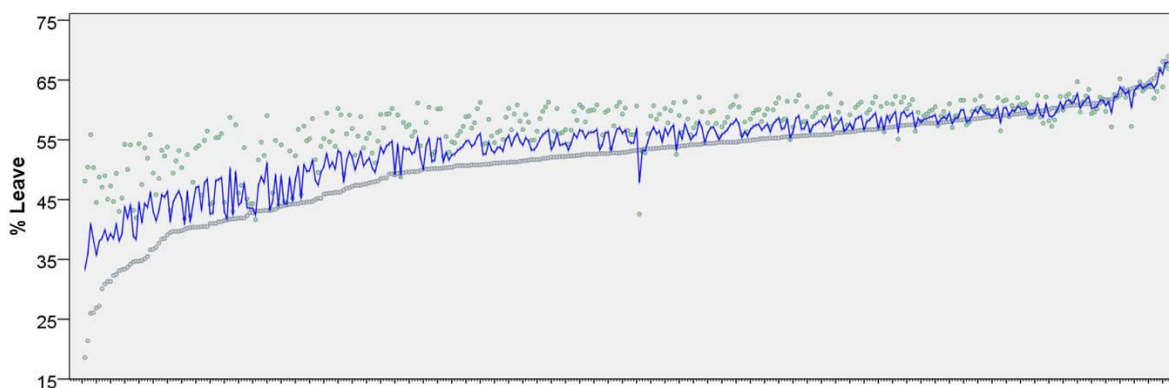

Figure 4: Initial starting point of local authority estimates on June 23

Before referendum results were declared it was expected that each counting area would announce the number of ballot papers issued. Although this number would include some votes that would be subsequently rejected (the national figure was only 25,359 rejected ballots) after three counting areas (excluding Gibraltar) announced provisional turnout the model re-calculated national turnout at 69.3% (the final figure is 72.2%). After a further twelve turnout declarations the model's national estimate was 72.3%. Since the SPSS output file contained details of expected national vote we were able to identify where the 'winning' line was likely to be.

After 17 declarations covering an overall electorate of 1.4 million and a total vote of just short of a million the forecast pointed to a leave victory of 55% to 45% for remain from an estimated vote of 32.8 million. These declarations were rather skewed including some very small counting areas and a relatively large proportion of results from the north east of England and Scottish authorities.

Table 3: Early declarations and percentage leave vote

Local Authority	Initial Estimate	Actual Result
Gibraltar	50.0	4.1
Newcastle-upon-Tyne	45.8	49.3
Orkney Islands	43.0	36.8
Clackmannanshire	44.0	42.2
Sunderland	54.5	61.3
Isles of Scilly	53.3	43.6
Swindon	57.3	54.7
Broxbourne	63.2	66.3
Kettering	60.0	61.0
Shetland Islands	42.8	43.5
South Tyneside	54.6	62.1
West Dunbartonshire	42.4	38.0
Dundee	38.1	40.2
Comhairle Nan Eilean Siar	40.9	44.8
East Ayrshire	43.9	41.4
Merthyr Tydfil	53.2	56.4
Stockton-on-Tees	55.9	61.7

When the number of declared results reached the fifty mark the model forecast was 54% for leave. This was based on 3.2 million votes. Expected turnout was forecast to reach 70%. It is interesting to delve a little deeper into how the model was reacting to the markedly different results from London compared to the rest of England and Wales. Figure 5 shows the model adjustments as results were being declared. Towards the left side of the distribution the red markers relate to a number of results from Scotland that were less favourable towards leave than each of our poll-based estimates. At the other end of the distribution, however, the actual support for leave was higher than that estimated. The graph also shows a small highlighted cross section which forms the basis for Figure 6 that highlights the cases of Stockport and Harrow. The original estimates for these two authorities, one located in Greater Manchester, the other a London borough, shows very little difference

between them in terms of the estimates where one set of polling data shows them at 45% while the other suggests a leave vote of about 55%. The simple average has the leave vote placed at 50%. However, as stated earlier, the model adjusts these starting positions as results are declared and it does so separately because all of the London boroughs are being modelled separately. In fact, the first London borough to declare (other than the extremely small electorate voting in the City of London area) was Lambeth at 02.20 a.m. with Wandsworth following five minutes later. Subsequently, after the first fifty declarations had been made the amended estimate for the leave vote in Stockport had risen to 57% (with no weight being given to the poll which placed it at 45%) but in Harrow it was revised downwards to 42% (with no weight given to the second poll).
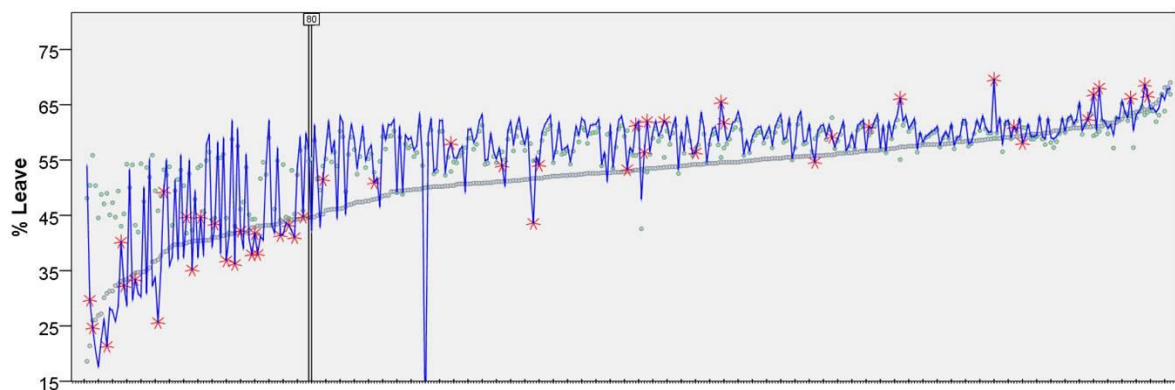

Figure 5: Model output after 50 declarations

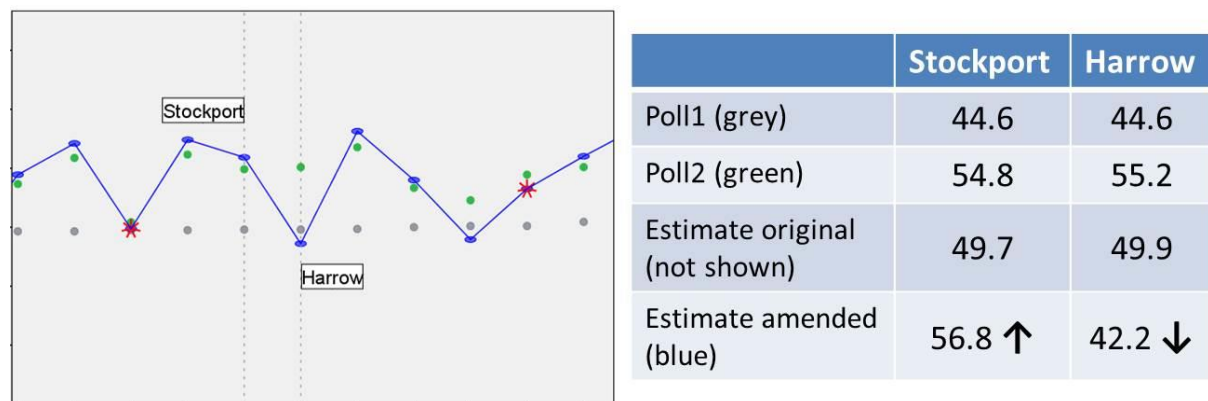|	Stockport	Harrow
Poll1 (grey)	44.6	44.6
Poll2 (green)	54.8	55.2
Estimate original (not shown)	49.7	49.9
Estimate amended (blue)	56.8 ↑	42.2 ↓

Figure 6: Revising estimates based on regional patterns of support for leave

The first on-air forecast was broadcast on Sky News at 02.15 hours, Thursday June 24 and put leave on 56%. This was followed an hour later with a lower figure of 53% for leave and a final forecast of 52% broadcast at 04.17 hours[7].

---

[7] In fact the PA News wires were reporting Sky News political analyst as predicting leave on 52% as early as 03.17 but this has not been verified.

It is instructive to observe the performance of the model that was being used by BMG. This used precisely the same method but using two separate polls to those being used at Sky HQ. Figure 7 uses the time stamp for the declarations although in some cases Sky News was ahead of the results being announced by the Press Association which BMG was wholly reliant upon. The red line denotes the model used at Sky with the blue one indicating the movement in BMG's modelling. There were few results declared in the first three hours after the polls closed but during this period the two lines climb towards a clear forecast for leave. Shortly after 2.00 a.m., however, the two lines move rapidly downwards. We believe that this is because it is around this time that a number of large electorate areas declare, including Glasgow and South Lanarkshire in Scotland, Caerphilly in Wales as well as Lambeth and Wandsworth in London. The BMG-based model was consistently running closer to the final outcome of 52% for leave before that team, with no broadcast obligations, stopped collecting actual results at about 04.30 a.m. The black line represents the third model which began with a set of estimates based on Sky data and corrected these estimates on the basis of average error on the estimate. This model starts with leave on 52% but then moves downwards because the first set of results had over-estimated the likely leave vote. However, it then veers back towards the leave decision and after about fifty results were declared it returns to around the 52% mark.
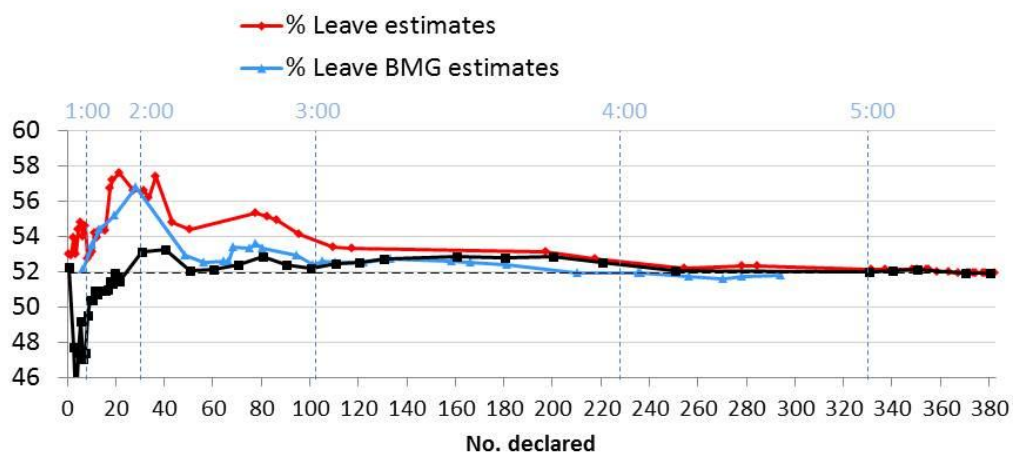

Figure 7: Comparing three models

From this it is tempting to ask why we should go to the effort of compiling estimates and build complex models that attempt to control for the complexity of results when a simple swing model works well enough given time. Indeed, as an exercise we took as our starting two simple measures of likely leave support for each local authority. The first was the level of support for Eurosceptic parties (UKIP and other anti EU parties) at the 2014 European parliament elections while the second value was based on our calculations of UKIP's vote share at the 2015 general election. Figure 9 compares how the model works with these simple starting parameters with the estimates derived from the polling data. Although there is some erratic movement at the outset the green line settles quite quickly into a clear forecast for leave and after 60 or so results are declared it is sending the same message that we received on the night of June 23. The most important conclusion to be drawn from using a basic model, however, is that while it arrives at the correct forecast given enough results it may be providing a somewhat erratic narrative before then which in the context of doing live broadcasting is not particularly helpful!

Conclusions

So, how well did the model perform both in terms of estimating the likely level of turnout (and hence the establishment of 'the winning line') and the level of support for leave (why everybody was watching). Figure 8a compares our estimate of turnout with actual turnout for each counting area. The initial estimate of national turnout was 61% and was based on survey respondents' stated likelihood of voting. Of course, once the first authorities declared their turnout the estimates were moved quickly in the right direction. The correlation between estimate and actual turnout is 0.78. Using only the general election turnout as a guide to likely referendum turnout the correlation falls to 0.65. This appears to justify the decision to use three measure of turnout and the mean deviation from the national average to calculate the estimates for each local authority.
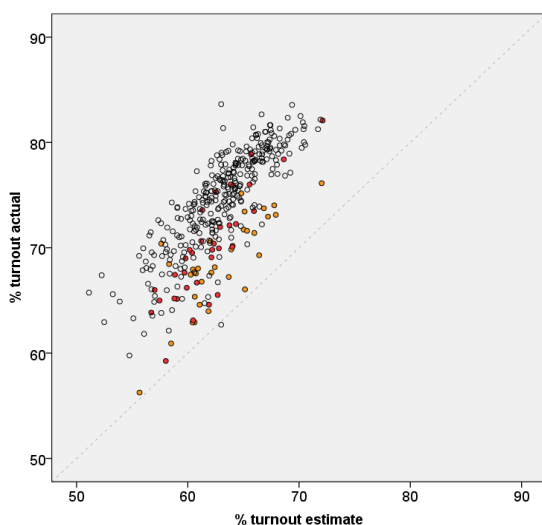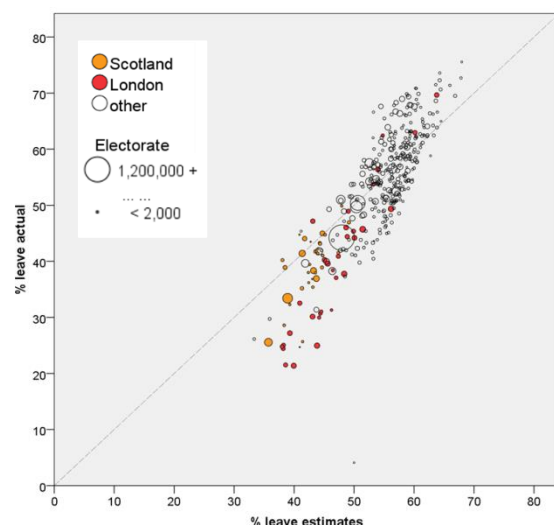

Figure 8 a: Turnout estimates and actual   *Figure 8b: Leave vote estimates and actual*

Figure 8b compares the percentage leave vote estimate with the actual outcome in the 382 counting areas. Each circle's size is related to the size of its electorate although they are not strictly to scale. The correlation between the estimate and actual leave vote is 0.87. It is interesting to note that among the circles away from the diagonal there are a number from Scotland (orange) and London (red). Obtaining reliable estimates for Scotland using our method for interpreting the survey data had always caused concern and these data appear to confirm that. Similarly, in the case of London there is a clear regional effect at work here and again one that could not be entirely captured beforehand using our method of devising estimates.

Overall, however, the model performed well, as it had done in the case of Scotland's independence referendum two years earlier. In the absence of an exit poll it provided initial estimates of what we might expect from each counting area as well as the variance in turnout that might occur. When the first declarations were made, first turnout followed by actual results, the model used this information to establish both the national turnout and crucially which side had won the referendum. Using the separate models in parallel assisted in the process and permitted an early and accurate forecast that Britain had voted to leave the European Union.