A Moving Target? An Analysis of Variation in Polling Error Across Post-war British General Elections Focusing on the Significance of Electoral Context

A Moving Target? An Analysis of Variation in Polling Error Across Post-war British General Elections Focusing on the Significance of Electoral Context

In this article, we argue that electoral context affects the projection mechanisms inherent in polling. This insight applies both to the estimation of party vote shares by pollsters and to poll-driven substantive political expectations. To test this argument, we analyse 794 in-campaign polls covering the UK's 21 post-war general elections, as well as an updated version of Jennings and Wlezien's (2018) international polling dataset. We demonstrate that the election level houses a substantial portion of the observed variance in polling error. This finding is valid across several modelling approaches and a range of measures of polling accuracy both within and beyond the UK. Within the UK, we show that the election level is a particularly important locus of variance when it comes to analysing whether polls give rise to misleading substantive expectations about election results.

Keywords: polling error; electoral context; multi-level modelling; British general elections; intra-class correlation

Introduction

In this article, we seek to bring the elections that polls are designed to predict more explicitly to the forefront of theories and analyses of polling error. The idea that electoral context affects polling accuracy is not a new one. Indeed, discussions of the features of individual elections appear in some of the earliest attempts to make sense of major polling failures (see, for example: Wilks et al., 1949). More recently, the importance of electoral context for analysing polling error has been foregrounded by comparative analyses that include variables pertaining to both political systems and specific aspects of individual elections (Jennings and Wlezien, 2018; Sohlberg and Branham, 2020).

However, relatively little systematic attention has been given to the research question that animates this article: to what extent is polling error is an election-level phenomenon? Addressing this question allows us to assess the substantive contention that some elections are simply more difficult for pollsters to predict accurately than others. To the extent that this is the case, it makes sense to think about polling error as having a hierarchical data structure, in which individual polls are clustered into elections which are more or less difficult to accurately predict. We explore this contention empirically, analysing the proportion of variance in polling error that is clustered at the election-level. We demonstrate that the election level should be formally accounted for to properly model polling error within and beyond the UK.

We also make a theoretical contribution by explicating mechanisms through which electoral context affects polling error. We argue that electoral context affects the reliability of the projection techniques associated with polling. Such techniques are required to produce vote share estimates based on survey data and are also (albeit often implicitly) central to the creation of substantive political expectations based on polls. We contend in this paper that the alignment between political context and widespread polling techniques varies across elections, which drives election-level variation in polling error.

Formal polling analyses typically investigate the extent to which pollsters' estimates of vote share estimates deviate from actual vote shares. This makes sense, given that polls are generally reported as party vote percentages. However, notable polling failures also encapsulate what Sturgis et al. (2016: 7) describe as "poll-induced expectation(s)" about the post-election distribution of political power across parties. In the British case, such expectations centre on identifying which party will win the most seats and whether it will be able to form a single-party majority government. These expectations result from (often implicit) poll-to-seat projections. Polling error in this sense arises when poll-based political projections prove misleading.

We analyse both types of polling error in this article – demonstrating that polls' vote share estimate errors cluster at the election level both in British general elections and internationally. We find that election-level clustering in the UK is particularly pronounced for the likelihood of a poll producing misleading projections of substantive political outcomes. More generally, we make it clear that an appreciation of the role and importance of electoral context should be integral to the theory, practice, and evaluation of polling.

Electoral Context in the Study of Polling Error

Two issues lie at the heart of the literature analysing polling error. Firstly, there is a disjuncture between the theory of probability sampling and the reality of conducting a poll. Buchanan (1986: 224), summing up this problem, states that "merely describing [the] ideal sampling arrangement is enough to show the contrast to what pollsters are really doing". Under this rubric fall the most consistent and pronounced sources of British polling error – a combination of unrepresentative samples and problematic weighting procedures (see, for instance: Mellon and Prosser, 2017; Prosser and Mellon, 2018; Sturgis et al., 2016). Another source of error that arises in this domain is geographic heterogeneity. Spatial clustering of party support and uneven geographic distribution of trends both complicate sampling and make seat share prediction less certain (Curtice, 2015).

Secondly, pollsters must account for the processes underlying turnout and vote choice. In doing so, they are faced with issues arising from the variable reliability of respondents' answers. Social desirability bias can lead respondents to over-report their turnout likelihood (see, for instance: Karp and Brockington, 2005) or to be less willing to report certain vote intentions, an intuition typically captured in the phrase 'shy' voters (Pickup et al., 2011). Relatedly, many respondents are unsure of their vote intention when polled, and pollsters' techniques for estimating how such voters will eventually break have been linked to polling error in the UK (Fisher and Shorrocks, 2018).

These issues are well-rehearsed in a voluminous literature, where the central question is how pollsters can adapt their methods to deal with them (for detailed reviews see: Crespi, 1988; Moon, 1999; Tudor 2017). In the UK, large swathes of this type of research responded to dramatic polling misses in the 1970 (Abrams, 1970; Koff, 1972) and 1992 (detailed in: Fisher and Lewis-Beck, 2016) elections, as well as the perceived polling failures of the 2010s (see, for example: Curtice, 2016; Mellon and Prosser, 2017; Prosser and Mellon, 2018; Sturgis et al., 2016).

The significance of electoral context is discernible in numerous post-mortem analyses of polling failures. Such analyses are often focused on attributing blame (see Crewe, 1992 for a rather playful example of this emphasis). Durand and Blais (2019) provide a conceptual distinction between a polling 'miss' (which occurs due to patterns of voting behaviour) and a polling 'failure' caused by problematic methodology. One of the key causes of a polling miss in this framework is a last-minute shift in the opinions of voters. In the British context, such 'late swing' has been invoked as absolving the pollsters of blame (Abrams, 1970; Crewe, 1992; Moon, 1999).

However, other election-specific factors are less easy to disentangle in terms of the blame game. For instance, when the American Association for Public Opinion Research (AAPOR) convened an ad hoc committee to explain the failure of pollsters to foresee Donald Trump's election in 2016, they found that a key problem was that poll weightings did not adequately account for the pronounced political relevance of educational attainment in that election (Kennedy et al., 2016). Along similar lines, Mellon and Prosser (2017) attribute the 2015 UK polling failure primarily to the over-weighting of politically disengaged voters due to their under-representation in polling samples. In the political context of 2015, this sampling flaw caused an over-estimation of Labour support.

Although election-level factors are widely discussed in commentary on polling failures in individual elections, comparative analyses that investigate the contention that elections may vary systematically in their predictability are rare. Jennings and Wlezien (2018) conclude that political context can predict some variation in poll performance. They find that the type of electoral system (proportional representation versus majoritarian) and type of election (presidential versus legislative) are both predictors of polling error. Sohlberg and Branham (2020) build on this work, separating relevant election characteristics into institutional versus non-institutional categories. They find that the non-institutional factors of electoral change and vote-buying prevalence affect poll performance, whereas turnout levels are not significant predictors of polling error. While these analyses represent welcome and important steps towards understanding the relationship between electoral context and polling error, this approach is still nascent.

Although much of the analytical focus in the polling error literature rests on the accuracy of pollsters' vote share estimates, elections with high levels of this type of polling error can slip under the radar. For instance, Prosser and Mellon (2018: 766) note that the 1951, (October) 1974, 1983, 1992, 1997, and 2001 British general elections all featured high levels of polling error but did not generate the same levels of attention as other polling failures because vote share estimates were not biased towards the "wrong winner". Indeed, the conceptual distinction between polls that are substantively misleading and the error associated with pollsters' estimated vote shares is well-understood (Mitofsky, 1998). However, to our knowledge no comparative empirical analyses of polling error focus on the substantive political implications of polls.

In this article, we provide a more generalised theoretical account of the relationship between electoral context and polling error than has emerged heretofore. In doing so, we seek to incorporate the interplay of polling projection methods and electoral context into an overarching mechanism that leads to election-level clustering of polling error. We also take an original approach to analysis – measuring the proportion of variance in polling error that lies at the election level both within and beyond the UK. These contributions are necessary steps forward in understanding how electoral context can affect polling error. By exploring the UK case in greater depth, we are well-placed to further develop the early insights of Sohlberg and Branham (2020) concerning the non-institutional drivers of polling error. Our UK focus also enables us to address a key contradiction of the polling error literature outlined above – namely that it tends to be driven by elections where polls are substantively misleading in terms of seat shares whereas its analytical focus is largely on the accuracy of pollsters' vote share estimates.

How and Why Electoral Context Matters for Polling Error

In this section, we elaborate a generalised theoretical mechanism that connects electoral context to variation in polling error. We also set out two hypotheses that follow from our line of argumentation, and which our analysis will test. We begin by unpacking the two concepts around which our argument centres – electoral context and polling projection mechanisms. We then outline two routes through which electoral context might affect the reliability of polling projection mechanisms. A direct route concerns those aspects of electoral context affect the ability of those intending to vote to cast their ballot. However, we contend that the more important route lies at the intersection of prevalent polling projection methods and variable aspects of electoral context. We then move to a discussion of how this relationship might play out when it comes to the reliability of the substantive political expectations that polls generate. We argue that election-level clustering of polling error is likely to be more pronounced when it comes to the accuracy of seat-based, politically salient aspects of election prediction than for vote share estimates.

The relationship between electoral context and polling error

We use the term ‘electoral context’ to refer to the specific political configurations and dynamics that a given election encapsulates. While there may be overlap across elections in terms of the main parties competing, the personalities of key candidates, campaign strategies, patterns of electoral behaviour among voters, the media environment, and the unfolding of events prior to and during the campaign timeline – each of these elements is variable, and the constellation of such factors is never identical from one election to the next.

We contend electoral context can affect the mechanisms that the facilitate the projection of election outcomes based on survey data. The conceptualisation of polls as vote projections rests on the presentation of polling results as national-level vote share estimates. Such estimates can only be generated from a pre-election survey if three projection mechanisms have been applied. This implies that polling figures are future-oriented, rather than serving solely as snapshots of public opinion. We think that we are on relatively solid ground here; the very endeavour of evaluating polling accuracy based on election results would be nonsensical if polls are conceptualised as having no bearing on the future.

There are three major categories of projection mechanism that facilitate the production of vote share estimates by pollsters. The first is sample-to-population projection. While probability sampling is the theoretical basis of this inferential process, polling companies are usually forced to deviate from this ideal due to the difficulty of securing a fully representative survey frame and the vagaries of non-response. This means that both the sampling strategies of polling companies and the weighting systems used to adjust survey figures are important for the accuracy of the overall projection. The second element of polling projection involves estimating respondents' turnout likelihood (Hillygus, 2011; Traugott and Tucker, 1984), an issue toward which no small amount of attention has been dedicated by the polling industry and scholarship over the years (for a review see: Kenett, Pfeffermann, and Steinberg, 2018). The third projection mechanism connects respondents' reports of their future voting choice to estimates of their actual voting choice, a problem that is particularly pronounced when respondents indicate uncertainty (Fisher and Shorrocks, 2018).

How might electoral context affect the reliability of pollsters' projection mechanisms? In the first place, there are electoral-contextual factors that can diminish the reliability of voters' assessment of their voting behaviour, and thereby directly reduce the reliability of polling projection mechanisms. Particularly important here are electoral contexts where access to voting (or having one's votes counted) is restricted or politicised (see, for a discussion of this issue in the American context: Shah and Smith, 2021). When this is the case, voters who intend to vote might find themselves unable to do so, throwing off survey-based vote projections.

However, we contend that most drivers of polling error lie at the *intersection* of electoral context and polling projection methods. Election-level variation across an array of dimensions that affect the three above-discussed polling projection mechanisms has been widely observed. In terms of sample-to-population projection, the political salience of socio-demographic and other variables used to select and adjust samples waxes and wanes across elections (Tilley and Evans, 2017). Election-level variation in both levels of turnout and the underlying socio-demographic predictors of turnout is common (Franklin, 2004). Such variation can affect the accuracy of turnout projection. Factors that influence the reliability of voters' reports of their future vote choice include the social desirability of indicating support for a given party or candidate (Brown-Iannuzzi et al., 2019); strength of partisanship (Tilley, 2002); in-campaign voter volatility (McAllister, 2002); campaign strategies (Vavreck, 2009); and the occurrence of in-campaign 'shocks' (Wlezien and Erikson, 2001; Fieldhouse et al., 2020). All of these factors are variable aspects of electoral context.

The specific methods used by pollsters to project vote shares from survey data are also variable across elections. They evolve over time partly due to changing communication technologies and emerging statistical analysis affordances (see Prosser and Mellon, 2018 for a discussion of this process in the UK). The dynamics of recent elections also weigh into pollsters' methods. For instance, during the 2017 general election pollsters applied a more 'aggressive' (Sturgis and Jennings, 2017: np) set of turnout adjustment procedures in the light of the 2015 polling failure.

As such, each election is characterised by a combination of electoral context affecting the three categories polling projection mechanisms and a prevalent set of polling projection methods. We contend that large, systematic errors in polling are most likely when electoral context bears on widely-adopted polling projection methods – activating them to an unforeseen extent and in a consistent direction. A perfect example of such a dynamic can be found in Mellon and Prosser's (2017) account of the 2015 UK polling miss – a pre-existing widespread polling projection method (the over-weighting of under-sampled disengaged voters) aligned unexpectedly with political context to produce an under-estimate of Conservative Party vote share.

Overall, we believe that there is reason to suspect that electoral context can affect polling error, either directly or in combination with widespread polling practices. If this is the case, then we should observe that:

H1: A substantial portion of polling error is clustered at the election-level.

Electoral Context and the Reliability of Poll-Based Political Projections

Extant analyses of polling error largely focus on the disparity between a poll's predicted vote shares and the election results. However, as we discussed above, many UK elections characterised by high levels of this sort of polling error have not been classed as polling failures. This discrepancy can be explained through an alternative conceptualisation of polling error – one that assesses whether polls prove politically misleading. While these conceptualisations of polling error are obviously related, they do not fully overlap (Mitofsky, 1998). For example, polls conducted for the second round of the 2017 French presidential election correctly called Emmanuel Macron as the victor but overestimated his margin of victory by an average of ten percentage points (Enten, 2017). On the other side of the coin, polling in the 2016 US Presidential election produced vote share projections that proved relatively accurate at the national level, but largely failed to forecast Donald Trump as the eventual winner (Prosser and Mellon, 2018).

In this sub-section, we elaborate why the effects of electoral context on polling error should be most pronounced when it comes to the reliability of the substantive political projections that polls can be used to generate.

The mechanism at play in the creation of substantive political projections from polls in the UK leads from vote share estimates to expected party seat distributions. The strategies underlying such projections are multiple and of varying degrees of sophistication – they range from the intuitive prognostications of pundits to the statistical modelling approaches of experts (see for example: Hanretty, Lauderdale, and Vivyan, 2014). YouGov's use of multilevel regression and post-stratification (MRP) to produce seat share estimates (see, for instance: YouGov, 2019) is a notable example of a pollster providing a direct seat share distribution projection. Once a seat distribution projection is produced, the political calculus for creating a substantive political projection in the UK political system straightforward. It involves identifying the largest party in terms of seat share and assessing whether it will enjoy a parliamentary majority.

While election-level factors that exacerbate vote share projection error also contribute to the creation of misleading substantive expectations, some aspects of electoral context contribute to uniquely to the latter type of polling error. Most importantly, elections vary considerably in their marginality. In post-war UK general elections, the percentage difference in vote share between the largest and second largest party has ranged between 14.8% (1983) and 0.6% (February 1974). Close elections will, by their very nature, be more difficult to call correctly than landslides. Furthermore, the operation of the UK's first-past-the-post system creates a thorny 'votes to seats' problem for poll-based forecasting (Fisher et al., 2011) and this problem becomes more pronounced as vote fragmentation and geographic heterogeneity of changes in party support increase. These aspects of British electoral context act in addition to the mechanisms that give rise to vote share projection error discussed in the previous sub-section, leading us to hypothesise that:

H2: Error in poll-based political projections in UK general elections clusters at the election-level to a greater extent than error in estimated vote shares.

Data, Methods, and Analysis

Dataset

Our primary dataset captures 794 in-campaign polls across the 21 post-war general elections in the United Kingdom. The campaign is defined as the period from the formal dissolution of Parliament to the day of polling, as per Sanders (2003). Our UK polling data is supplemented by a global polling dataset capturing 21,432 polls across 400 national elections in 50 countries for the purposes of comparison (Jennings and Wlezien, 2018). We supplement the Jennings and Wlezien dataset with polling data from more recent elections, carrying over their party numbering conventions. We removed all interpolated values so that we only analyse data capturing published polling figures. Full details of the procedures and sources used to compile these datasets are provided in Section A1 of the Online Appendix to this article.

Operationalisation of Polling Error

Our approach to measuring polling error captures the conceptual distinction between vote share error and poll-based political projection error. To analyse vote share error, we compare polling figures with the relevant election outcome. For each set of polling numbers, we calculate the Mean Absolute Error (MAE) score, the most widely used metric in the field of polling error analysis (see: Jennings, Lewis-Beck, and Wlezien, 2020 for a recent discussion of some of the strengths and limitations of this metric in applied analysis). *MAE* has the advantage of capturing error across multiple party categories (in our analysis, these are Labour, the Conservatives, the Liberal Party/Liberal Democrats, and ‘Other’). To improve the robustness of our analysis, we complement *MAE* with two alternative measures. These are: Mosteller et al.’s (1949) Measure 5 (M5), which captures the absolute difference in the margin between the two leading parties in a poll and the actual margin between them on election day and Martin, Traugott, and Kennedy’s (2006) Measure A, which captures the degree to which polls over- or under-estimate the vote shares of the two leading parties. The formulae used in calculating all three of these values are provided in the Section A2 of the Online Appendix to this article.

Given the majoritarian structure of the UK political system (Lijphart, 1997), there are two substantively important political outcomes to be projected from polling data, both of which centre on seat distributions. The first of these involves the identity of the largest party in terms of seat share. The second is whether a single party enjoys a parliamentary majority. We deem a poll-based political projection to be ‘correct’ if it produces seat projections that correctly predict both outcomes. Poll-based seat distribution projections that fail to correctly predict either one or both outcomes are coded as ‘incorrect’.

This approach requires that a seat distribution projection must be created for each poll, and the substantive expectations resulting from the projected distribution must be assessed against the relevant election outcome. Seat distribution projection is not straightforward given the UK’s first-past-the-post electoral system. Our preferred approach involves mapping the changes to party vote share across constituency-level results from the previous election. We use ratio swing (Awan-Scully, 2014), more commonly referred to as proportional swing (Butler and Kavanagh, 1979), in which the ratio of the change in support for each party between elections is taken and applied across all constituencies individually, altering their predicted vote shares proportionally. We follow conventional practice in excluding Northern Irish consistencies from this calculation.

We deploy two alternative approaches to seat projection, to increase our confidence that our findings are not driven by a specific seat share projection approach and to reflect the fact that that techniques for converting polling figures into seat distributions are not constant across the elections that we consider. Both alternative approaches are more straightforward than our preferred method. The first of these is mapping vote changes across constituency-level results employing the additive uniform national swing formula derived by Baxter (2007). The second uses a power-exponent law (Shugart and Taagepera, 1989), where the ratio of seats gained by the two leading parties in an election is taken to be equal to the ratio of their vote shares raised to an exponent. Although the best-known example of such a law is the 'cube rule' where the exponent is 3 (Laakso, 1979), an exponent of 2.6 has been found to be the most effective (Shugart, 2008) for UK elections. We therefore use 2.6 as the exponent in our calculations. Again, full procedures and formulae for the calculation of these seat share projections are provided in the Section A2 of the Online Appendix to this article.

Throughout the analysis, we focus on the findings arising from our preferred measure of vote share error (*MAE*) and our preferred approach for generating seat distributions (the ratio swing method). While we refer to the Appendices for the reproduction of these results using alternative operationalisations, we note here that none of these alternative approaches generate findings that differ substantively from those that we report in the paper.

Descriptive Analysis

Figure 1 displays a series of boxplots characterising *MAE* over all 794 polls in our UK dataset, grouped by election. The width of the boxplots is proportionate to the square root of the number of polls captured for each election. We present two orderings of these data. In the top pane, the elections are ordered temporally, whilst in the bottom pane they are ordered by the size of the median *MAE*. This presentational choice allows us to tell two substantive stories about the distribution of *MAE* across the 21 elections studied.

Looking at the over-time ordering, we can see a relatively random pattern of levels of polling error, with a steady increase in the volume of polls as we approach the 2019 election. However, when we order by median *MAE*, we can see that there are significant discrepancies across elections. These initial findings concur with our contention that vote share error is clustered at the election-level and is also congruent with recent studies that find little evidence for an over-time decline of polling accuracy (Jennings and Wlezien, 2018; Prosser and Mellon, 2018). We note that these trends are replicated across both *M5* and *Measure A* and therefore appear to be a feature of the underlying data, rather than an artefact of a particular polling error measurement strategy (see: Online Appendix, Section A4 for details).

FIGURE 1 ABOUT HERE

When we run a similar descriptive analysis on our measure of whether polls give rise to correct or incorrect substantive political projections in Figure 2, we see a more pronounced version of the same story. Indeed, many elections sit at the logical extremes on this measure in the sense that where either all polls were ‘correct’ (1950; 1951; 1966; October 1974; 1983; 1987; 1997; 2001; 2005) or ‘incorrect’ (1970; 2015). We note that our characterisation of these data is robust to alternative seat projection mechanisms (see: Online Appendix, Section A4). There are, however, some interesting contrasts that demonstrate the value of using both approaches to polling error in tandem. For instance, the polls for the 2015 election all generated ‘incorrect’ political projections (projecting a hung parliament rather than the eventual Conservative majority result) but did not generate very high *MAE* scores.

FIGURE 2 ABOUT HERE

TABLE 1 ABOUT HERE

In combination, Figures 1 and 2 provide substantive plausibility for hypothesis 2, as the election-level clustering of error is more extreme for poll-based political projections than for estimated vote shares. The numerical values that underlie Figures 1 and 2 are provided in Table 1.

Inferential Analysis

We run a series of null multi-level models over our data to estimate the portion of total polling error variation that can be accounted for by the higher (election-level) tier of our data (Hox, 2002). Using a multi-level approach presents us with a range of options and decisions regarding both data treatment and model estimation. We treat our data structure as hierarchically nested (Osborne, 2000), with polls nested within the individual elections to which they relate. As such, our data is two-level in nature with no instances of cross classification.

Our preferred modelling approach for estimated vote share error analysis is Bayesian, a decision that is largely driven by the relatively small group *n* (that is, number of elections) in our dataset. Frequentist approaches, such as maximum likelihood and restricted maximum likelihood models, can bias parameter estimates downward for data where group *n* is relatively small (El-Horbaty and Hanafy, 2018) and tend to produce large standard errors (Goldstein, 2011). We therefore use Bayesian Markov chain Monte Carlo (MCMC) estimation, which is effective for the analysis of multi-level models (Carlin and Louis, 1996). Our preferred model uses half-Cauchy priors for the variance parameters, as recommended for MCMC multi-level modelling (Gelman, 2006). It is important to note that our results do not stand or fall on these modelling decisions; Sections A3 and A5 of the Online Appendix demonstrate their robustness across alternative modelling strategies.

The key test statistic that we derive from our models is the intra-class correlation coefficient (ICC), which measures the percentage of total variation in the data that is accounted for by between-group (in our case, between-election) variation (Gelman and Hill, 2006). The ICC is calculated as shown in equation 1, where $\sigma^2_{between}$ represents between-group variance and σ^2_{within} represents within-group variance (Bobak, Barr, and O'Malley, 2018). We employ a range of additional ANOVA-based measurement techniques in Section A5 of the Online Appendix to ensure the robustness of our findings, including ICC1 (Bliese, 2000), eta-squared (Shieh, 2012), and omega-squared (Albers and Lakens, 2018).

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}} \tag{1}$$

We provide 95% confidence and credible intervals for our range of ICC estimates. For our MCMC model outputs, these are provided by the BRMS R package (see: Bürkner, 2017 for details).

As regards interpretation, an ICC value of 0 indicates that the values taken by observations in a dataset are not affected by their membership within the higher tier grouping. Conversely, a value of 1 indicates that the observed data only varies between these groups (Sommet and Morselli, 2017). The ICC value above which multi-level modelling strategies are considered advisable is 0.05 (Raykov, 2011). We therefore take this value as a test statistic for the evaluation of hypothesis 1. As such, the cut-off point for testing this hypothesis is whether the 95% confidence or credible interval for the ICC value includes values that are below 0.05.

TABLE 2 ABOUT HERE

Table 2 presents the ICC values and 95% credible intervals given by our MCMC models across our three measures of estimated vote share error. Results for modelling our UK data are presented in the first row. Subsequent rows present results when we apply our models to the updated Jennings and Wlezien (2018) dataset as well as to individual countries for which we had at least 500 polls across 10 elections.

In the case of the UK, between-election variation is estimated to account for 29.2% to 32.7% of total estimated vote share error. The lower end of this range comfortably exceeds the 5% cut-off point for the use of multi-level analysis. Looking beyond the UK, the average ICC value across the entirety of our dataset ranges from 50% to 68.1%, while the ICC values of individual countries range from 19.5% to 88.7%. Once again, none of the lower bounds of the 95% credible intervals are below 5%. Substantively, these findings are corroborated across alternative measures of vote share error using a wide range of alternative modelling strategies and ICC calculations which we present and discuss in the Online Appendix (Sections A3 and A5). As such, we conclude that polling error is substantially clustered at the election level across all of our cases. These findings mean that we are unable to reject hypothesis 1.

There is also indicative evidence that the proportion of variance clustered at the election level varies between countries, although this insight should be treated cautiously, given the different sample sizes at play. Overall, we can see that the influence of election-level differences on continuous polling error in the United Kingdom is slightly below the international average, which may reflect the relative maturity of the British polling industry and the political stability of post-war elections in the UK.

The last part of our analysis focuses on the accuracy of poll-based political projections in the UK. Because the variables that we use to operationalise this outcome are dichotomous in nature, an alternative modelling approach is needed to estimate ICC values, along with 95% confidence and credible intervals. The two most prominent methods used in the multi-level modelling of dichotomous variables are Laplace approximation and adaptive Gauss-Hermite quadrature (AGHQ), with the latter being most applicable to two-level models such as our own (McNeish, 2016).

Bayesian MCMC estimation is applicable to models concerned with dichotomous variables. However, for such models, Bayesian MCMC numerical integration is inferior to AGHQ, as Monte Carlo integration estimates vary from one approximation to the next, which means that hybridisation with other quadrature procedures are required (Owen, 2013). Consequently, AGHQ is the preferred estimative method for the generalised linear null models in our analysis.

Equation 1 is not applicable to binomial generalised linear models with a logit link, so a different ICC calculation is required. Given that our error measure is a discretisation of an underlying threshold continuous variable, we calculate ICC in relation to it using the latent variable approach outlined in equation 2, where $\sigma^2_{between}$ again represents between-group variance, but within-group variance is held constant at $\pi^2/3$ (Nakagawa, Johnson, and Schielzeth, 2017). The latent variable approach has the added benefit of being both calculatively comparable to ICC measures for linear models (Leckie et al., 2020), as well as allowing for the practical comparison of results between linear and logit models (see, for example: Fernée and Konstantinos, 2021).

As before, we provide 95% confidence and credible intervals alongside our ICC estimates. These are calculated using the logit transformation via the *estat icc* function in Stata (Nakagawa and Schielzeth, 2010), as well as via model outputs in the case of our Bayesian estimates.

$$ICC = \frac{\sigma^2_{between}}{\sigma^2_{between} + (\pi^2/3)} \qquad (2)$$

Table 3 displays the ICC values and 95% confidence and credible intervals that result from following these procedures using our three approaches to creating poll-based seat projections. According to our preferred measure and modelling approach (indicated in bold in the first row), the estimated ICC value is 87% and the 95% confidence interval for this value ranges from 74% to 95%. Looking across the entire suite of models and measures, the lowest ICC estimate within the 95% confidence and credible intervals is 68%. As such, the proportion of error variation explained by the election level is substantially higher when assessing poll-based political projections within the UK (ranging from 68% to 95%) than vote share error (ranging from 19.7% to 47.8%). These findings mean that we cannot reject hypothesis 2.

TABLE 3 ABOUT HERE

Conclusion

Crewe (1992: 478) explains that "opinion polls are not (…) the mere fluff of elections. They influence the timing, strategy and course of election campaigns and, to that extent, the result." As election nights unfold in the UK, both exit polls and the subsequent tallying of votes are often framed against poll-based expectations, allowing political actors, commentators, and members of the public to understand the extent to which they have been misled by in-campaign polling (for a detailed account, see: Wilks-Heeg and Andersen, 2020). As such, analyses of polling error are fundamentally evaluative enterprises. Reputations are at stake when companies engage in in-campaign polling, and unusually large polling errors, especially when they create misleading political expectations, can lead to a blame-game once the votes have been counted.

This study's major contribution to the literature on poll evaluation lies in making the theoretical and empirical case that polling error is a dynamic process of interplay between polling projection methods and electoral context. In empirical terms, we demonstrate that a substantial portion of the observed variance in polls' vote share error is clustered at the election level. We show that this is true both within and beyond the UK. Within the UK, we also find the election level houses a higher portion of the variance in the likelihood of a poll generating misleading political projections concerning the election result. Our theoretical approach provides a way of approaching why this is the case and can, we believe, lead to a more considered and balanced understanding of polling hits and misses. In terms of developing this work further, the major implication of our findings is that scholars should incorporate analyses of election-level variables into their evaluations of polling accuracy.

It would be remiss to come this far without identifying likely suspects that should be the focus of future investigations along these lines. There are a range of relatively fixed factors that the literature point towards – it goes without saying that no analysis should ignore the execution of the fundamentals of polling (Weisberg, 2009), so variation in polling practices over time should be considered. In addition, marginal elections are always going to be more difficult for pollsters to call. Analysts of polling error should also bear in mind that polling accuracy improves in well-understood ways as the campaign approaches its conclusion (Erikson and Wlezien, 2012) and that those campaigns characterised by high levels of uncertainty and volatility (especially when there is a late movement of voters in a systematic direction) will generate higher levels of polling error.

While all of these factors should therefore feature in polling error analysis, we contend that political instability in general and change relative to the previous election in particular are the most likely aspects of electoral context to drive polling error. These contextual dynamics create the optimal conditions for a mismatch between electoral context and widespread polling projection methods to occur. This intuition receives initial confirmation in Sohlberg and Branham's (2020: 11) conclusion that "elections conducted around large changes in party support are particularly difficult to poll". However, their analysis centres on electoral volatility in terms of parties' overall vote share. We would argue that such shifts in party support levels are a proxy for the key mechanism: changes in the *drivers* of electoral behaviour. When electoral context alters the foundations upon which polling projection rests, pollsters may find themselves fighting the last war (that is, using techniques that reflect the electoral context of the previous election) which sets them up for a polling failure.

Acknowledgements: The authors would like to thank participants at the 5th Leuven-Montreal Winter School on Elections; The 2019 Political Studies Association Annual Conference; and the 2019 Annual Elections, Public Opinion and Parties Conference in the University of Strathclyde for their helpful comments and suggestions on earlier iterations of this paper. We would also like to thank colleagues at Swansea University's Research Seminar for their feedback on an early version of this work. The authors are also indebted to the anonymous reviewers, who provided extensive and highly constructive feedback on the paper.

References

Abrams, Mark. 1970. 'The opinion polls and the 1970 British general election.' *Public Opinion Quarterly* 34: 317-324.

Albers, Casper, and Daniel Lakens. 2018. 'When Power Analyses Based on Pilot Data are Biased: Inaccurate Effect Size Estimators and Follow-up Bias'. *Journal of Experimental Social Psychology*, 74: 187-195.

Awan-Scully, Roger. 2014. 'A Bit More on Uniform National Swing – and an Alternative'. Available at: https://blogs.cardiff.ac.uk/electionsinwales/2014/11/17/807/, accessed February 6, 2021.

Baxter, Martin. 2017. 'Additive Model.' Available at: https://www.electoralcalculus.co.uk/newmodel.html#additive, accessed January 25, 2019.

Bliese, Paul D. 2000. 'Within-Group Agreement, Non-Independence, and Reliability', in *Multilevel Theory, Research, and Methods in Organizations*, edited by

Katherine J. Klein and Steve W. J. Kozlowski, 349-381. San Francisco: Jossey-Bass.

Bobak, Carly A., Paul J. Barr, and A. James O'Malley. 2018. 'Estimation of an inter-rater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement', *BMC Medical Research Methodology* 18:93.

Brown-Iannuzzi, Jazmin, Maxine Najle, and Will Gervais. 2019. 'The illusion of political tolerance: Social desirability and self-reported voting preferences.' *Social Psychological and Personality Science* 10: 364-373.

Buchanan, William. 1986. 'Election Predictions: An Empirical Assessment'. *The Public Opinion Quarterly* 50.2: 222-227.

Butler, David and Dennis Kavanagh. 1980. *The British General Election of 1979*. London: Macmillan Press.

Bürkner, Paul-Christian. 2017. 'brms: An R Package for Bayesian Multilevel Models Using Stan.' *Journal of Statistical Software* 80: 1-28.

Carlin, Bradley and Thomas Louis. 1996. 'Identifying prior distributions that produce specific decisions, with application to monitoring clinical trials.' In Donald John, Kathryn Chaloner, and Geweke Arnold Zellner (eds) *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*. New York: John Wiley & Sons, pp. 493-503.

Crespi, Irving. 1988. *Pre-election polling: Sources of accuracy and error*. New York: Russell Sage Foundation.

Crewe, Ivor. 1992. 'A Nation of Liars? Opinion Polls and the 1992 Election'. *Parliamentary Affairs* 45.4: 475-495.

Curtice, John. 2015. 'A Return to Normality? How the Electoral System Operated', *Parliamentary Affairs* 68(suppl. 1): 25–40.

Curtice, John. 2016. 'EU Referendum – How the Polls Got it Wrong Again'. *The Conversation*, June 26. Available at: http://theconversation.com/eu-referendum-how-the-pollsgot-it-wrong-again-61639, accessed June 25, 2020.

Durand, Claire and André Blais. 2019. 'Quebec 2018: A Failure of the Polls?'. *Canadian Journal of Political Science* 53.1: 133-150.

El-Horbaty, Yahia, and Eman Hanafy. 2018. 'Some Estimation Methods and Their Assessment in Multilevel Models: A Review', *Biostatistics and Biometrics Open Access Journal* 5: 1-8.

Enten, Harry. 2017. 'Macron Won, But the French Polls Were Way Off.' *FiveThirtyEight*. Available at: https://fivethirtyeight.com/features/macron-won-but-the-french-polls-were-way-off/, accessed June 24, 2020.

Erikson, Robert S., and Christopher Wlezien. 2012. *The timeline of presidential elections: How campaigns do (and do not) matter*. Chicago: University of Chicago Press.

Fernée, Christianne L. and Konstantinos P. Trimmis. 2021. 'Detecting Variability: A Study on the Application of Bayesian Multilevel Modelling to Archaeological Data. Evidence from the Neolithic Adriatic and the Bronze Age Aegean'. *Journal of Archaeological Science* 128: 105346.

Fieldhouse, Edward, Jane Green, Geoffrey Evans, Jonathan Mellon, Christopher Prosser, Hermann Schmitt, and Cees van der Eijk. 2020. *Electoral Shocks: The volatile voter in a turbulent world*. Oxford: Oxford University Press.

Fisher, Stephen and Michael Lewis-Beck. 2016. 'Forecasting the 2015 British general election: The 1992 debacle all over again?' *Electoral Studies* 41: 225-229.

Fisher, Stephen, Robert Ford, Will Jennings, Mark Pickup, and Christopher Wlezien. 2011. 'From polls to votes to seats: Forecasting the 2010 British general election.' *Electoral Studies* 30: 250-257.

Fisher, Stephen and Rosalind Shorrocks. 2018. 'Collective Failure? Lessons from Combining Forecasts for the UK's Referendum on EU Membership'. *Journal of Elections, Public Opinion and Parties* 28.1: 59 – 77.

Franklin, Mark. 2004. *Voter turnout and the dynamics of electoral competition in established democracies since 1945*. Cambridge: Cambridge University Press.

Gelman, Andrew. 2006. 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper).' *Bayesian analysis* 1: 515-534.

Gelman, Andrew, and Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Goldstein, Harvey. 2011. *Multilevel Statistical Models: Fourth Edition*. New Jersey: John Wiley & Sons

Hanretty, Christopher, Lauderdale, Benjamin and Nick Vivyan. 2014. 'Presenting a new model for forecasting the 2015 General Election' *London School of Economics British Polling and Policy Blog*. Available at:

https://blogs.lse.ac.uk/politicsandpolicy/presenting-a-new-model-for-forecasting-the-2015-general-election/, accessed March 3, 2021.

Hillygus, D. Sunshine. 2011. 'The Evolution of Election Polling in the United States.' *Public Opinion Quarterly* 75: 962-981.

Hox, Joop, J. 2002. *Multilevel Analysis: Techniques and Applications – Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Jennings, Will and Christopher Wlezien. 2018. 'Election polling errors across time and space.' *Nature Human Behaviour* 2.4: 276-283.

Jennings, Will, Michael Lewis-Beck, Christopher Wlezien. 2020. 'Election forecasting: Too far out?.' *International Journal of Forecasting* 36.3: 949-962.

Karp, Jeffrey A. and David Brockington. 2005. 'Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries'. *Journal of Politics* 67.3: 825–40.

Kenett Ron S., Danny Pfeffermann, David M. Steingberg. 2018. 'Election Polls – A Survey, A Critique, and Proposals'. *Annual Review of Statistics and its Application*. 5: 1 – 24.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney et al. 2018. 'An evaluation of the 2016 election polls in the United States.' *Public Opinion Quarterly* 82.1: 1-33.

Koff, Stephen P. (972. Public Opinion Polls and the 1970 British Election. *Il Politico* 37: 466-482.

Laakso, Markku. 1979. 'Should a Two-and-a-half Law Replace the Cube Law in British Elections?'. *British Journal of Political Science*, 9: 355-384.

Leckie, George et al. 2020. 'Variance Partitioning in Multilevel Models for Count Data'. Available at: https://arxiv.org/abs/1911.06888, accessed July 1, 2021.

Lijphart, Arend. 2012. *Patterns of Democracy: Government Forms and Performance in Thirty-six Countries - Second Edition*. London: Yale University Press.

Martin, Elizabeth A., Michael W. Traugott, and Courtney Kennedy. 2005. 'A Review and Proposal for a New Measure of Poll Accuracy'. *Public Opinion Quarterly* 69: 342-269.

McAllister, Ian. 2002. 'Calculating or capricious? The new politics of late deciding voters.' In David Farrell and Rüdiger Schmitt-Beck (eds.) *Do Political Campaigns Matter?* London: Routledge.

McNeish, Daniel. 2016. ‘Estimation methods for mixed logistic models with few clusters.’ *Multivariate Behavioral Research* 51: 790-804.

Mellon, Jonathan and Christopher Prosser. 2017. ‘Missing Nonvoters and Misweighted Samples: Explaining the 2015 Great British Polling Miss.’ *Public Opinion Quarterly* 8: 661–687.

Mitofsky, Warren. 1998. ‘Was 1996 a worse year for polls than 1948?’ *The Public Opinion Quarterly* 62: 230-249.

Moon, Nick. 1999. *Opinion polls: History, theory and practice*. Manchester: Manchester University Press.

Mosteller, Frederick et al. 1949. *The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-election Polls and Forecasts*. New York: Social Science Research Council.

Nakagawa, Shinichi and Holger Schielzeth. 2010. ‘Repeatability for Gaussian and non-Gaussian Data: A Practical Guide for Biologists.’ *Biological Reviews of the Cambridge Philosophical Society* 85: 935-56.

Nakagawa, Shinichi, Paul C. D. Johnson, and Holger Schielzeth. 2017. ‘The Coefficient of Determination R^2 and Intra-class Correlation Coefficient from Generalized Linear Mixed-effects Models Revisited and Expanded.’ *Journal of the Royal Society Interface* 14: 134.

Osborne, Jason. 2000. ‘Advantages of Hierarchical Linear Modeling.’ *Practical Assessment Research Evaluation* 7: 1-4.

Owen, Art. 2013. ‘Monte Carlo Theory, Methods and Examples.’ Available at: https://statweb.stanford.edu/~owen/mc/Ch-quadrature.pdf, accessed September 10, 2019.

Pickup, Mark., Matthews, J. Scott, Jennings, Will, Ford, Robert and Stephen D Fisher. 2011. ‘Why did the polls overestimate Liberal Democrat support? Sources of polling error in the 2010 British general election.’ *Journal of Elections, Public Opinion and Parties* 21: 179-209.

Prosser, Christopher, and Jonathan Mellon. 2018. ‘The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses.’ *Government and Opposition* 53: 757-790.

Raykov, Tenko. 2011. ‘Intraclass correlation coefficients in hierarchical designs: Evaluation using latent variable modelling.’ *Structural Equation Modelling* 18: 73-90.

Sanders, David. 2003. 'Pre-election polling in Britain, 1950–1997.' *Electoral Studies* 22: 1-20.

Searle, Shayle R. 1971. *Linear Models*. New York: Wiley.

Shah, Paru and Robert S. Smith. 2021. 'Legacies of Segregation and Disenfranchisement: The Road from Plessy to Frank and Voter ID Laws in the United States' *RSF: The Russell Sage Foundation Journal of the Social Sciences* 7: 134-146.

Shieh, Gwowen. 2012. 'A Comparison of Two Indices for the Intraclass Correlation Coefficient'. *Behavioral Research Methods* 44: 1212-1223.

Shugart, Matthew Soberg. 2008. 'Inherent and Contingent Factors in Reform Initiation in Plurality Systems', in *To Keep or to Change First Past the Post?*, edited by André Blais, 7-61. New York: Oxford University Press.

Shugart, Matthew Soberg and Reign Taagepera. 1989. *Seats and Votes*. New Haven: Yale University Press.

Sohlberg, Jacob and J. Alexander Branham. 2020. 'Just a difficult election to poll? How context affects polling accuracy.' *Survey Methods: Insights from the Field (SMIF)*. Available at: https://surveyinsights.org/?p=13652, accessed 02 March, 2020.

Sommet, Nicolas and Davide Morselli. 2017. 'Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS.' *International Review of Social Psychology* 30: 203-218.

Sturgis, Patrick and Will Jennings. 2017. 'Who Will Turnout and Who Will Not? The Indicator that Could Make or Break GE2017 Poll Predictions'. Available at: http://eprints.lse.ac.uk/80345/1/blogs.lse.ac.uk-Who%20will%20turnout%20and%20who%20will%20not%20The%20indicator%20that%20could%20make%20or%20break%20GE2017%20poll%20predictions.pdf, accessed February 17, 2021.

Sturgis, Patrick, Baker, Nick, Callegaro, Mario, Fisher, Stephen, Green, Green, Jennings, Will, Kuha, Joni, Lauderdale, Ben and Patten Smith. 2016. 'Report of the inquiry into the 2015 British general election opinion polls.' London, England: Market Research Society and British Polling Council.

Tilley, James. 2002. 'Political Generations and Partisanship in the UK, 1964–1997.' *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165: 121-135.

Tilley, James, and Geoffrey Evans. 2017. ‘The new politics of class after the 2017 general election.’ *Political Quarterly* 88.4: 693-697.

Traugott, Michael W., and Clyde Tucker. 1984. ‘Strategies for predicting whether a citizen will vote and estimation of electoral outcomes.’ *Public Opinion Quarterly* 48.1: 330-343.

Tudor, Jack. 2017. ‘Are Certain Elections More Predictable than Others? A Series of Bivariate Analyses of the Impact of Electoral Characteristics Upon the Predictability of British National Elections from 1945 to 2017’. Available at SSRN: https://ssrn.com/abstract=3601958, accessed June 25, 2020.

Vavreck, Lynn. 2009. *The message matters: The economy and presidential campaigns*. Princeton: Princeton University Press.

Wall, Matthew, and Jack Tudor. 2019. ‘An expected surprise? An evaluation of polls and seat forecasts during the campaign’. Available at: http://www.electionanalysis.uk/uk-election-analysis-2019/section-2-voters-polls-and-results/an-expected-surprise-an-evaluation-of-polls-and-seat-forecasts-during-the-campaign/, accessed June 25, 2020.

Weisberg, Herbert F. 2009. *The total survey error approach: A guide to the new science of survey research*. Chicago: University of Chicago Press.

Wilks-Heeg, Stuart and Peter Anderson. 2020. ‘The Only (Other) Poll That Matters? Exit Polls and Election Night Forecasts in BBC General Election Results Broadcasts, 1955-2017’. *Political Studies* 0: 1-21. Available at: https://journals.sagepub.com/doi/full/10.1177/0032321720906324, accessed February 17, 2021.

Wlezien, Christopher, and Robert Erikson. 2001. ‘Campaign Effects in Theory and Practice.’ *American Politics Research* 29: 419-436.

Wolak, Matthew E., Daphne J. Fairbairn, and Yale R. Paulsen. 2012. ‘Guidelines for Estimating Repeatability’. *Methods in Ecology and Evolution* 3: 129-137.

Table 1: Election-level breakdown of the number of in-campaign polls, average MAE value, and the proportion of 'incorrect' polls (using ratio swing approach).

Election	*Number of Polls*	*Average MAE*	*Proportion of 'Incorrect' Polls (%)*
1945	4	3.50	25
1950	10	1.54	0
1951	6	1.86	0
1955	5	0.83	20
1959	16	1.31	31
1964	18	1.96	61
1966	21	2.03	0
1970	20	2.70	100
1974 (Feb)	27	3.08	48
1974 (Oct)	27	2.56	0
1979	34	3.52	6
1983	47	3.72	0
1987	41	2.02	0
1992	48	3.20	92
1997	52	3.30	0
2001	36	3.55	0
2005	57	2.18	0
2010	84	2.93	26
2015	101	2.27	100
2017	70	3.20	81
2019	70	2.16	13
Overall	794	2.54	29

Table 2: Bayesian MCMC null model ICC estimates using half-Cauchy priors for party vote distribution polling error with standard errors and 95% credible intervals.

	MAE (standard error)	*95% CI*	*M5 (standard error)*	*95% CI*	*A (standard error)*	*95% CI*	*N (number of elections)*
UK	0.327 (0.069)	0.197 – 0.465	0.321 (0.070)	0.203 – 0.478	0.292 (0.068)	0.200 – 0.466	794 (21)
All elections in updated Jennings and Wlezien (2018) dataset	0.681 (0.013)	0.654 – 0.707	0.565 (0.016)	0.534 – 0.597	0.500 (0.021)	0.457 – 0.540	21,238 (400)
Australia	0.310 (0.065)	0.200 – 0.454	0.290 (0.059)	0.187 – 0.420	0.308 (0.054)	0.222 – 0.434	1,375 (29)
Canada	0.512 (0.068)	0.384 – 0.649	0.574 (0.066)	0.441 – 0.700	0.360 (0.066)	0.241 – 0.500	1,549 (24)
France	0.718 (0.071)	0.554 – 0.834	0.625 (0.075)	0.467 – 0.759	0.787 (0.053)	0.667 – 0.875	563 (19)
Germany	0.372 (0.086)	0.227 – 0.564	0.358 (0.081)	0.219 – 0.535	0.500 (0.021)	0.457 – 0.540	3,177 (16)
Netherlands	0.539 (0.095)	0.355 – 0.727	0.595 (0.089)	0.415 – 0.762	0.635 (0.096)	0.434 – 0.812	1,617 (13)
New Zealand	0.536 (0.085)	0.369 – 0.702	0.195 (0.062)	0.101 – 0.345	0.374 (0.083)	0.215 – 0.540	717 (15)
Norway	0.736 (0.070)	0.583 – 0.858	0.313 (0.081)	0.178 – 0.496	0.348 (0.093)	0.187 – 0.553	1,206 (14)
Spain	0.887 (0.042)	0.785 – 0.948	0.858 (0.050)	0.738 – 0.933	0.536 (0.116)	0.300 – 0.754	1,122 (11)
USA	0.508 (0.042)	0.426 – 0.592	0.514 (0.043)	0.430 – 0.598	0.500 (0.035)	0.460 – 0.596	3,838 (58)

Table 3: ICC values for null models of 'incorrect' versus 'correct' polls with standard errors in parentheses and 95% confidence and credible intervals.

	2.6	*95% CI*	*UNS*	*95% CI*	*Ratio*	*95% CI*
Frequentist Models						
Mixed Effects GLM (AGHQ)	0.908 (0.042)	0.799 – 0.965	0.850 (0.058)	0.710 – 0.936	**0.873 (0.055)**	**0.736 – 0.952**
Mixed Effects Logit (Laplace)	0.913 (0.048)	0.763 – 0.972	0.845 (0.062)	0.682 – 0.933	0.873 (0.061)	0.701 – 0.953
Mixed Effects Logit (QRD)	0.909 (0.046)	0.771 – 0.968	0.847 (0.060)	0.690 – 0.932	0.869 (0.058)	0.710 – 0.948
Bayesian Models						
Weakly Informative Priors	0.917 (0.043)	0.802 – 0.969	0.892 (0.053)	0.753 – 0.962	0.886 (0.054)	0.745 – 0.955
Non-Informative Priors	0.947 (0.036)	0.842 – 0.984	0.903 (0.052)	0.764 – 0.966	0.919 (0.052)	0.771 – 0.975
Half-Cauchy Priors	0.936 (0.039)	0.828 – 0.979	0.883 (0.055)	0.738 – 0.954	0.903 (0.051)	0.767 – 0.965

Figure 1: Boxplots describing *MAE* grouped by election, ordered by time (top pane) and median level of error (bottom pane).

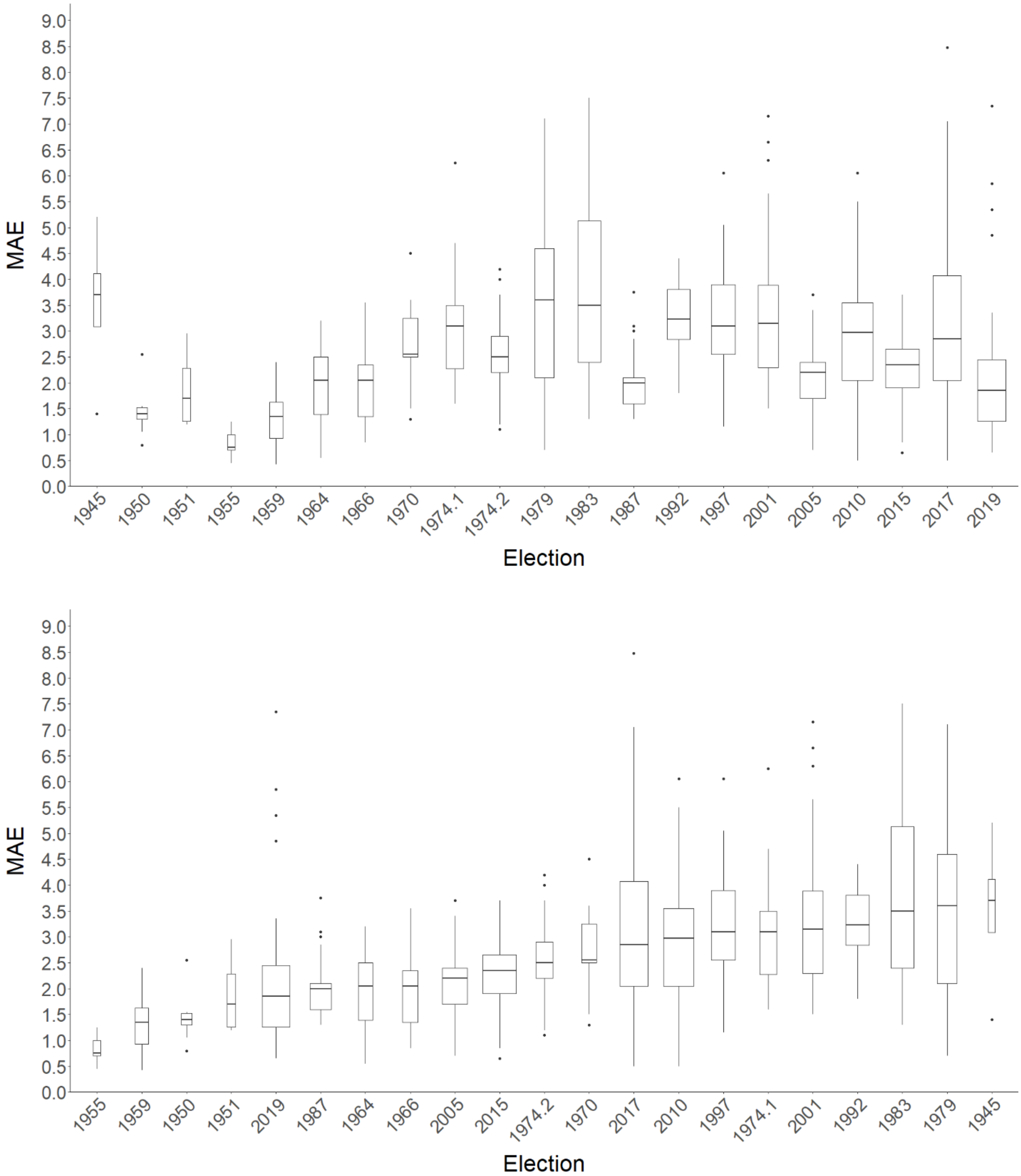

Figure 2: Stacked bar charts showing the number of (in)correct polls calculated using ratio swing in each election over time (top pane) and their proportion, ordered from highest to lowest (bottom pane).