Model-Based Pre-Election Polling for National and Sub-National Outcomes in the US and UK

Benjamin E Lauderdale, London School of Economics*
Delia Bailey, YouGov
Jack Blumenau, University College London
Douglas Rivers, Stanford University and YouGov

May 15, 2019

Abstract

We describe a strategy for applying multilevel regression and post-stratification (MRP) methods to pre-election polling. Using a combination of contemporaneous polling, census data, past election polling, past election results, as well as other sources of information, we are able to construct probabilistic, internally consistent estimates of national vote and the sub-national electoral districts that determine seats or electoral votes in many electoral systems. We report on the performance of three applications of the general framework conducted and publicly released in advance of the 2016 UK Referendum on EU Membership, the 2016 US Presidential Election, and the 2017 UK General Election.

1 Introduction

Election polling in the US and UK has suffered several high profile failures in recent years, some real and some merely perceived. While the accuracy of polling in established democracies in national elections has in fact been roughly constant for the past half century (Jennings and Wlezien, 2018), there are nonetheless periodic, high profile polling misses (Sturgis et al., 2016; Rivers and Wells, 2015). This is hardly surprising given the quality of data used by pre-election polling, little of which comes from probability samples, and none of which involves high response rates. But even if these data quality problems were corrected, national polling does not always answer the questions of interest, which often concern electoral outcomes that depend on results in a very large number of sub-national units (states, electoral districts or constituencies) that cannot be polled individually at reasonable expense. Both the problems of sample representativeness and the problem of sub-national elections can be addressed by better modelling (Wang et al., 2014).

In this study, we describe our approach to applying the logic of multilevel regression and post-stratification (MRP; Gelman and Little, 1997; Park et al., 2004) to pre-election polling. We begin by describing the typical approaches taken to electoral polling and sub-national electoral

*Please send correspondence to b.e.lauderdale@lse.ac.uk

results prediction, and the benefits of addressing these problems in a common modelling framework. We then state the problem as a three-way decomposition: 1) estimating the distribution of vote choice conditional on turnout and demographic and political types, 2) estimating the distribution of turnout conditional on demographic and political types, and 3) estimating the population joint distribution of demographic and political types. This approach allows us to generate sub-national estimates for electoral quantities of interest and sub-population estimates for understanding patterns of vote intention and change across politically relevant groups. In practice, this approach may improve national vote share estimates as well, although this is difficult to test without application to a very large number of elections.

We report here on three applications of this analysis strategy—to the 2016 UK Referendum on EU Membership, the 2016 US Presidential Election, and the 2017 UK General Election—using estimates that we published online in advance of each event. While the details of performance vary across these applications, in general all three models performed well. The magnitude of the errors on the national vote share margin between the top two alternatives were all reasonable (2.6% EU 2016, 1.7% US 2016, 0.9% UK 2017), but we acknowledge that it is possible to get lucky on a single outcome quantity in three elections. The sub-national estimates successfully captured the novel patterns of voting in the referendum as well as non-uniform patterns of swings in the US and UK elections. In the EU referendum, we generated local authority level estimates that correlated with the results at r=0.92, despite the lack of any directly comparable past electoral results. In the US Presidential Election, the point estimate was for a narrow Clinton victory. However, the model correctly identified Trump's electoral college advantage—even though the electoral college advantage in 2012 was in Obama's favour—and therefore that Trump's most likely victory scenarios involved losing the national popular vote. In the UK general election, the model clearly outperformed other pre-election estimates (Hanretty, 2017) in predicting the winning party in each parliamentary constituency, and we also show that our predictions were more accurate than estimates produced after the election which apply national and regional uniform swings based on the true vote swings that were not known in advance of the election. The detailed performance assessments that we report below identify several areas of potential improvement, which we discuss more extensively in the conclusion.

2 Election Polling and Prediction

2.1 'Election fundamentals' models

Absent some kind of contemporaneous information about voting intention (whether polling or other kinds of more indirect signals) one can forecast election results by using models of the relationship between past election outcomes and "fundamental" economic and political conditions like economic growth, unemployment, and incumbent party tenure in office. There is a long tradition of these models, and several scholars employed regression models of this type in the run-up to the 2015 UK general election (Fisher and Lewis-Beck, 2016) and before the 2012 and 2016 US presidential elections (Campbell, 2012, 2016).

Unfortunately these models have severe and unavoidable limitations: with few historical data points (typically no more than 20) correctly specified prediction models necessarily provide

very uncertain predictions (Lauderdale and Linzer, 2015). Further, these fundamentals models typically target national-level vote shares as their main prediction quantity of interest, and only address subnational outcomes – the main focus for us – as a secondary concern.[1] While these models can form a prior expectation Linzer (2013) for analysis involving polling data, the latter is required in order to form predictions with a useful level of precision.

An alternative motivation for these models is that they can also be used to make statements about which factors appear to drive election outcomes, something that may be of substantive interest to political scientists. The MRP approach we describe below does not suggest which economic and political indicators best correlate with election outcomes at the aggregate level, but (as we describe in section 5.4) it does allow us to produce detailed descriptions of the ways in which patterns of turnout and political support amongst key demographic groups change between elections, something that is also often of great interest to political scientists.

2.2 Traditional electoral polling

Traditional electoral polling methods follow the logic of social surveys, despite the fact that electoral polling is almost never based on probability samples. Nearly all UK pollsters use methods that can be understood as involving some kind of more or less sophisticated quota sampling followed by some set of post-stratification adjustments (Sturgis et al., 2016). Practice in the US is more varied, but broadly follows the same procedures. First, a non-probability sample is collected by telephone or online. Second, a set of weights $W_i(X_i)$ are estimated that match the sample marginal distributions of measured covariates to known eligible voter population targets. This is typically done using raking/calibration (Sturgis et al., 2016). Third, turnout probability weights T_i are constructed from stated intention to turnout, for each sampled respondent. In practice, these are often binary, reflecting some simple cutoff rules in stated intention to turnout, but more generally can reflect best estimates of the probability that the particular respondent will turnout. Finally, the estimate of the national vote share for party k is formed by multiplying these two sets of weights by the probability V_{ik} that the respondent will vote for party k.

$$\frac{\sum_i V_{ik} T_i W_i(X_i)}{\sum_i T_i W_i(X_i)} \tag{1}$$

Like the turnout probability, these V_{ik} are usually binary assessments based simply on stated vote intention, but can reflect an estimate of the probability that the particular respondent will in fact vote for a given party. Within this general procedure, there is a variety of practice regarding exactly how each component of the above equation is generated from the polling response, particularly surrounding the treatment of people who say they do not know how they will vote, and people who express varying confidence that they will turnout.

The historical performance of election polling is mixed. Jennings and Wlezien (2018) show that the predictive performance of national polls in established democracies has been stable over the past half century, even as sampling methods and response rates have evolved. The magnitude of historical errors clearly indicates that there are typically systematic biases on top of sampling variability. Aside from accuracy concerns, one important limitation of these tradi-

[1] Though see Hummel and Rothschild (2014); Lauderdale and Linzer (2015).

tional methods is that they do not provide reliable estimates for public opinion in electorally relevant sub-national geographic units, unless separate polls are conducted in each sub-national geographic unit. This is usually prohibitively expensive. In the UK, for example, conducting a 1000 person poll in every parliamentary constituency would require polling 650,000 individuals. Further, because raking/calibration match marginal rather than joint distributions of characteristics, these methods also may not provide reliable estimates for non-geographic subpopulations (age groups, education levels, etc) that may be of interest in advance of an election.

2.3 Poll aggregation models

In contexts where many polls are conducted and published, several academic and non-academic researchers have demonstrated that it is possible to effectively aggregate those polls to produce ensemble estimates. Poll aggregation methods aim to correct for pollster-specific biases in order to estimate the true level of support for different voting alternatives (Jackman, 2005), and combine national-level information with sub-national information to construct estimates of the probability of different election outcomes (Silver, 2017; Linzer, 2013; Hanretty et al., 2016a). Poll averaging replaces the assumption that an individual pollster has unbiased procedures with an assumption that the average pollster is unbiased (Wright and Wright, 2018). The plausibility of this assumption varies from election to election, but recent elections have provided a series of high-profile cases of systematic polling bias in both the UK (Sturgis et al., 2016; Fisher and Lewis-Beck, 2016) and the US (Kennedy et al., 2018).

Nevertheless, in US presidential elections, poll aggregation has proved effective at translating the many state polls conducted by different pollsters into national estimates. While no pollster fields polls in every state, at least a few polls are conducted in most states over the period of the campaign, and many polls are conducted in competitive states. In contrast, in a UK election, with 650 constituencies and a short campaign period, constituency polling does not generally occur to any significant degree.[2] A large number of national polls are conducted, some with substantial sample sizes, but this still translates into small numbers of responses at the constituency level.

2.4 Multilevel regression and post-stratification

In recent years, researchers in statistics and political science have shown that it is possible to construct high quality sub-national (small area) estimates of public opinion using multilevel regression and post-stratification methods, typically referred to as MRP (Gelman and Little, 1997; Park et al., 2004). These methods rely on multilevel regression models to utilise small numbers of observations in each sub-national unit in order to discover patterns in opinion as a function of the demographic composition and other measurable features of those sub-national units. These patterns are then mapped out onto all sub-national units through post-stratification of fitted values from the model. There have been applications of this methodology to measuring public opinion on a variety of politically relevant geographies, including US states (Lax and Phillips, 2009), US congressional districts (Tausanovitch and Warshaw, 2013), German electoral districts

[2] In only one election cycle (2015) have a non-trivial number of constituency polls been conducted, and even then only about 20% of constituencies over the entire year preceding the election.

(Selb and Munzert, 2011), Swiss Cantons (Leemann and Wasserfallen, 2017), UK parliamentary constituencies (Hanretty et al., 2016b), and others.

These studies of public opinion have typically targeted the full adult population. Pre-election polling has the additional step of distinguishing voters and non-voters among the voting eligible population. The only published academic study that we know of applying MRP to pre-election polling is by (Wang et al., 2014). In their study of the 2012 US presidential election, those authors used an unusually large sample—750,000 responses—from an unusually unrepresentative data source—an XBox poll with a 93% male and 65% 18-29 years old sample—to estimate vote shares for Barack Obama and Mitt Romney in 50 US states plus the District of Columbia. While an impressive demonstration of using modelling to rescue low quality "big" data, this is well outside the scope of most pre-election polling, which involves much smaller, but less unrepresentative, samples. Further, it is not the most convincing illustration of the power of model-based methods for generating small area estimates, as that study used an average of 15,000 observations per state. Such massive sample sizes reduce the need to leverage patterns across sub-national units, which is a major potential upside to model-based approaches to analysing survey data.

In this paper, we use a much smaller, but higher quality, sample, to get estimates of the 2016 US presidential election that are comparable in quality to those Wang et al recovered for the 2012 US election and the most successful 2016 US polling aggregators. For the UK general election we use a smaller overall sample, and generate estimates for eight parties in 632 constituencies that were far more accurate than any other pre-election analysis as well uniform swing heuristics applied to the actual national or regional swings. The UK general election example particularly demonstrates the potential of MRP, because there are few viable alternatives to model-based methods in electoral systems with large numbers of first-past-the-post electoral districts.

3 Methods

3.1 Decomposition of the Problem

We denote the electoral alternative chosen by individual i as an unordered categorical variable $V_i \in 1, \ldots, K_{vote}$. In our examples, these correspond to {Leave, Remain}, {Clinton, Trump, Johnson, Stein, Other} or {Conservative, Labour, Liberal Democrat, etc}. Whether an eligible elector turns out to vote is a binary variable $T_i \in 0, 1$. Finally, a 'voter type' is a vector of measurable characteristics $X_i \in \mathcal{X}$ for an eligible voter i. These might include age, gender, education, vote in preceding elections, geographic location, etc. For each voter type, there are three important quantities that we would like to know:

1. Conditional voting distribution: $p(V_i|X_i, T_i = 1)$. What proportion of each type will vote for each of the alternatives among those who do vote?
2. Conditional turnout distribution: $p(T_i = 1|X_i)$. What proportion of each voter type will turn out to vote?
3. Poststratification frame: $f(X_i)$ or $p(X_i)$. How many or what proportion of eligible voters are of each type?

The proportion of voters turning out to vote is then:

$$\frac{\sum_{X_i \in \mathcal{X}} p(T_i = 1|X_i) f(X_i)}{\sum_{X_i \in \mathcal{X}} f(X_i)}$$

and the vote share for alternative k is:

$$\frac{\sum_{X_i \in \mathcal{X}} p(V_i = k|X_i, T_i = 1) p(T_i = 1|X_i) f(X_i)}{\sum_{X_i \in \mathcal{X}} p(T_i = 1|X_i) f(X_i)} \tag{2}$$

Note that we can sum over relevant subsets of types rather than all types $\in \mathcal{X}$, in order to calculate turnout or vote counts/shares on geographic or demographic subsets of electorate.

The crucial distinction between this approach, and the weighting approach followed by most polling, is seen by comparing Equations 1 and 2. In the weighting approach, the sum is over the sampled observations; in the modelling approach, the sum is over the set of voter types. Thus, when weighting, the key estimation step is the construction of weights $W_i(X_i)$ for each observation in the sample that match sample and population moments of a set of covariates for which the population distribution is known. In the modelling approach, the primary estimation exercise is to construct an outcome model $p(V_i = k|X_i, T_i = 1)$ that describes vote choice as a function of voter types, over which the population distribution is known.

This is not the only decomposition we could adopt for this problem. There are two simpler two-component decompositions that have been previously applied. The first of these is that used by Wang et al. (2014) when they post-stratify to an exit poll from the previous presidential election. The exit poll sample is a sample from $p(T_i = 1, X_i) = p(T_i = 1|X_i)p(X_i)$ at the last election, thus combining the second two components of our three-way decomposition into a single step of estimating the demographic distribution of those who will turnout, given the assumption that it will be the same as in the previous election. The other possible two-component decomposition treats non-voting as a voting choice category symmetrically with the voting alternatives, estimating $p(V_i, T_i = 1|X_i) = p(V_i = k|X_i, T_i = 1)p(T_i = 1|X_i)$ in a single step.

The reason that we explicitly specify the three-component decomposition here is that we separately model turnout rates and vote choices for demographic types, using different data sources. We do this because several of the problems facing pre-election non-probability samples are far more severe for turnout than for vote choice. First, turnout self-reports have well-known social desirability biases (Bernstein et al., 2001; Holbrook and Krosnick, 2010), while vote choice only sometimes has a significant social desirability bias. Second, not turning out to vote is associated with unit non-response to political surveys (Jackman and Spahn, ming), while vote choice is less systematically associated with unit non-response. These two points, taken together, suggest that typical pre-election surveys are likely to face more serious problems of representativeness and misreporting with respect to turnout than with respect to vote. Fortunately, while pre-election survey estimates of turnout are problematic, the demographics of turnout are usually broadly stable across elections. In this paper, we use post-election face-to-face surveys from the preceding comparable election to estimate turnout patterns at that election, and assume they will not change much in the present election.

There are risks associated with predicting turnout in the current election using demographic patterns of turnout from previous elections. If there are changes in the relative rate at which

different demographic or political groups turnout to vote across elections, then this could cause our estimates to be biased in unpredictable ways. This is therefore a potential weakness of our model. However, while our approach means that we cannot predict how turnout might be changing in the present election versus that preceding election, it also avoids the very large errors that can result from relying on uncalibrated prospective self-reports.[3] We provide more discussion of this point in the evaluation of our model below.

4 Applications

4.1 Political Context

The UK Referendum on EU Membership was held on 23 June 2016. The results of the referendum were reported for each of the 380 local authorities in England, Scotland and Wales (Great Britain), for Northern Ireland, and for Gibraltar. We modelled the 380 local authorities in Great Britain and added fixed priors for the relatively small number of votes in Northern Ireland and Gibraltar. The only electoral outcome of consequence was whether the national vote share for Leave was greater than 50% of the valid votes. 51.9% of votes were for Leave, a 3.8% margin of victory over Remain.

The 2016 US Presidential Election was held on 8 November. The results of the election were reported at the county level in nearly all of the 51 states, however the electoral outcomes relevant to determining the vote count in the electoral college were the state plurality winners, plus the plurality winners in the congressional districts of Maine and Nebraska. We modelled all states and congressional districts. Hillary Clinton won 48.2% of the national popular vote versus Donald Trump's 46.1%, a 2.1% national vote margin, however due to narrow Trump victories in several key states, he won states and districts awarding 306 electoral votes versus 232 for Clinton.[4]

The 2017 UK General Election was held on 8 June. The results of the election were reported at the level of the 650 parliamentary constituencies. Our modelling only concerned the 632 constituencies in Great Britain,[5] the remaining 18 seats in Northern Ireland elect MPs from a different set of parties. In Great Britain, the Conservative party received 43.5% of the vote to 41.0% for Labour, a national vote margin of 2.5%. The electoral outcome is determined by the plurality winner in each constituency. Among the parties competing in Great Britain, Conservatives won 317 seats (-13 versus 2015), Labour 262 (+30), the Scottish National Party 35 (-21), the Liberal Democrats 12 (+4), Plaid Cymru 4 (+1), Greens 1 (nc), UKIP 0 (-1) and Other 1 (nc).

All three of these votes were close relative to the magnitude of historical polling errors and also in the sense that the key electoral outcomes were uncertain before the election. Indeed all three had outcomes that were surprising to the majority of election observers and led to substantial financial market movements as the results became clear. We reported our estimates for the EU referendum online in the form of an article on YouGov's website[6] on June 21, and

[3] Implicit in this approach is that most consequential changes in results from election to election are due to changes in vote choices rather than changes in who turns out, or at least that these are the only changes we can reliably predict.

[4] These are the nominal electoral vote totals before seven members of the electoral college failed to vote for the candidate they were slated to vote for. The official record for each candidate is 304 for Trump and 227 for Clinton.

[5] Great Britain is used here as a shorthand for England, Scotland and Wales.

[6] `https://yougov.co.uk/news/2016/06/21/yougov-referendum-model/`

provided updates with the final days' data via Twitter. For the US presidential election, we posted daily updates on YouGov's website from 3 October until a final release on 7 November using data through 6 November.[7] For the UK general election, we posted daily updates on YouGov's website from 31 May until a final release on 7 June using data through 6 June.[8].

4.2 Frame Construction

Our information about the population joint distribution of demographic variables is captured in the form of a "frame", a rectangular dataset representing a very large sample of the population with micro-level data of the variables of interest, adjusted through weighting to maintain consistency with known marginal targets. The creation of a frame for each application is described in detail below. The input data files vary by application, but the overall logic of the construction process is similar across them.

First, a large face-to-face survey or census of the population is selected as the base frame, with micro data responses on basic demographics like age, gender, educational attainment, race and ethnicity, and geographic locale. This is typically a publicly available Census file, or an annual population survey which provides detailed information at low-level geographic areas. Then, additional geographies of interest (target) are imputed onto the base dataset at the lowest level geography available. The prior probability for the target geographic units is estimated by the proportion of population of the source geography within the target. The likelihood for a given respondent is estimated by published tables describing the joint distribution of demographics in the target geography if available, and assuming independence of published marginal tables otherwise. The new geographies are sampled from a multinomial distribution, with probability equal to the prior times the likelihood. Then, for political applications, voting behaviour is imputed onto the frame from a separate survey using multinomial regression. Finally, the frame base weights are raked to published vote totals, population demographics, and estimates of the electorate size.

In Table 1 we report the specific steps taken in each application, and the data sets used. The frames generated in this way are approximations to the true population distributions of the included demographic, geographic and past vote variables that they include. For nearly all variables, we have known marginal distributions of all variables at the geographic level we are interested in, or good approximations thereof. The information about the conditional distributions comes from a variety of sources, and is less reliable. Nonetheless, this information is important to include. Leemann and Wasserfallen (2017) show that in MRP applications like this one, using the product of the marginal distributions to define the post-stratification distribution yields identical estimates to the true joint distribution, so long as the model is linear and has no interactions (for logistic models the equivalence is approximate). Put differently, without information about the conditional distributions, we could only reliably apply a linear and additive model. Such a model would be unsatisfactory in these applications, there is very clear evidence of significant interactions in the vote choice models for all three applications.[9]

[7] https://today.yougov.com/us-election/

[8] https://yougov.co.uk/uk-general-election-2017/ When the Times published our initial estimates on 31 May that the Conservative party was likely to lose its majority, the value of the £ immediately declined by a half percent. http://www.bbc.co.uk/news/business-40101566

[9] For example, entering 2015 general election vote and 2016 referendum vote additively into the vote choice model

	2016 UK Referendum on EU Membership
0	Base file: 2011 UK Census Microdata Individual Safeguarded Sample (Local Authority)
1	Impute local authority district conditional on local authority group
2	Impute constituency conditional on local authority district
3	Impute 2015 general election turnout conditional on demographics (2015 BES face-to-face validated vote)
4	Impute 2015 general election vote choice for voters conditional on demographics and region (YouGov)
5	Rake to regional margins for constituency by vote and region by age by gender by qualifications
6	Post-stratify on vote, constituency and census marginals
	2016 US Presidential Election
0	Base file: 2012 American Community Survey (ACS)
1	Impute congressional district conditional on PUMA and demographics
2	Rake to ACS demographics and observed state level turnout
3	Impute registration and turnout conditional on demographics and state
4	Rake to demographics and turnout by state for 18+ citizens
5	Rake 2012 exit poll to ACS demographics and actual vote by state
6	Impute 2012 vote conditional on demographics by region
7	Rake to demographics and 2012 vote by state for voters
8	Rake to demographics and 2012 vote by congressional district for voters
9	Increment age by 4, apply survival weights
10	Impute education for 18-21 year olds
11	Impute registration for 18-21 year olds from the Current Population Survey (CPS) 2012
12	Rake to 2016 census population projections for demographics
	2017 UK General Election
0	Base file: 2016 Annual Population Survey (Jan - Dec)
1	Impute constituency conditional on region and demographics
2	Rake YouGov 2015 general election data to demographics and known vote totals at regional level
3	Rake YouGov EU referendum data to demographics and known vote totals at regional level
4	Impute 2015 turnout using pooled 2010 and 2015 BES face-to-face validated vote
5	Impute 2015 vote for voters using weighted YouGov data
6	Impute referendum vote conditional on 2015 vote + demographics using weighted YouGov data
7	Rake to constituency margins for demographics, 2015 vote, and estimated referendum vote

Table 1: Procedure followed to generate population frames for each application.

4.3 Turnout Model Specification

For each application, we estimated the conditional probability of turnout $p(T_i = 1|X_i)$ as a function of covariates using a face-to-face probability survey conducted after one or more prior election. For the EU referendum model, we used the 2015 British Election Study (BES) post-election survey observations for which vote validation was completed using the marked electoral register after the election. For the US presidential election model, we used the Current Population Survey (CPS) round completed after the 2012 US presidential election, relying on self-reported voter turnout. For the UK general election model, we pooled both the 2010 and 2015 BES post-election surveys, after verifying that the demographics of turnout in both surveys were largely similar. This yielded turnout model data sets for the three applications of 2955, 68167, and 6449 observations, respectively.

For all three applications, the turnout model took the form of a multilevel binary logistic regression model. For the EU referendum model and the US presidential election model, no survey weights were used; for the UK general election model, BES weights were used via a quasi-likelihood approach. The variables used in each model (including interactions) are listed in Table A2 in the appendix.

None of these data sets provided a wholly satisfactory measure of turnout in the election preceding the one being studied: for the BES data there is a self-reported recall of behaviour five years previous, for the CPS there is no measure at all. As a result, for all three data sets, we imputed turnout at the previous election in such a way as to not distort the demographic patterns of turnout in the data, while also yielding a high level of serial correlation in voter turnout. In each case, we randomly assigned previous election turnout in the turnout data set, conditional on reported/validated turnout in the observed election, such that the transition rates between turning out and not turning out matched either our priors (UK) or those from state voter files (US). These imputed values then became regressors in the turnout model, ensuring that our modelled electorate mostly (but not entirely) consisted of individuals who voted in the previous election. Once we fit these models to the relevant data, we then used the model to construct turnout probabilities/weights $p(T_i = 1|X_i)$ for each observation in the post-stratification frame.

4.4 Vote Choice Model Specification

For each application, we estimated the conditional probability of voting for each alternative $p(V_i|X_i, T_i = 1)$ as a function of covariates using data collected from YouGov's online panel. Respondents were selected from YouGov's panel on a daily basis, using YouGov's sample matching procedure (Rivers and Bailey, 2009).

The vote choice prompts were as follows. For the EU referendum, the question was "Should the United Kingdom remain a member of the European Union or leave the European Union?" with alternatives of "Remain a member of the European Union", "Leave the European Union", "Would not vote" and "Don't know". For the US election, among those who did indicate an intention to vote in a preceding question, we asked "Who will you vote for in the election for President in November?" with alternatives of "Hillary Clinton (Democrat)", "Donald Trump (Republican)",

for the 2017 UK general election does not fit the data as well as allowing the association with referendum preferences to vary by 2015 vote choice.

"Gary Johnson (Libertarian)", "Jill Stein (Green)", "Other", and "Not sure". For the UK general election, we provided a list of the candidates standing for election in the respondent's constituency, with the form "<Name> - <Party>", plus "Other", "Will not vote" and "Don't Know".

For purposes of modelling vote choice among voters, we excluded "Not sure" and "Don't Know" responses. For the US presidential model, there were five outcome categories: Clinton, Trump, Stein, Johnson and Other. For the UK general election model, there were eight outcome categories: Conservative, Labour, Liberal Democrat, UKIP, Green, SNP, PC and Other. We estimated separate models of the same form for England, Scotland and Wales, so not all eight outcome categories were used in any given model. We modelled the probability of voting for parties that were not standing in a respondent's constituency as zero. The variables used in each model (including interactions) are listed in Table A1 in the appendix. The models all include individual vote at the previous election plus interactions thereof that we deemed to be politically relevant in each case. For example, in the US Presidential election, we interacted 2012 election vote with race of the respondent, as we did not expect to see similar patterns of switching between parties for black and white respondents (on the logistic scale). For the UK general election, we interacted 2015 vote with constituency-level vote shares, as switching to and from different parties is predicted by the relative competitiveness of parties in a given constituency.

For all three applications, we used a relatively large time window of data, but modelled time trends within that data window. For the EU referendum and US presidential election we used a 14 day window, for the UK general election we used a 7 day window. Our final estimates in these three applications were based on 48738, 81246, and 55707 panelist responses, respectively.[10]

Once we fit these models to the relevant data, we then used the model to construct fitted vote choice probabilities $p(V_i|X_i, T_i = 1)$ for each observation in the post-stratification frame. When constructing fitted values for the purposes of post-stratification, we set the date variable to the most recent day, thus "adjusting" for time trends within the data window. Although we account for time trends within our survey data, we do not attempt to model how current levels of support are likely to evolve between the time of the survey and election day (as in, for example, Hanretty et al. (2016a)). Because of this, at the time they are produced, our estimates are best understood as forecasts of what the election result would be if the election were to be held immediately. Our method is therefore best suited for making predictions immediately before elections are in fact held. If our model was applied well in advance of the election, the accuracy of the approach would depend on how stable vote intention is over the election period.

Finally, in all applications, in order to capture our best guess of the scale of non-sampling errors that we could not explicitly model, we also added additional gaussian noise at the post-stratification stage, at both the national level and the sub-national level.

4.5 Aggregation

To form estimates, we generate turnout frequency weights by multiplying the population frame frequency weights by the fitted turnout probabilities for each observation in the post-stratification frame. The resulting turnout frequency weights are effectively a post-stratification frame for voters, rather than the electorate. We then multiply these turnout frequency weights by the fitted

[10] This approach can be used with smaller sample sizes, but a far more parsimonious vote model would have the be specified.

	Vote Choice	Result	Estimate	Low	High
EU Referendum	Leave	51.9	50.6	48.8	52.4
	Remain	48.1	49.4	47.6	51.2
US Presidential	Trump	46.1	44.1	43.0	45.2
	Clinton	48.2	47.9	46.8	49.1
UK General	Conservative	43.4	41.6	39.2	43.9
	Labour	41.0	38.2	36.1	40.6
	Liberal Democrat	7.6	9.0	7.9	10.3
	UKIP	1.9	3.5	2.9	4.1
	Green	1.7	2.0	1.7	2.4
	SNP	3.1	3.8	3.4	4.2
	Plaid Cymru	0.5	0.5	0.4	0.6

Table 2: National vote shares for major alternatives with mean posterior estimates and 95% predictive interval lower and upper bound.

vote choice probabilities for each observation in the post-stratification frame to get the estimated breakdown of each frame observation across the available vote choices. This can then be aggregated to the national level, the sub-national level, or by any other demographic variable that we have in the post-stratification frame in order to form estimates for that level of aggregation. We saved full posterior distributions for the electorally relevant sub-national aggregates by post-stratifying at each iteration of the MCMC simulation.

We implemented the turnout model, the vote model and the post-stratification in Stan (Carpenter et al., 2017). Given that we were posting estimates daily, and several aspects of the model were computationally expensive (large sample sizes, many parameters and multinomial outcomes) we used multiple, relatively short simulation chains estimated in parallel. The referendum model used four chains of 500 iterations (250 iteration initial burn-in discarded), the presidential model used 36 chains of 25 iterations (25 iteration burn-in), and the general election model used 4 chains of 125 iterations (75 iteration burn-in). Despite the extremely short chains used for the US election, and the modest samples overall, we did not see evidence of instability across runs. All three applications had estimation times that were roughly eight hours, which was our target to facilitate daily updates. Shorter estimation times for equivalent models may have been possible using alternative parameterisations that yield better performance in Stan.

5 Results

5.1 National Vote Shares

Figure 1 shows the election results in comparison to the posterior distributions for the key electoral margin in each election (Leave - Remain; Trump - Clinton; Conservative - Labour). In all three cases, the national margins fall in a region of the posterior distribution with a reasonable level of density, suggesting that our national-level uncertainty in the margin was plausibly calibrated, at least given what we can learn from three results. Table 2 shows the election results in comparison to the mean posterior and central 95% posterior interval for the voting alterna-

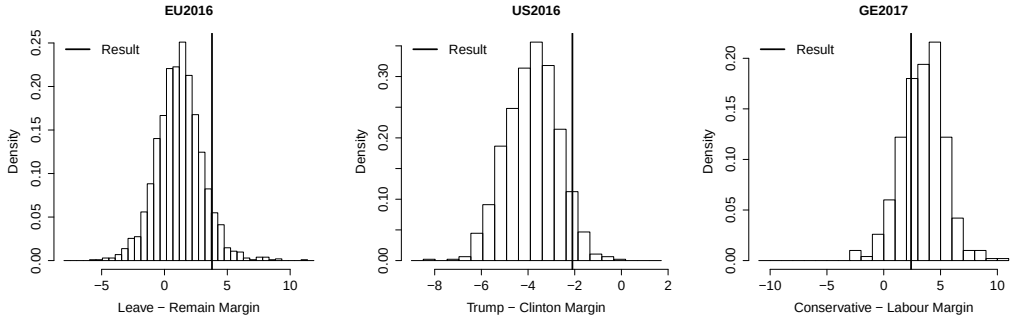

Figure 1: Posterior predictive distributions and results for national vote share margin, in the referendum (left), the presidential election (center), and in the general election (right).

tives. The largest error for any individual party in any of the applications was Labour, which we underestimated by 2.8 percentage points. In contrast to the national margins, many of the party vote shares fell outside the interval estimates. In both the US and UK elections this was due to the "major parties" overperforming the estimates while the minor parties generally underperformed. This may reflect a more general problem in polling of first-past-the-post electoral systems, where tactical decisions to vote for major parties may not be expressed to pollsters.

5.2 Sub-National Vote Shares

While for the EU Referendum, the national vote totals were the only relevant electoral outcome, for the US Presidential Election and the UK General Election, it is the sub-national vote totals that matter. In the US presidential election, the plurality winner in each state (plus those in Maine and Nebraska's congressional districts) determines the distribution of electoral college votes and thus the winner of the election. In the UK general election, the plurality winner in each of the 650 constituencies is seated in parliament. The posterior distribution of our model over vote shares for each candidate in each of these sub-national geographies thus implies a posterior distribution over these electoral outcomes.

Figure 2 shows the posterior distributions of (nominal) electoral college votes for Clinton in the 2016 US presidential election and seats for Con, Lab, LD, SNP and PC in the 2017 UK general election. The number of electoral votes secured by Clinton was at the very low end of the range we observed in our posterior sample, owing to her surprising loss of Michigan combined with her losing all of the predicted close states. Nonetheless, her result was not outside the range of our simulated election results. The seat totals for all the UK parties were generally located within the posterior distributions we estimated, with the largest surprise being Plaid Cymru's single seat gain.[11]

Looking at the vote share predictions in each reporting electoral unit, our estimates broadly tracked the results in the relevant sub-national areas in all three applications (Fig 3), however we see significant undercoverage of our interval estimates and some degree of attenuation bias. In all three cases we underestimated voting alternatives where they were strong and overestimate

[11]The PC went from 3 to 4 seats by gaining Ceredigion constituency by a margin of 104 votes or 0.2% The Ceredigion constituency has four competitive parties, which is extremely unusual, and thus is likely to deviate from the patterns of voting for those parties in other constituencies.

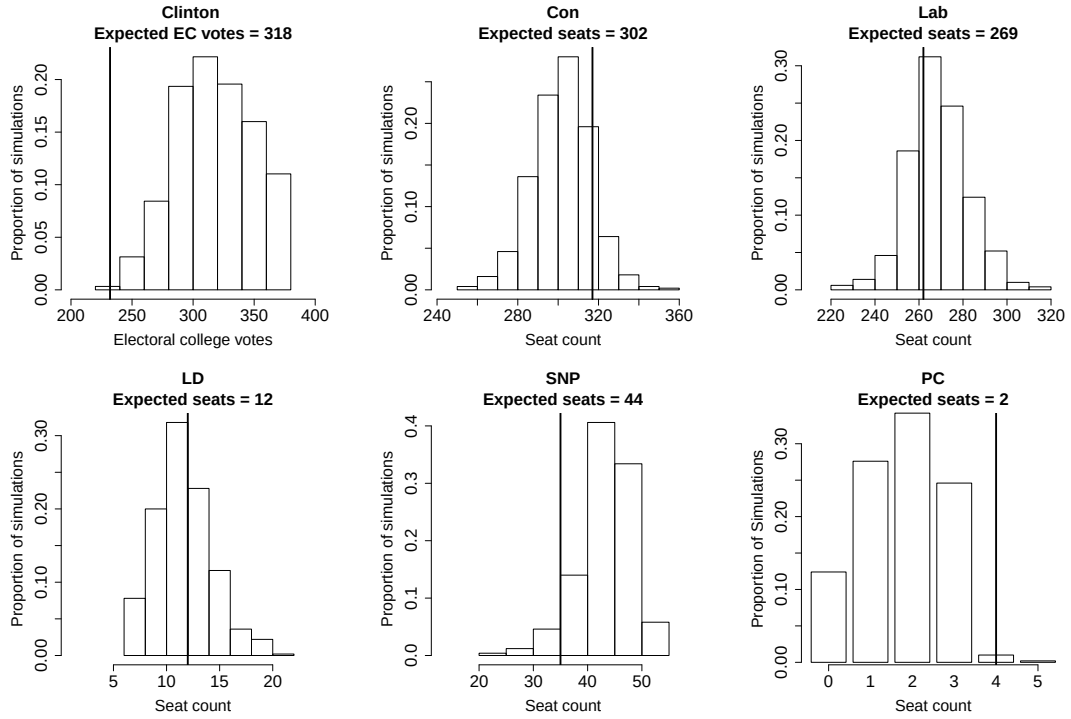

Figure 2: Posterior distributions of electoral votes in the US presidential election (top left) and for major and minor parties in the UK general election, with actual results depicted by a vertical line.

them where they were weak. A simple linear regression estimate of the marginal change in our prediction as a function of the marginal change in the result is 0.75 for the referendum, 0.79 for the presidential election, and 0.80 for the general election.

We suspect there were a few reasons for the general pattern of attenuation bias, as well as the differences across applications.[12] The first reason is that in general random effects and multilevel models tend to have this kind of attenuation bias. Such models reduce estimation variance, but are biased towards the overall mean in their predictions. The second reason is that individual-level behaviour is not fully explained by individual-level characteristics: individuals in politically extreme places vote differently from those in moderate places, even conditional on their observed characteristics and lagged vote. This meant that at the individual-level, the patterns of switching among supporters of each party were very different in different states. In the US model, we included an interaction of congressional district 2012 vote share and individual 2012 vote to try to capture this. In the UK general election model, we added an interaction of constituency vote share with individual lagged referendum and 2015 vote to enable to the model to discover this kind of pattern, which may have helped reduce (but not eliminate) attenuation bias. Because the referendum was cross-cutting with respect to prior election results, these contextual effects were probably smaller in that application. One of our key conclusions from these applications is that careful model specification is essential in applying MRP to pre-election polling, additive models without interactions involving political context performed less well in

[12] Wright and Wright (2018) show that almost all pre-election forecasts in the US in the 2016 election were subject to similar levels of attenuation bias.

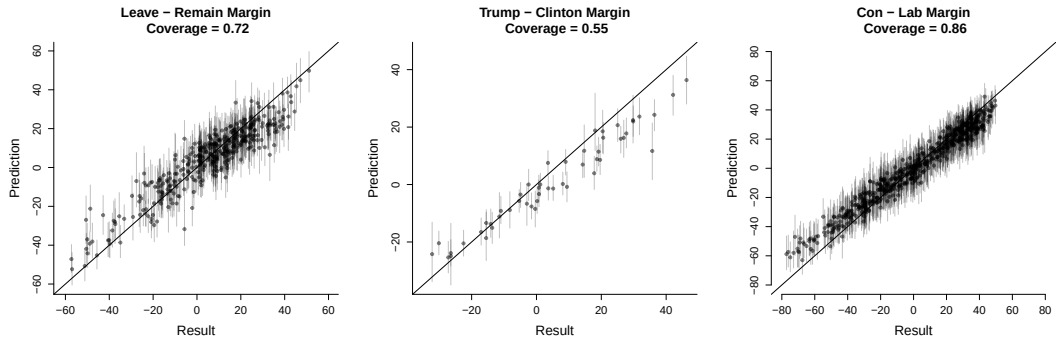

Figure 3: Predicted margin by actual margin, for local authorities in the referendum (left), states in the presidential election (center), and constituencies in the general election (right).

our model testing.

We can report some comparisons of fit to similar applications. Preceding the EU referendum, Chris Hanretty published local authority level estimates before the election based on an alternative multilevel regression (but not post-stratification) strategy, which generated estimates that correlated with both our estimates and the results at the same r=0.92. His estimates only aimed to generate the relative levels of the local authority reporting areas in a hypothetical 50-50 referendum, not the absolute level of support.

As previously discussed, Wang et al. (2014) used a multilevel regression and post-stratification approach in 2012 to map responses from a survey conducted on XboxLive onto electoral outcomes in 51 Electoral College races (excluding the Maine and Nebraska congressional district races). They report that the "mean and median absolute errors of our [Obama vote share] estimates on the day before the election are just 2.5 and 1.8 percentage points, respectively." When we assess our state-level estimates by these two metrics, using Clinton vote share, we calculate absolute errors of 2.5 and 1.9 percentage points, respectively. The 2012 election was a much easier election for prediction than 2016, as changes in relative state-level results were smaller between 2008 and 2012 than between 2012 and 2016.

For 2016, we can compare our estimates to those produced by a number of forecasters (table 3), most of whom pooled all (or most) publicly released state and national polls. We use the root-mean-squared error (RMSE) as our main evaluation metric and, for the 2016 US election, we focus on the RMSE for the Democrat-Republican margin.[13] Our RMSE was similar to, but slightly worse than, those of FiveThirtyEight.com, the New York Times, and the Princeton Election Consortium (PEC).[14] Since these forecasters pooled all publicly available state and national polls, this implies that the informational content of our MRP analysis was nearly that of all other published polling for the election. All of these forecasts and our model provided a slight improvement on the RMSE that would have resulted from applying the correct 2012-2016 national vote share swing to the 2012 state-level margins, which indicates that all of these analyses were able to recover some information about relative state-level movements versus 2012.

[13] We prefer RMSE over the mean or median absolute error used by Wang et al. (2014) because it enables comparisons to sampling variability as well as bias-variance decompositions.

[14] We do not report assessments of the probabilities of state-level victory because there is no standard way to assess the correlation of the state-level outcomes. For example, by Brier score, the PEC is among the best forecasts, but in fact the PEC indicated that a Clinton victory was certain, because (like the Brier score metric) it failed to take into account the possibility that state-level errors would be correlated.

Table 3: Comparison to uniform swing model and forecasts based on aggregations of state-level polling (RMSE) for the 2016 US Presidential election.

Model	RMSE (Top Two Margin)
538 (polls plus)	7.0
Princeton Election Consortium	7.0
New York Times	7.0
538 (polls only)	7.1
YouGov	7.3
UNS	7.5
PollSavvy	8.0
2012 results	8.1
HuffPost	10.7

Table 4: Comparison to uniform swing models and Hanretty forecast (RMSE and % correctly predicted) for the 2017 UK general election.

Model	Con	Lab	LD	UKIP	Green	SNP	PC	Other	% correct
YouGov model	4.4	4.7	2.5	2.4	0.9	3.3	3.6	1.6	92.9
Uniform swing (Regional)	4.6	4.1	3.6	3.8	1.9	3.8	2.9	1.9	91.6
Uniform swing (Country)	5.4	4.1	3.8	4.3	2	3.8	2.9	1.9	91.8
Uniform swing (GB)	5.9	4.7	3.8	3.8	1.8	11.9	3.3	1.9	91.1
2015 results	8.3	10.6	3.9	12.4	3	13.6	3.3	1.9	89.7
Hanretty	5.3	6.1	3.7	1.9	1.9	4.7	4.1	2.3	86.2

For the UK election, we provide comparisons with both the pre-election estimates produced by Chris Hanretty (Hanretty, 2017) pooling public national and regional polls, and also with estimates constructed after the election which apply uniform swings to the party vote shares based on the actual election swings from 2015 to 2017. We use three different measures of swing for this comparison: at the national level (Great Britain), at the country level (England, Scotland and Wales), and at the regional level. Table 4 presents the party-specific RMSE and percentage of constituency winners correctly predicted for each of these approaches. Compared to Hanretty's pre-election forecast, our model has a lower RMSE for all parties except for UKIP, and we correctly predict 92.9% of constituency results compared with 86.2% for the Hanretty model. Our model has lower RMSEs for most parties and correctly predicts a larger percentage of constituency results versus any of the uniform swing models, even though the uniform swing models use true swings that were not known in advance of the election. This indicates that our model was able to measure variation in swings across different kinds of constituencies, even within UK regions. One illustration of this was the model's ability to correctly predict gains for both major parties: Labour gains in several seats that had been held by the Conservatives for decades (Kensington, Canterbury) and also Conservative gains from Labour in constituencies the latter had held for decades (Middlesbrough South and East Cleveland, Stoke-on-Trent South). A further comparison that we can make is to classification accuracy that one would obtain if one were to predict that each constituency were retained by the incumbent party. As table 4 makes clear, we also correctly classify a greater number of constituencies (92.9%) than if we were to use the 2015 results alone (89.7%).

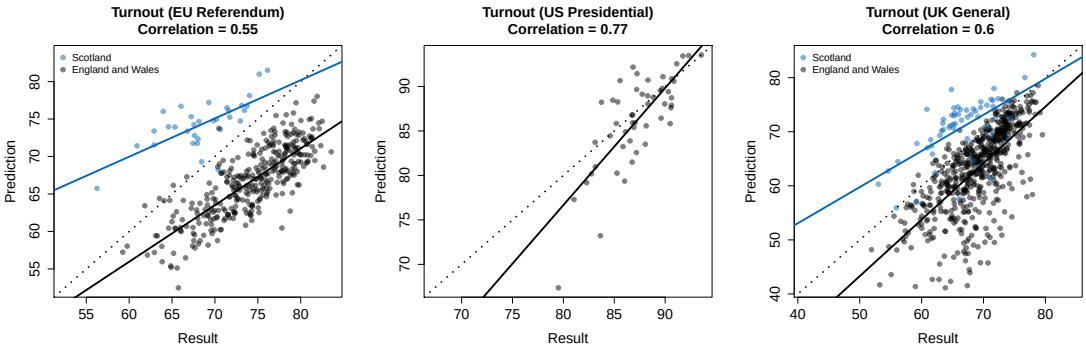

Figure 4: Predicted turnout by actual turnout among the registered electorate, for local authorities in the EU referendum (left), states in the US presidential election (center), and constituencies in the UK general election (right).

5.3 Sub-National Turnout

We have only indirect ways of assessing the performance of our turnout model. To assess individual-level turnout, we would have to rely on exit poll data, which is variably available across the different cases and can be problematic. We can, however, look at the extent to which the turnout rates that we expected at the sub-national level match the observed turnout rates at that level. In all three applications we see moderate correlations, with some indirect evidence that turnout was responsible for some prediction error.

In both the EU referendum and the UK general election, the turnout estimates had one obvious source of error: the relative turnout in Scotland was down and the relative turnout everywhere else in the UK was up relative to the 2015 general election. Given the SNP's takeover of almost all Scottish parliamentary seats in 2015 this turnout shift was perhaps not surprising, but we made no explicit efforts to model it in either case. Within Scotland, and within the rest of the UK, the turnout model performed decently at estimating relative turnout rates, suggesting the the demographic patterns of relative turnout were not substantially changed other than this national-level discrepancy. For the referendum, the estimated Leave share rises from 50.6 in our pre-election estimates to 51.0 if we replace the turnout estimates with the true turnout rates for each local authority, reducing the error on the margin from 2.6% to 1.8%. For the 2017 general election, the same calculation moves our estimate of Con 41.6 - Lab 38.2 (Con +3.4) to Con 41.2 - Lab 39.0 (Con +2.2). This is a better match to the true margin between the two parties (Con +2.5) because turnout was indeed up in Labour strongholds in relative terms, however the levels of support are still too low for both major parties. In the US presidential election, the corresponding analysis is less informative because of the smaller number of sub-national units, and the national vote margin only changes by 0.1% when using the true state turnouts.

There is no clear pattern across the three applications, and as noted above this is a very indirect test of whether the turnout model was responsible for the errors we see. The real risk is not so much mispredicting the relative turnout across electoral units, but rather mispredicting the relative turnout across groups within electoral units. With data from the 2016 CPS in the US and the 2017 British Election study in the UK equivalent to the data sets we used to fit the turnout model, we could evaluate the performance of the model if we use those data in place

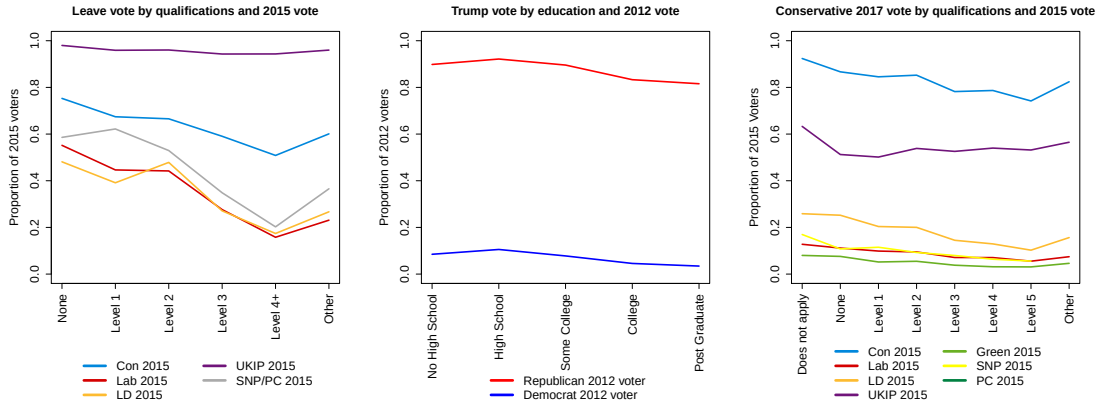

Figure 5: Vote choice by qualifications/education and previous election vote in the EU referendum (left), in the US presidential election (center), and in the UK general election (right).

of the data from the preceding elections. However, as we noted when we set out our estimation strategy, the question is not whether such errors occurred, but whether it is possible to do better prospectively by using self-reported likelihood to turnout measures. This is an important area for future development.

5.4 Demographic Patterns

While our focus here has been on the performance of these methods for electoral prediction, a major benefit of MRP for pre-election polling is that it can reveal politically important demographic shifts in voting patterns that are occurring in the electorate. A common theme in these three elections were shifts in voting by age and by education. Here we show the interaction of education with vote in the preceding election. In the EU referendum, educational qualifications were very strongly predictive of referendum vote within supporters of a 2015 party. Aside from 2015 UKIP voters, who supported Leave at very high rates regardless of education level, among 2015 supporters of all other parties, there are very large differences in support for Leave versus Remain by education. Among 2015 Conservative voters, the difference between those with no qualifications (less than GCSEs or equiv) support for Leave was about 20 points higher than among those with Level 4+ (BA or higher qualifications). For Labour, Liberal Democrat and SNP/PC 2015 supporters, these differences were even larger, reaching 30-40 points. In the US presidential and UK general elections, education predicted patterns of switching to a considerable extent. Trump disproportionately gained low education Obama voters and retained low education Romney voters. In the 2017 general election, the Conservatives retained about 90% of their 2015 voters with lower levels of qualifications, but only 80% of those with higher levels of qualifications. They also gained larger shares of the Liberal Democrat and Labour voters with low levels of qualifications than those with high.

6 Conclusion

As we have demonstrated, using MRP to conduct pre-election polling estimates for sub-national electoral units is promising, but there are several remaining challenges highlighted by the performance of the models that we document. First, there is a general problem of attenuation bias: the tendency to underpredict voting alternatives where they are strong and overpredict them where they are weak. We were able to partially mitigate this through the use of cross-level interactions of individual and constituency/district vote, with varying success across applications. Fortunately, attenuation bias tends to lead to larger errors in less competitive electoral units, and to balance out in the aggregate national estimates. Second, we see undercoverage of the constituency/district level results in all our applications. The difficulty of incorporating uncertainty due to non-sampling errors into the model estimation makes it difficult to generate properly calibrated interval estimates, even when error magnitudes are not large. Third, a major limitation of our strategy is that we do not attempt to model shifts in turnout patterns at either the individual or aggregate level. While this avoids large errors in the modelled voting population, and we believe there are good reasons to be skeptical of naive use of self-reported likelihood to turnout, given suitably rich panel data on likelihood to vote, combined with validated turnout data, it should be possible to improve on this strategy.

We have not provided explicit performance comparisons of the approach we describe versus classical methods for adjustment and turnout modelling. This is because it is difficult to do so in a way that is fair. A comparison on the basis of the national vote totals provides three meaningful data points, and it is easy for either our methods or standard methods to get lucky through counterbalancing sources of error. Further, if the same information on the marginal distributions of the same variables are used for weighting as we used for our post-stratification frame, the aggregate estimates will be nearly identical. A comparison on the basis of sub-national numbers will favour our method because classical methods are not designed to produce these estimates. Our primary aim is not to demonstrate that weighting based methods are worse than model based methods, but rather to demonstrate how MRP methods can be applied to pre-election polling and that they can provide sufficiently high quality sub-national vote share estimates in order to model electoral outcomes in a variety of electoral systems.

Comparing the two elections where the sub-national units matter for electoral outcomes, it is clear that our strategy performed better in the UK general election than in the US presidential election. This may be for idiosyncratic reasons having to do with the elections and how we constructed the models. However it may be that constructing good estimates for 51 states is actually more difficult than for 632 constituencies, because the latter support more covariates at the second level of the multilevel regression and provide greater opportunity for errors to counter-balance. The widely varying state electoral votes make close US elections extremely sensitive to a few errant state results. Even though we had far more data per US state than per UK constituency, the difficulty of modelling was also greater, as well as the sensitivity of the aggregate prediction to sub-national prediction errors. The situations in which our approach is likely to be most successful relative to alternatives are national legislative elections where there are a large number of sub-national electoral units, such that polling them individually is infeasible, but where there are likely to be systematic patterns across those units like those we

identified in the UK election case. Elections to the US House of Representatives, the UK House of Commons, the Canadian House of Commons, the Australian House of Representatives, and others fit this structure, although the details of polling and modelling an instant runoff election (as in Australia) are more complicated. The advantages to MRP approaches are more limited in proportional systems where the exact distribution of votes across areas is less important, although the structure we propose is still useful in those contexts for correcting sample bias.

In general, while the setup costs of moving pre-election polling to a model-based approach are substantial, in these cases we believe that the payoffs were also substantial. In each application the errors of the national vote share estimates were low, and our estimates were close to the key results in what turned out to be a close vote. In the two UK cases, the model got the key national electoral outcome right, and in the US case identified that a Trump electoral college victory would be combined with Clinton winning a narrow popular plurality. While the state-level estimates for the US election suffered from significant attenuation bias that needs to be better addressed, the constituency-level estimates for the UK general election outperformed all benchmarks. Indeed, they were the only pre-election seat forecast that correctly indicated that the Conservatives would fail to secure a majority.

Table A1: Variables included in the vote choice models for each application. The number of levels is given in brackets for categorical variables.

2016 EU Referendum	2016 US Presidential Election	2017 UK General Election
I – GE2015 vote [7]	**I** – 2012 vote [5]	**I** – GE15 vote [9]
I – Qualifications [6]	**I** – Qualifications [5]	**I** – Qualifications [8]
I – Age [14]	**I** – Age [15]	**I** – Age [14]
I – Gender [2]	**I** – Gender [2]	**I** – Gender [2]
I – Days ago [14]	**I** – Days ago [14]	**I** – Days ago [7]
	I – Race [4]	**I** – Political Attention [8]
	I – Marital status [5]	**I** – EU16 vote [3]
C – Constituency [632]	**D** – Congressional District [436]	**C** – Constituency [632]
C – Region [11]	**S** – State [51]	**C** – Region [11]
C – Population density	**S** – Region [9]	**C** – Incumbency [3]
C – % EU passport	**D** – District 2012 vote	**C** – Standing [2]
C – % Born in UK	**S** – State 2012 vote	**C** – Incumbent EU16 position [2]
C – % Christian		**C** – % 'Leave' 2016
C – % Muslim		**C** – % Long term unemployed
C – % Industry agriculture		**C** – % Industry manufacturing
C – % Degree		**C** – Population density
C – % Retired		**C** – GE15 share$_p$
C – % Asian		
C – % Black		
C – % Employed		
C – % Long term unemployed		
C – Deprivation index		
C – UKIP 2015 share		
I * I – GE15 vote * Qualifications	**I * D** – 2012 vote * District 2012 vote	**I * I** – EU16 vote * GE15 vote
I * I – GE15 vote * Age	**I * I** – 2012 vote * Days ago	**I * I** – Age * GE15 vote
	I * S – 2012 vote * Region	**C * I * I** – GE15 share$_p$ * GE15 vote * EU16 vote
	I * I – 2012 vote * Qualifications	**C * I * I** – GE15 share$_p^2$ * GE15 vote * EU16 vote
	I * I – 2012 vote * Race	
	I * S – Race * State	
	I * S – Race * Region	
	I * I – Race * Gender	
	I * I – Race * Educ	
	I * I – Race * Age	
	I * I – Qualifications * Age	
	I * I – Qualifications * Gender	
	I * S – Gender * Region	

Note: **I** = Individual-level variable; **C** = Constituency-level variable; **D** = Congressional district-level variable; **S** = State-level variable

Table A2: Variables included in the turnout models for each application. The number of levels is given in brackets for categorical variables.

2016 EU Referendum	2016 US Presidential Election	2017 UK General Election
I – GE2015 vote [7]	**I** – 2012 turnout [3]	**I** – GE15 turnout [3]
I – Qualifications [6]	**I** – Education [5]	**I** – Qualifications [8]
I – Age [14]	**I** – Age [15]	**I** – Age [14]
I – Gender [2]	**I** – Gender [2]	**I** – Gender [2]
	I – Marital status [5]	**I** – Political Attention [8]
	I – Race [4]	**I** – EU16 turnout [3]
C – Constituency [632]	**S** – State [51]	**C** – Constituency [632]
C – Region [11]		**C** – Region [11]
C – Population density		**C** – Population density
C – % EU passport		
C – % Born in UK		
C – % Christian		
C – % Muslim		
C – % Industry agriculture		
C – % Degree		
C – % Retired		
C – % Asian		
C – % Black		
C – % Employed		
C – % Long term unemployed		
C – UKIP 2015 share		
C – Deprivation index		
	I * I – Age * Qualifications	**I * I** – Age * Qualifications
	I * I – Race * Gender	
	I * I – Race * Qualifications	
	I * I – Race * Age	
	I * S – Race * State	
	I * I – State * Qualifications	
	I * I – State * Age	
	I * I – State * Gender	

Note: **I** = Individual-level variable; **C** = Constituency-level variable

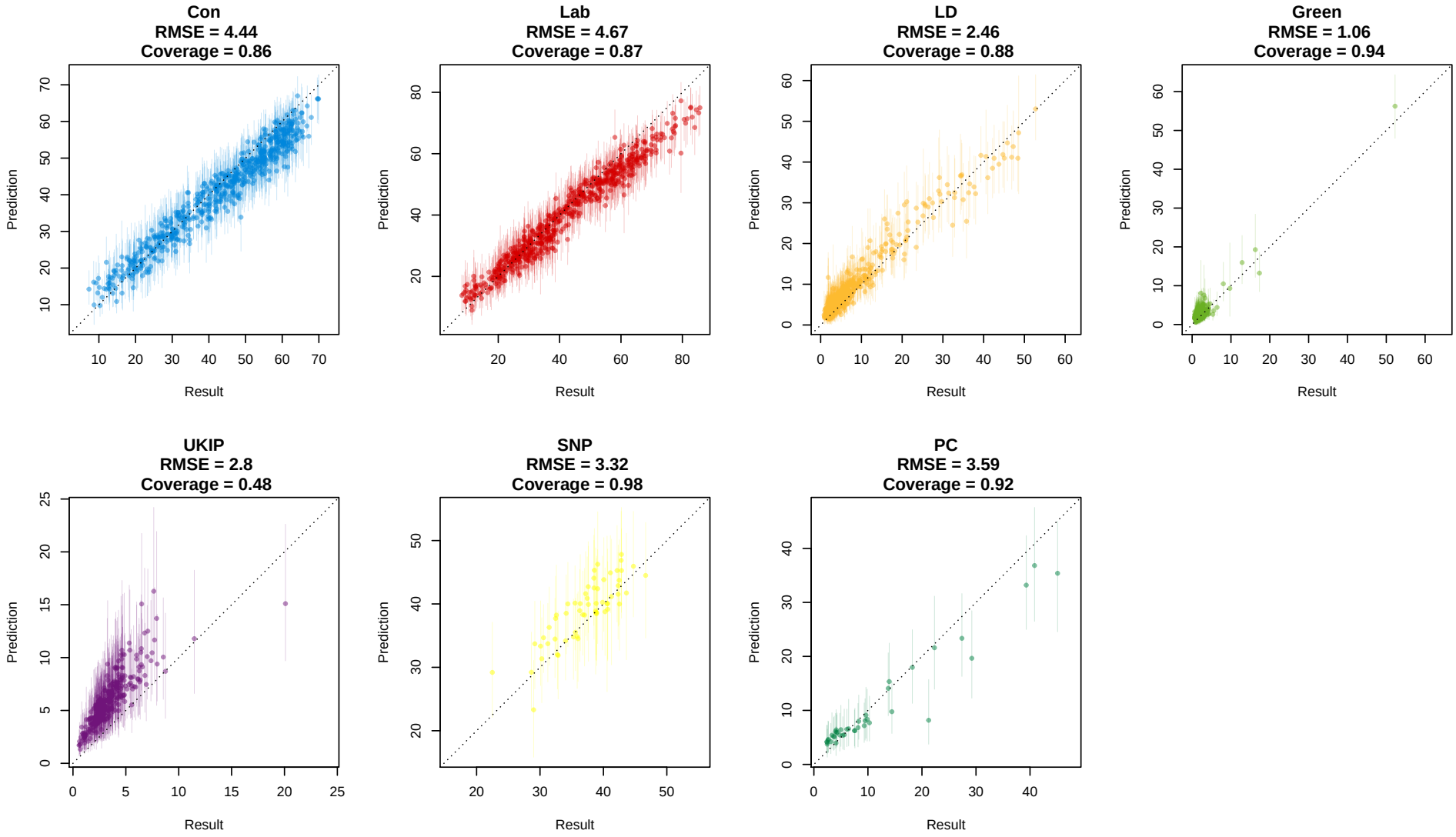

Figure A1: Predicted vs actual vote shares for UK parties by parliamentary constituency in the 2017 UK general election.

References

Bernstein, R., A. Chadha, and R. Montjoy (2001). Overreporting voting: Why it happens and why it matters. Public Opinion Quarterly 65, 22–44.

Campbell, J. E. (2012). Forecasting the 2012 american national elections: Editor's introduction. PS: Political Science and Politics 45(4), 610–613.

Campbell, J. E. (2016). Forecasting the 2016 american national elections. PS: Political Science and Politics 49(4), 649–654.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. Journal of Statistical Software 76(1).

Fisher, S. D. and M. S. Lewis-Beck (2016). Forecasting the 2015 british general election: The 1992 debacle all over again? Electoral Studies 41, 225–229.

Gelman, A. and T. C. Little (1997). Poststratification into many categories using hierarchical logistic regression. Survey Methodology 23(2), 127–135.

Hanretty, C. (2017). Electionforecast.co.uk.

Hanretty, C., B. Lauderdale, and N. Vivyan (2016a). Combining national and constituency polling for forecasting. Electoral Studies 41, 239–243.

Hanretty, C., B. E. Lauderdale, and N. Vivyan (2016b). Comparing strategies for estimating constituency opinion from national survey samples. Political Science Research and Methods, 1–21.

Holbrook, A. L. and J. A. Krosnick (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. Public Opinion Quarterly 74(1), 37–67.

Hummel, P. and D. Rothschild (2014). Fundamental models for forecasting elections at the state level. Electoral Studies 35, 123–139.

Jackman, S. (2005). Pooling the polls over an election campaign. Australian Journal of Political Science 40(4), 499–517.

Jackman, S. and B. Spahn (Forthcoming). "why does the american national election study overestimate voter turnout?". Political Analysis.

Jennings, W. and C. Wlezien (2018). Election polling errors across time and space. Nature Human Behaviour 2(4), 276–283.

Kennedy, C., M. Blumenthal, S. Clement, J. D. Clinton, C. Durand, C. Franklin, K. McGeeney, L. Miringoff, K. Olson, D. Rivers, et al. (2018). An evaluation of the 2016 election polls in the united states. Public Opinion Quarterly 82(1), 1–33.

Lauderdale, B. E. and D. Linzer (2015). Under-performing, over-performing, or just performing? the limitations of fundamentals-based presidential election forecasting. International Journal of Forecasting 31(3), 965–979.

Lax, J. R. and J. H. Phillips (2009). How should we estimate public opinion in the states? American Journal of Political Science 53(1), 107–121.

Leemann, L. and F. Wasserfallen (2017). Extending the use and prediction precision of subnational public opinion estimation. American Journal of Political Science 61(4), 1003–1022.

Linzer, D. A. (2013). Dynamic bayesian forecasting of presidential elections in the states. Journal of the American Statistical Association 108(501), 124–134.

Park, D. K., A. Gelman, and J. Bafumi (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. Political Analysis 12(4), 375–385.

Rivers, D. and D. Bailey (2009). Inference from matched samples in the 2008 u.s. national elections. American Association of Public Opinion Research – Joint Statistical Meetings.

Rivers, D. and A. Wells (2015). Polling error in the 2015 uk general election: An analysis of yougov's pre and post-election polls. Technical report, YouGov.

Selb, P. and S. Munzert (2011). Estimating constituency preferences from sparse survey data using auxiliary geographic information. Political Analysis 19(4), 455–470.

Silver, N. (2017). 538 2016 election forecast.

Sturgis, P., N. Baker, M. Callegaro, S. Fisher, J. Green, W. Jennings, J. Kuha, B. Lauderdale, and P. Smith (2016). Report of the inquiry into the 2015 british general election opinion polls. Technical report, British Polling Council.

Tausanovitch, C. and C. Warshaw (2013). Measuring constituent policy preferences in congress, state legislatures, and cities. The Journal of Politics 75(2), 330–342.

Wang, W., D. Rothschild, S. Goel, and A. Gelman (2014). Forecasting elections with non-representative polls. International Journal of Forecasting 31(3), 980–991.

Wright, F. A. and A. A. Wright (2018). How surprising was trump's victory? evaluations of the 2016 us presidential election and a new poll aggregation model. Electoral Studies 54, 81–89.