Hierarchical model for forecasting the outcomes of binary referenda☆

Arkadiusz Wiśniowski[a,1], Jakub Bijak[b], Jonathan J. Forster[b,c], Peter W. F. Smith[b]

[a]*Social Statistics Department, University of Manchester*
[b]*ESRC Centre for Population Change, University of Southampton*
[c]*School of Mathematics, University of Southampton*

Abstract

A Bayesian hierarchical model is proposed to forecast outcomes of binary referenda based on opinion poll data acquired over a period of time. It is demonstrated how the model provides a consistent probabilistic predictions of the final outcomes over the preceding months, effectively smoothing the volatility exhibited by individual polls. The method is illustrated using opinion poll data published before the Scottish independence referendum in 2014, in which Scotland voted to remain a part of the United Kingdom, and subsequently validate it on the data related to the 2016 referendum on the continuing membership of the United Kingdom in the European Union.

Keywords: opinion polls, forecasting, Bayesian inference, statistical modelling, United Kingdom, Brexit

1. Introduction

Opinion polls provide predictions of the outcomes of voting events such as elections or referenda, based on samples drawn from the population that is eligible to vote. Forecasting from opinion polls can be of widespread interest: to the public, to

---

☆Online supplementary material is available for this article.

[1]Corresponding author. Mailing address: Social Statistics Department, University of Manchester, M13 9PL Manchester, United Kingdom. Email: a.wisniowski@manchester.ac.uk, Tel: +44 161 275 4738.

September 13, 2018

the media, to those running campaigns, or to individuals standing for election. The forecasts can also aid decision making, for example on the amount of money spent on campaigning, by providing a description of uncertainty about the outcome of the voting event.

In this article, we propose a method for predicting the outcome of two referenda recently held in the United Kingdom. The first one is the Scottish independence referendum, held on 18 September 2014, and the second one is the referendum on UK's membership in the European Union (EU) that took place on 23 June 2016. We use a Bayesian hierarchical model to capture the dynamics of the opinion polls. Share of the votes in the polls are assumed to be sampled from a multinomial distribution with overdispersion and a possibility of polling company bias. Next, logit-transformed probabilities of voting for Yes, No and those Undecided are assumed to follow a stationary Ornstein-Uhlenbeck process. We provide a framework that can be subsequently generalised to forecast voting outcomes on the polling day where there are more than two options, such as parliamentary elections with several parties, or to forecast turnout as well as the outcome.

The paper is structured as follows. In the next section, we review the literature on opinion poll forecasting, and provide a general background for both Scottish independence and EU membership referenda. Section 3 contains a description of the opinion polls data used for making predictions of referendum outcomes. In Section 4 we present the forecasting model. The results are described and analysed in Section 5, with detailed outcomes presented for the Scottish referendum, and the EU membership referendum used for external model validation. In the last section we conclude and suggest issues for further research.

2. Background

2.1. Forecasting from opinion polls

Forecasting the result of voting events has long been of great interest in democratic countries. The most prominent examples in the statistical literature describe methods for predicting the share of votes for political parties and presidential candidates in national elections, based on pre-election opinion polls (Campbell, 1992; Jackman, 2005, 2009; Lock and Gelman, 2010; Silver, 2012; Linzer, 2013). There are also models for forecasting the outcome of an election, such as a UK general election, based on exit polls (Brown et al., 1999; Curtice and Firth, 2008). These models have been successful in predicting the outcome before the final result is declared. The earlier the polls take place before the event, the less precise they are, as opinion changes over time (e.g., Campbell and Wink, 1990; Gelman and King, 1993; Campbell, 1996). Nevertheless, they can be useful in identifying trends in preferences of the electorate.

In many models used to forecast US national election outcomes, covariate information, called 'fundamentals', is used to create initial forecasts. These, in turn, are used as prior input into a binomial model (Campbell, 1992; Lock and Gelman, 2010; Linzer, 2013). Fundamentals are based on theories of retrospective voting, according to which voters tend to punish current authorities for economic or social crises (e.g., Kinder and Kiewiet, 1981; Nadeau and Lewis-Beck, 2001; Duch and Stevenson, 2008, citations after Linzer, 2013). Hence, such fundamentals include economic growth, unemployment rate and geographical variation (e.g., Campbell, 1992; Linzer, 2013).

Campbell (1992) predicted the outcome of a presidential election for each state in the US. He applies a linear regression with 16 covariates measured at national, regional and state levels, which include, amongst others, macroeconomic variables, opinion polls from early campaign, state's voting record in the previous two elections, incumbency of a candidate and measures of partisan shifts over time. The

3

assumption of using linear regression to predict an outcome bounded between zero and one is justified by the fact that the predicted shares are likely to lie around 0.5 rather than the boundaries.

Lock and Gelman (2010) analysed deviations of opinion on the state level from the nationwide average. A Bayesian model used for forecasting is based on a normal approximation to a binomial outcome. Their method integrates past election data to form prior distributions for each state outcome, and combines them with information from the state-level opinion polls to create posterior distributions of the shares of vote before the election. To account for overdispersion due to survey issues, such as weighting or clustering as well as uncertainty about opinion shifts between the forecast and the election day, the prior for variance includes information from past opinion polls averaged over states.

The model proposed by Linzer (2013) extends the Lock and Gelman (2010) approach by explicitly introducing binomial variability and allowing the model to borrow strength not only across states, but also over time. This is achieved by introducing a random walk model, similar to that of Jackman (2005), for national-level and state-level effects. The logit of the share of the votes is the sum of these effects. Again, historical state-level forecasts are used to form a prior distribution for the state-level effects.

In the models described above, the historic data on actual elections can be utilised for forecasting the outcome of presidential or parliamentary elections, as well as assessing the performance of the opinion polls in predicting their outcomes. The 2014 Scottish referendum had no direct precedent, this being the first time that the choice of full independence had been offered to the Scottish electorate. Therefore, there are no data with which to assess the influence of fundamentals on the opinion of the electorate. Furthermore, within the time frame for analysis, fundamental macroeconomic variables do not vary greatly. We consider that the influence of other factors, such as the position of the UK government and the main opposition party on

4

the independence question (both pro-Union), can be more effectively incorporated in the analysis through expert-based priors.

A Bayesian hierarchical model for the evolution of public opinion over time, based on a series of opinion polls, which does not utilise fundamentals, is the state-space model described by Jackman (2005, 2009). This model assumes that the result of the poll, denoted by y, is a normal variable with mean that includes 'true' intentions and a polling company bias (also called 'house effects'), and a variance approximated from the central limit theorem as $y(1 - y)/n$ where n is a sample size, as in Lock and Gelman (2010). This model, however, does not allow for overdispersion. Polls are pooled over time assuming random walk models for the true intentions. The polling company bias is estimated by assuming that the trajectory converges to the known result of the vote on the day of the election.

An alternative approach of Bunker and Bauchowitz (2016) is based on a two-stage model. In the first stage, the polls results are aggregated using weighting procedures that account for accuracy of the pollster, margin of error resulting from the size of the poll and timing of the poll. In the second stage, the variability of the voting preferences is taken into account by using a multinomial distribution with a conjugate Dirichlet prior.

There are also a number of articles in the literature that provide models for forecasting the outcome of the voting event on the British general election night (Brown et al., 1999; Curtice and Firth, 2008). These models are useful to predict the outcome right before the event taking place or before the final result has been published. In general, they estimate the deviations of votes from the known outcome of the previous election. The earlier the polls take place before the event, the less precise they are, as opinion changes over time (e.g., Campbell and Wink, 1990; Gelman and King, 1993; Campbell, 1996). Nevertheless, they can be useful in identifying trends in preferences of the electorate.

In our model, we extend Jackman's approach to allow us to model simultaneously

turnout and the binary outcome of the poll. We also use a logistic model which implies that the variance is not a data-dependent approximation as it is in Jackman (2005) and Lock and Gelman (2010). We model the variability of the polls over time by using a stationary Ornstein-Uhlenbeck process, which accounts for the fact that polls were infrequent in the early stage of the campaign but very frequent close to the referendum day. Following Lock and Gelman (2010), we also explicitly allow for overdispersion due to the survey issues and the fact that variability between polls is not simply a function of the underlying time-varying proportions. Finally, in our forecasts we assume that the outcome of election remains unknown to the forecasters.

2.2. The Scottish referendum of 2014

Scotland and England have had a shared sovereign since James VI and the Union of the Crowns of 1603. However, after the Acts of Union of 1707 the Scottish Parliament has been incorporated into the Parliament of the Kingdom of Great Britain, with its seat in Westminster, effectively ending the era of independence of both countries. In modern times, the question of self-rule and possible independence of Scotland has returned to the political agenda over 250 years later, mainly since the 1970s, when the referendum on establishing a separate Scottish Parliament first took place on 1 March 1979. Although 52% of voters were in favour, it did not meet the special condition outlined in the Scotland Act 1978 that 40% of the entire electorate have to vote in favour (BBC, 1997a).

As a result of the referendum on 11 September 1997, in 1999 Scottish Parliament (Holyrood) was convened by the Scotland Act 1998 (BBC, 1997b; Scottish Parliament, 2014a). The Scottish Parliament has the power to make laws in so-called Devolved Matters, such as education, justice or health. The Reserved Matters are outside the competence of the Scottish Parliament and remain with the UK Parliament in Westminster. They include, amongst others, defence, foreign policy, energy, and immigration (Scottish Parliament, 2014b).

6

The independence referendum that took place on 18 September 2014 was enabled by the Scottish Independence Referendum Act 2013 (Scottish Parliament, 2013). It asked residents of Scotland a question: 'Should Scotland be an independent country?'. If the majority of respondents had answered positively, Scotland, according to the Scottish Government White Paper *Scotland's Future* (Scottish Government, 2013), would have been able to decide about all Reserved Matters, with the monarch of the UK remaining the head of state.

The motivation for developing a model to forecast the outcome of the referendum arose from research on future Scottish migration. As mentioned above, migration policy implemented in Scotland is currently that of the entire UK (McCollum et al., 2014). Forecasts of Scottish migration, prepared before the referendum, had to take into account the uncertainty of the referendum outcome (Wiśniowski et al., 2014). This was achieved by incorporating the probability that Scotland will vote 'Yes' in the forecasting framework. This probability was used for averaging the forecasts of Scottish migration under two scenarios: independence and *status quo*. The model developed in the current paper was designed to provide this probability for particular dates prior to the referendum.

2.3. The UK membership in the EU referendum of 2016

On 1 January 1973, the UK has joined the European Communities (in 1992 transformed into the European Union; see e.g. Church and Phinnemore, 1995). This was confirmed by the 1975 European Communities referendum which saw 65% turnout and 67% voters supporting the membership (BBC, 1975). The 2016 referendum on the continuation of the membership in the European Union (EU) was called by the then Prime Minister David Cameron, amidst the rounds of renegotiating the conditions of UK's future EU membership (UK Parliament, 2016; Nugent, 2017).

The referendum took place on 23 June 2016 and, despite the opinion polls suggesting the outcome in favour of the Remain camp (e.g. BBC, 2016), 51.9% of those eligible voted to leave the EU, with turnout just above 72% (Electoral Commission,

2016). The inconsistency of the polls with the final outcome spurred a discussion on the reasons for discrepancies between the opinion polls and the outcomes of voting events, similar to the one after the 2015 general election (Sturgis et al., 2016). The opinion polls data for the EU membership referendum are used to estimate the probability of the "Leave" outcome and validate the model developed for forecasting the Scottish referendum.

3. Data

3.1. Scottish independence opinion polls data

The data on opinion polls for forecasting Scottish referendum were obtained from the UK Polling Report website[2] for each of the ten polling companies. These companies apply various methodologies for polling: YouGov – online polling from a recruited panel of respondents; Survation – online interviews, telephone and face-to-face; TNS-BMRB – face-to-face computer assisted interviews at respondents' homes; ICM – online polling with stratification by demographic characteristics; Ipsos-Mori – computer-assisted telephone interviews with quota sampling; Panelbase – online polling; Progressive – online polling and telephone interviews; Angus Reid – online polling; Ashcroft – documentation unclear but apparently face-to-face interviews, which seem to have been subcontracted to different companies; and Opinium – online interviews.[3] The first observation comes from 29 January 2012, the last polls were carried out on the eve of the referendum, on 17 September 2014. Categories of voters 'undecided' and 'not voting' have been merged together into one called 'Undecided', though the category of those 'not voting' was used in very few polls. Proportional corrections to the Undecided were made if the shares of votes did not

---

[2]Available at: http://ukpollingreport.co.uk/scottish-independence-referendum (last accessed on 22 September 2014).

[3]This information was obtained from the websites of the polling companies, current as of 22 September 2014.

sum to one. In the upper plot of Figure 1 we present the percentage of respondents to the opinion poll who declared an intention to vote in favour of independence, out of the total number of those expressing intention to vote. The lower plot shows the percentage of respondents who were classified as 'Undecided' at the time of each poll.
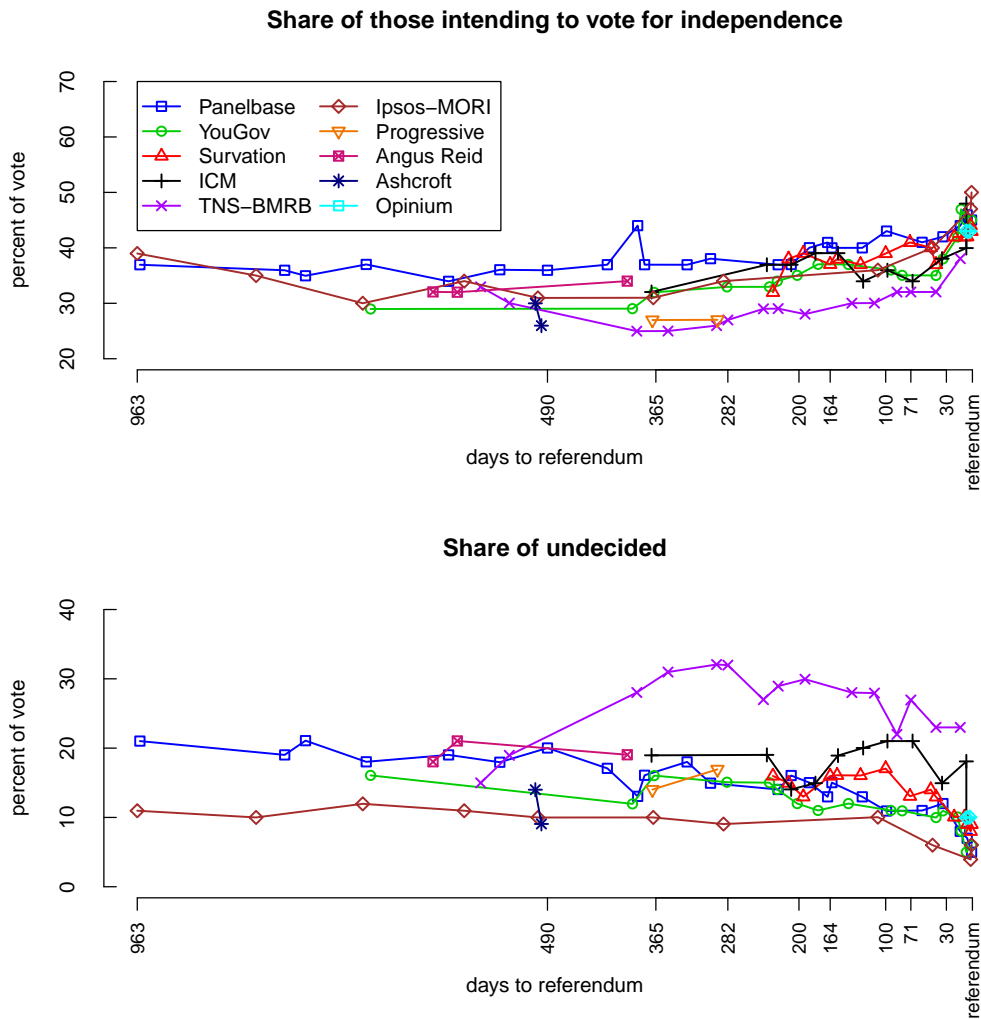
Share of those intending to vote for independence

Share of undecided

Figure 1: Opinion polls results by polling company. Available at: http://ukpollingreport.co.uk/scottish-independence-referendum (accessed last on 22 September 2014).

3.2. EU membership opinion polls data

The data on opinion polls for the EU membership referendum were obtained from the same source as above[4], and are presented in Figure 2 for each of the ten polling companies. The last category "Other" contains polls from companies that polled only once or twice during the considered period. Here, the first observation comes from 9 September 2010 and the last six polls were collected on the eve of the referendum. For clearer picture of the last days before the referendum, figures in Online Supplementary Material depict the data with the time scale converted into the natural logarithm of the number of days left till the referendum day. Given that the data come from the same source as for the Scottish referendum, the same general remarks and caveats as discussed above hold with respect to their key features, interpretations and limitations.

3.3. Data assessment and assumptions

Throughout the paper, we assume that the main sources of error are the sampling error, variability over time and overdispersion. Possible causes of overdispersion include survey issues such as non-independence of respondents or differences in methodologies used by the polling companies, as well as other factors contributing to differences in polls carried out on the same day such as measurement error. We also allow for the fact that individual polling companies may exhibit a consistent bias across time, but we assume that the average bias across companies is zero. This strong assumption is a direct consequence of the fact that we did not know the outcome of the referendum at the moment of creating a forecast. If the final result is known, the approach of Jackman (2005) can be used to learn about the polling company biases. Alternatively, if there were relevant data on the historic performance of pre-referendum opinion polls, this assumption could be relaxed.

---

[4]Available at: http://ukpollingreport.co.uk/eu-referendum (last accessed on 8 September 2017).

Share of those intending to vote for Leave the EU
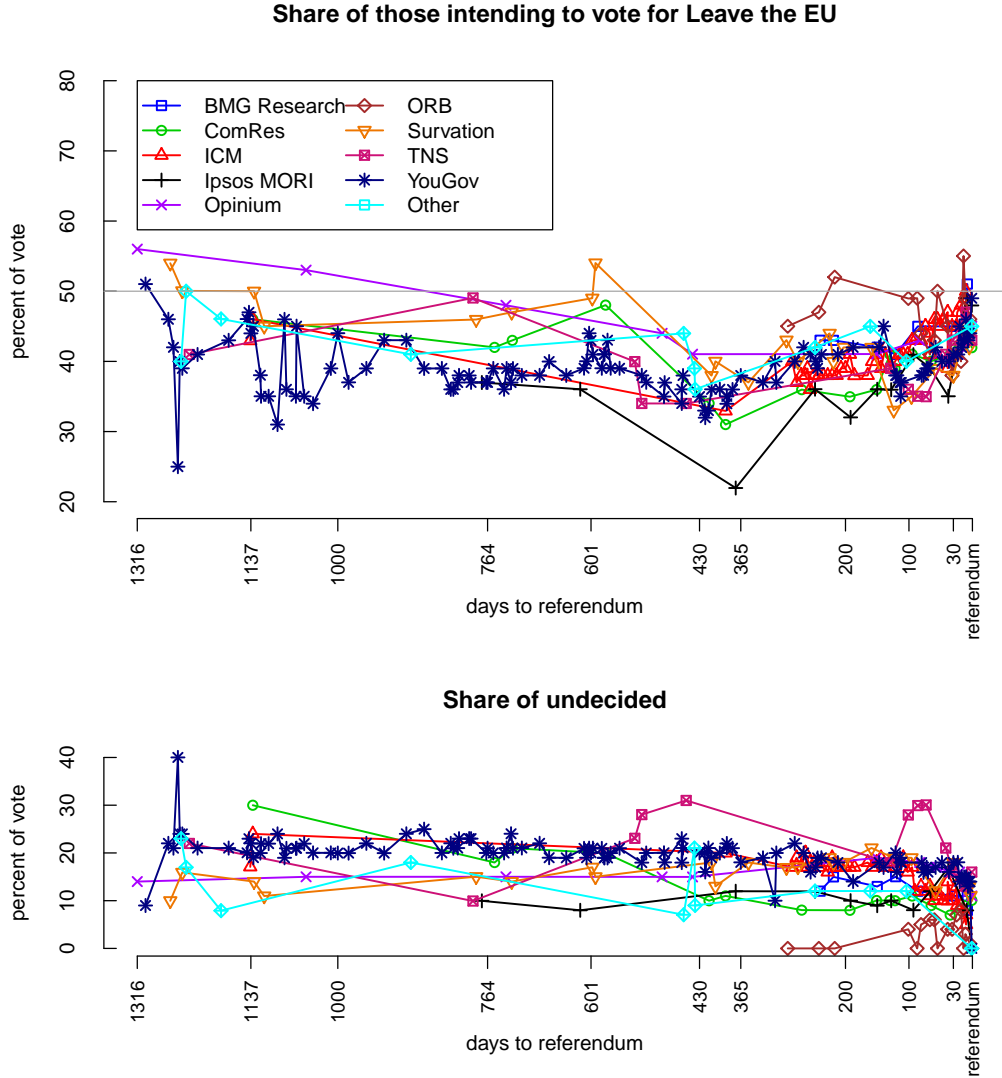

Share of undecided

Figure 2: EU membership referendum - opinion polls results by polling company

In general, the polling company bias can stem from sources such as the mode of data collection (Jäckle et al., 2010), item non-response (National Research Council et al., 2013) or type of survey (probability and non-probability surveys, see e.g. Sturgis et al., 2016; Baker et al., 2013), to name a few. For instance, face-to-face interviews may induce social desirability bias and reporting behaviours which do not actually occur. This bias tends to be reduced when using telephone interviews

(Presser and Stinson, 1998; DeBell et al., 2018). However, in the current application, we do not have access to such details for most of the used polls as we rely on available aggregate summaries of the data described in previous subsections.

We recognise that the usual polling techniques used for elections may not provide such accurate results for a referendum for which polling companies have no previous experience. The main two problems that may distort the accuracy of the opinion polls are overestimation of turnout and differential response rate (e.g., McAllister and Studlar, 1991; Lynn and Jowell, 1996; Wells, 2014). The differential response rate between prospective Yes and No voters, which, if present here, might be expected to indicate a reluctance to reveal a preference for Scotland remaining in the UK or UK leaving the EU, can lead to a bias in opinion poll results.

The one-off referendum, on one hand, may engage those who stay at home for parliamentary elections which would diminish the bias. On the other hand, polls usually overestimate the turnout, mainly due to the social desirability and acquiescence response biases which can put the validity of the entire prediction of the outcomes into question (DeBell et al., 2018). In their experimental study, Großer and Schram (2010) find that the turnout depends on the information from the opinion polls and floating (undecided) voters. A decrease in turnout can be expected if polls indicate a landslide victory of one option and vice versa, turnout increases with polls showing decreasing differences between the two options. Further, the floating voters are almost entirely responsible for the turnout boosts. Interestingly, in 2004, the pollster Populus categorized 35 percent of the electorate in the UK as floating voters (The Times, 2004).

4. Model specification

In this section we present the bespoke model for forecasting the referendum outcome based on the opinion poll data. Even though the model was initially designed to fit the data on the Scottish referendum, its generality has been subsequently

12

assessed and confirmed in the context of the referendum on the EU membership.

4.1. Forecasting model

In our application, we consider three categories for voting intentions: Yes (1), No (2) and Undecided (3). Let $y_i = (y_{i1}, y_{i2}, y_{i3})$ be a vector containing the numbers of persons declaring to vote for each category in the ith opinion poll taken at time $t(i)$ days before the referendum, by polling company $c(i)$, with a sample size n_i. For each poll i, $t(i)$ takes a value in $\{t_1, \ldots, t_D\}$, representing D distinct polling occasions, and $c(i)$ takes a value in $\{1, \ldots, 10\}$, as there are ten polling companies. To simplify, we assume that the results of a poll are published at the same time on a given day.

In particular, we assume that

$$y_i \sim \text{Multinomial}(n_i, p_i), \tag{1}$$

where $p_i = (p_{i1}, p_{i2}, p_{i3})$ is the corresponding probability vector. Let $q_i = (q_{i1}, q_{i2})$ be the logit transformation of the category Yes against No, and of the Yes and No categories jointly against the Undecided,

$$q_{i1} = \log\left(\frac{p_{i1}}{p_{i2}}\right), \quad q_{i2} = \log\left(\frac{p_{i1} + p_{i2}}{p_{i3}}\right). \tag{2}$$

The logit transformation in (2) ensures that the estimated probabilities p_{i1}, p_{i2}, p_{i3} remain bounded by 0 and 1.

For q_i we make the following assumptions. First, we assume overdispersion resulting from the issues discussed in Section 3.3. Second, we account for the possibility of polling company biases, and introduce parameter γ to capture it. Finally, we include a parameter ρ to allow for correlation between q_{i1} and q_{i2}. Hence, a hierarchical model for q_i is as follows:

$$\begin{bmatrix} q_{i1} \\ q_{i2} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \theta_1[t(i)] + \gamma_{c(i)1} \\ \theta_2[t(i)] + \gamma_{c(i)2} \end{bmatrix}, \Sigma\right), \tag{3}$$

13

where $\mathcal{N}(\mu, \Sigma)$ denotes a normal distribution with mean μ and variance matrix Σ, which we parameterise using marginal precisions τ_{21} and τ_{22} and correlation ρ. In (3), $\theta_1(t)$ and $\theta_2(t)$ are logits of the corresponding proportions in the general population at a time t days before the referendum. Such a specification assumes that the only difference between opinion polls on a given day t is due to the polling company biases $\gamma = (\gamma_{c1}, \gamma_{c2})$, $c \in \{1, \ldots, 10\}$. These effects are constrained to have a zero mean and are assumed to follow a univariate normal distribution:

$$\gamma_{ck} \sim \mathcal{N}\left(0, \tau_{3k}^{-1}\right), \quad k = 1, 2, \quad c \in \{1, \ldots, 10\}. \tag{4}$$

The zero mean here reflects the assumption, discussed in Section 3.3, that the average bias across polling companies is zero.

Next, we assume that the underlying logits $\theta_k(t)$ follow independent Ornstein-Uhlenbeck processes. For $k = 1, 2$, we define $\theta_k = (\theta_{1k}, \ldots, \theta_{Dk})$ as logit-transformed proportions in the general population on the dates of the opinion polls in our sample, so that $\theta_{jk} = \theta_k(t_j)$. Therefore, the model for θ_k is

$$\theta_{jk} \sim \mathcal{N}\left(\alpha_k^{t_j - t_{j-1}}\left[\theta_{j-1\,k} - f(t_{j-1}, \beta_k)\right] + f(t_j, \beta_k), \ \frac{1 - \alpha_k^{t_j - t_{j-1}}}{\tau_{1k}}\right), \tag{5}$$

where $j = 1, \ldots, D$ and $0 \leq \alpha_k < 1$. Function $f(t_j, \beta_k)$ is the marginal mean of q_{ik}, for any poll i taken on day t_j, and specifies the time trend using a vector parameter β_k. Parameter α_k controls the extent of serial correlation. For function f we consider constant, linear, quadratic and cubic trend specifications. For example, for the quadratic trend we have

$$f(t, \beta_k) = \beta_{k0} + \beta_{k1}t + \beta_{k2}t^2.$$

By setting the date of the referendum as the origin of the time variable t $(t = 0)$, the intercept parameter β_{10} is the marginal expected logit of the probability of a Yes outcome in the referendum. The forecasted share of votes in favour of independence, s_0, is given by

$$s_0 = \frac{\exp[\theta_1(0)]}{1 + \exp[\theta_1(0)]}, \tag{6}$$

14

where $\theta_1(0)$ is a forecast forward to $t = 0$. Hence, s_0 is the key parameter of interest taking into account both the underlying trend β_{10} and recent divergences of opinion from this trend.

The best fitting trend is selected by using the Deviance Information Criterion (Spiegelhalter et al., 2002). Time trend, autocorrelation and precision parameters are specific to the level of the model hierarchy k, as the odds Yes/No (q_{i1}) and (Yes+No)/Undecided (q_{i2}) can behave in different ways over time. In general, different specifications of the time trend f for both logits are also possible.

4.2. Prior distributions

For the model parameters, we assume the following prior distributions. For the key parameter β_{10}, the intercept in the mean model for Yes/No odds in (5), we utilise subjective expert information obtained from the Delphi questionnaire on future migration trends in Scotland (Shang et al., 2014; Wiśniowski et al., 2014). The Delphi survey was carried out in June 2013. Information was provided by eight out of 12 experts, who had been asked to state the probability that Scotland would gain independence. This probability can also be interpreted as a probability that the share of the votes will be larger than 50%. This provides partial prior information about the experts' uncertainty about the share of the votes which we utilise to construct the prior distribution for β_{10} as follows.

We assume that the prior distribution is an equally-weighted mixture of normal distributions, constructed for each of the eight experts. Since β_{10} represents the logit of the expected share of votes on the referendum day, we assume that the probability of share of votes being larger than 50% provided by expert l is

$$P_l = P\left(\log\frac{S_l}{1 - S_l} > 0\right) = 1 - \Phi\left(-\frac{\mu_l}{\sigma_l}\right), \quad l = 1, \ldots, 8, \tag{7}$$

where Φ is the distribution function of a standard normal distribution. Here S_l is a random variable representing the lth expert's opinion on the share of votes on a referendum day, which we assume to have a logistic-normal distribution with mean

μ_l and variance σ_l^2. As full information was not provided by the experts about their uncertainty concerning the share of the votes, there are 16 unknowns in eight equations (7). Therefore, we assume that each expert's opinion is equivalent to a poll of size n. In such a poll, we would have $nS_l \sim \text{Binomial}(n, p_l)$. Hence, by applying the delta method we obtain

$$\sigma_l^2 = Var\left(\log \frac{S_l}{1 - S_l}\right) \approx Var(S_l)\left(\frac{1}{p_l(1 - p_l)}\right)^2 = \frac{1}{np_l(1 - p_l)}. \qquad (8)$$

Since this variance is the largest for $p_l = 0.5$, we simplify (8) to $\sigma_l^2 \approx 4/n$, for $l = 1, \ldots, 8$. For our application, we assume that each expert is equal to $n = 125$ respondents in a representative opinion poll, so that a total of eight experts is equivalent to 1000 respondents, which is a typical size for an opinion poll. In the sensitivity analysis below we assess departures from this assumption by considering $n = 1$ as well as an approximately uniform prior.

For the other elements of the vector β_k we use normal distributions centred around zero with very large variances:

$$\beta_{ki} \sim \mathcal{N}(0, 10^4), \quad i, k = 1, 2.$$

These priors are extremely diffuse, which in principle allows opinion to exhibit variation in time which is much more rapid than might be considered plausible. However, when we use priors with much smaller variances restricting variation in opinion to a more plausible range, the final results were indistinguishable from those with the diffuse priors, indicating the lack of sensitivity to the precise values of the prior variances.

For α_k we assume a uniform distribution on $(0, 1)$, which prevents this parameter taking negative values or exhibiting explosive behaviour:

$$\alpha_k \sim \mathcal{U}(0, 1), \quad k = 1, 2.$$

For precisions τ_{1k}, τ_{2k} and τ_{3k}, we assumed truncated normal distributions:

$$\tau_{1k} \sim \mathcal{N}\left(0, 10^6\right) \mathbb{1}(\tau_{1k} \geq 0), \quad k = 1, 2,$$

$$\tau_{2k} \sim \mathcal{N}\left(0, 10^6\right) \mathbb{1}(\tau_{2k} \geq 0), \quad k = 1, 2,$$

$$\tau_{3k} \sim \mathcal{N}\left(0, 10^6\right) \mathbb{1}(\tau_{3k} \geq 0), \quad k = 1, 2,$$

where $\mathbb{1}(g)$ is an indicator function taking one if g is true, and zero otherwise. Finally, for the correlation parameter ρ, we assume

$$\rho \sim \mathcal{U}(-1, 1).$$

To select the prior distributions for the precisions including the values of the hyperparameters, we analysed the model with data simulated to resemble opinion polls related to the Scottish referendum. This simulation study demonstrated that the use of diffuse conjugate gamma priors (e.g. $\Gamma(10^{-3}, 10^{-3})$) for the precision parameters τ_{k1} and τ_{k2} results in poor mixing of the MCMC algorithm and undesirable levels of sensitivity. Hence, following recommendations of Gelman (2006), we investigated (i) diffuse half-normal prior distributions for precisions, and (ii) uniform distributions over large range for standard deviations. If the number of days with multiple polls was small, the overdispersion parameter τ_{k2} was weakly identified. This led to poor convergence of the MCMC algorithm and high correlation between τ_{k1} and τ_{k2}. The truncated normal distributions performed considerably better than the uniform distributions.

5. Results of forecasting referendum outcomes

5.1. General remarks

Predictive distributions of the referendum result, based on the opinion poll data and expert opinion, were obtained by using OpenBUGS (Lunn et al., 2009) and JAGS with `rjags` R package (Plummer et al., 2003, see code repository). As discussed in Section 4.1, s_0 in (6) in the context of the Scottish referendum is the

17

share of the Yes votes, in favour of independence, on the referendum day, and hence uncertainty about the referendum outcome is represented by the posterior distribution of s_0. The probability of an outcome favouring independence is therefore calculated as the probability mass of the predicted share of the Yes votes larger than 50 percent, i.e. $P(s_0 > 0.5)$. Similar operationalisation has been adopted for the EU membership referendum, where s_0 relates to the share of the Leave votes, favouring the withdrawal of the UK from the European Union.

Below, we first present the history of forecasts for the Scottish independence based on the data one year, and then 200, 100, 30 and seven days before the referendum. Next, we use the data that include the polls carried out within the week prior to the referendum as this was the period when nine additional polls were carried out. The time trend with the lowest DIC is quadratic for all subsets of the polling data. Finally, we apply the model to the EU membership referendum data, in order to externally validate the robustness of the method.

5.2. Scottish referendum: History of forecasts

Forecasts of the result of the referendum, together with the opinion poll data until one week before the referendum, are presented in the upper panels of Table 1 and Figure 3. In the last three columns of Table 1, we present the median and quartiles of the predicted share of Yes votes (excluding Undecided). We observe that the median is almost constant over time, at around 46 percent (with a drop to 44.7 percent one month before the referendum), but the predictive distribution of the share of votes for independence is shrinking with additional observations utilised in the model. An earlier version of our model[5], which did not include the polling company effect γ, produced undesirable sensitivity of the forecasts to the order in which different companies published their polls and led to an overestimation of the

---

[5]Presented in the Monkey Cage blog at the Washington Post: Scottish independence vote is too close to call (30 July 2014) and Odds of a Scottish Yes vote are fading fast (20 August 2014).

proportion in favour of independence at certain dates.

The results indicate that one year before the referendum, there was still room for a change in the opinion for both Yes and No camps. Closer to the referendum, however, the chances of any dramatic shifts were shrinking. The decrease observed one month before indicates the influence of the debate between the First Minister of Scotland, Alex Salmond, and the leader of the pro-Union campaign, Alistair Darling, on 5 August 2014, deemed to have been lost by the former. Even in the final days of campaigning, the forecasts remained at a similar level, which can be explained by relatively large differences amongst the polls carried out by various companies in the last week. Hence, the uncertainty of the predicted votes share remained rather large, with the interquartile range of 45.2–48.1 percent.

The probability of independence decreased over time from 23 percent one year prior to the referendum, to two percent with 30 days to go (and after the first TV debate on 5 August 2014). The upward shift in the polls conducted a week before the referendum led to an increase of the probability to five percent, which might have been a result of the second TV debate on 25 August, said to have been won by Alex Salmond. The changes in this probability over time are summarised in the upper panel of Table 1 and the corresponding distributions of the share of the Yes vote are presented in the upper panel of Figure 3.

5.3. Scottish referendum: On-the-eve forecasts

Our final prediction, based on the opinion polls reported up to 17 September 2014, i.e. on the eve of the referendum, is presented in Figure 4, together with the referendum outcome on 18 September, which was 44.7 percent in favour for independence. This outcome is approximately the 18th percentile of our predictive distribution. Despite the last-minute increase in the probability of independence to seven percent, the median share of Yes votes, 46.7 percent, remained at a similar level to the earlier predictions. The final result also suggests that the opinion polls carried out shortly before the referendum might have been biased towards independence, as
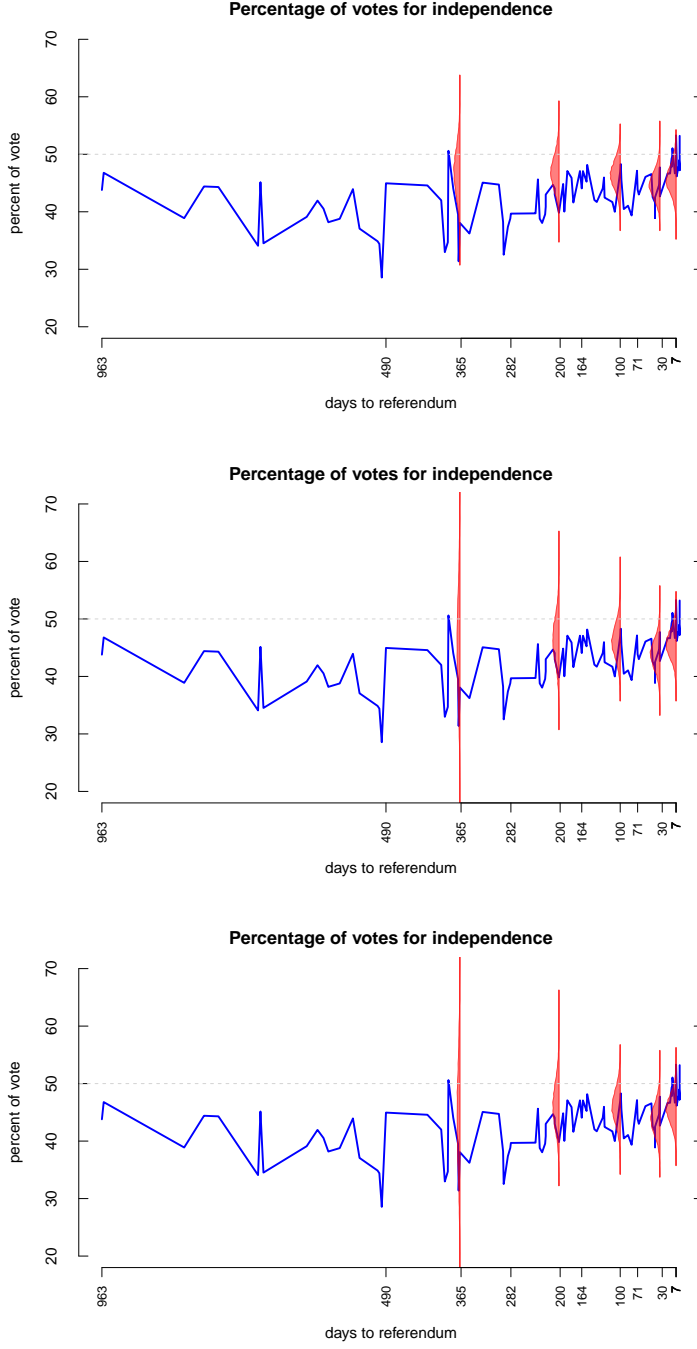
19

Figure 3: Time series of the opinion poll results, together with probability distributions of the percentage of the votes for independence on the referendum day; expert-based prior with $n = 125$ in upper panel, expert-based prior with $n = 1$ in the second panel, approximately uniform prior in the third panel; the probability distributions of the percentage of votes relate to the day of the referendum and are plotted on the time scale to denote the moment when the forecast was made; shares of votes relate to responses favouring independence and excluding undecided and those not voting.

Table 1: Probability of independence and quartiles of the Yes votes share

Days to referendum	Date of last poll in sample	Probability of independence	Percentage of Yes votes		
			1st quartile	Median	3rd quartile
Expert-based prior ($n = 125$)					
365	16/09/2013	23%	44.0%	47.0%	49.8%
200	28/02/2014	14%	44.8%	46.8%	48.8%
100	10/06/2014	8%	44.8%	46.5%	48.1%
30	15/08/2014	2%	43.2%	44.7%	46.3%
7	11/09/2014	5%	44.1%	45.7%	47.4%
1	17/09/2014	7%	45.2%	46.7%	48.1%
Sensitivity analysis					
Expert-based prior ($n = 1$)					
365	16/09/2013	27%	38.3%	44.4%	50.6%
200	28/02/2014	17%	43.4%	46.2%	48.9%
100	10/06/2014	9%	44.0%	46.0%	48.0%
30	15/08/2014	1%	42.0%	43.8%	45.6%
7	11/09/2014	4%	43.9%	45.6%	47.2%
1	17/09/2014	7%	44.7%	46.3%	47.9%
Approximately uniform prior					
365	16/09/2013	32%	39.7%	45.8%	51.8%
200	28/02/2014	18%	43.8%	46.4%	49.1%
100	10/06/2014	9%	44.2%	46.1%	48.1%
30	15/08/2014	1%	42.1%	43.9%	45.6%
7	11/09/2014	4%	43.9%	45.6%	47.3%
1	17/09/2014	5%	44.6%	46.3%	47.8%

they all predicted a higher share of the Yes votes than the final outcome.

The predicted share of the undecided on the day of election was 8.5 percent with the 95% interval being 6.8–10.6 percent. If we assume that all 'decided' vote, this provides an estimate of a turnout of 91.5 percent. The actual turnout was at a lower level of 84.6 percent which aligns with the findings about the overestimation of turnout described in Section 3.3. Finally, the comparison of the models with quadratic and cubic trends, which in the end were the two competing models, reveals that the differences in the predictions were minimal.

5.4. Scottish referendum: Sensitivity analysis of the priors

To assess sensitivity of the results to the expert-based prior on β_{10} we considered two priors where the expert opinion is moderated. In the first, we reduce the effective

Percentage of votes for independence

Referendum result

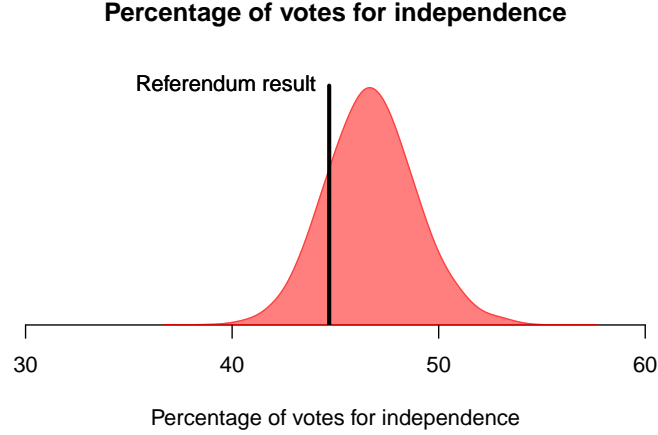30   40   50   60

Percentage of votes for independence

Figure 4: Probability distribution of the percentage of the votes for independence (red) on the eve of the referendum with the referendum result (black). Shares of votes relate to responses favouring independence and exclude those undecided or not voting.

weight of each expert to $n = 1$, and in the second we use a diffuse prior $\beta_{10} \sim \mathcal{N}(0, 2)$, which is approximately equivalent to a uniform prior on $(0, 1)$ for $E(s_0)$. The results are presented in the second and third panels of Table 1 and Figure 3. As would be expected, the expert-based prior has an influence on the predictive distributions at the earlier polling dates, in 2013 and the spring of 2014, effectively shrinking the predictive distributions. However, there is little sensitivity to the choice of the prior in the run-up to the referendum, as the polling evidence outweighs the earlier expert opinion.[6] As a result, the discrepancy between the predicted results close to the referendum day is negligible.

---

[6]For the on-the-eve prediction, the true referendum outcome is close to the 25th percentile of the predictive distributions obtained by using either the expert-based prior with $n = 1$ or the approximately uniform prior.

22

5.5. Model validation: EU membership referendum

The EU membership referendum carried out in the United Kingdom in 2016 enabled us to carry out an external validation exercise for the model proposed in Section 4. As in this case we did not have access to expert opinion which would be comparable to the information elicited for the Scottish referendum (Section 3), the estimation only includes a vague prior distribution for the parameter β_{10}. As before, the best-fitting trend has been also selected by using the Deviance Information Criterion, with the caveat that forecasts done for different pre-referendum cut-off dates could have been based on different trend functions. In particular, the on-the-eve predictions based on all three trend functions yielded similar predictions as well as approximately the same Deviance Information Criteria.

The specific results of the forecasts are presented in Figures 5-6, as well as in Table 2. Overall, as was the case with the Scottish independence referendum, there is a visible narrowing of predictive intervals over time. Towards the end of the prediction horizon, and especially on the eve of the referendum, the predictions formally remained inconclusive, with the inter-quartile range of the predictive distributions covering 50 percent of the Leave vote. Nevertheless, as seen especially in Table 2 and Figure 6, the odds were only slightly for the Leave vote winning. As it happened, the final share of the Leave vote was 51.9 percent. The predicted share of the Undecided vote, a proxy for the share of the electorate that did not vote, was 10.2 percent, with a 95% predictive interval between 7.9 to 13.2 percent. Hence, the model predicted the turnout of 89.8 percent which is again quite well above the actual turnout of 72.2 percent. Here, a likely explanation might be an estimated lower turnout of the young voters aged 18-34 compared to those aged 55-74 (Ipsos-Mori, 2016). A speculative hypothesis for that behaviour might be that, on the day of the referendum, a victory of the Remain camp has been expected as revealed by, e.g., the last published poll which showed 48 to 46 percent for Remain[7]. This

---

[7]Available at: https://ig.ft.com/sites/brexit-polling/ (last accessed on 13 August 2018).

explanation would align with the findings of Großer and Schram (2010). Overall, as in the case of the Scottish referendum, the model performed reasonably well, clearly indicating the uncertainty around the polling outcome.
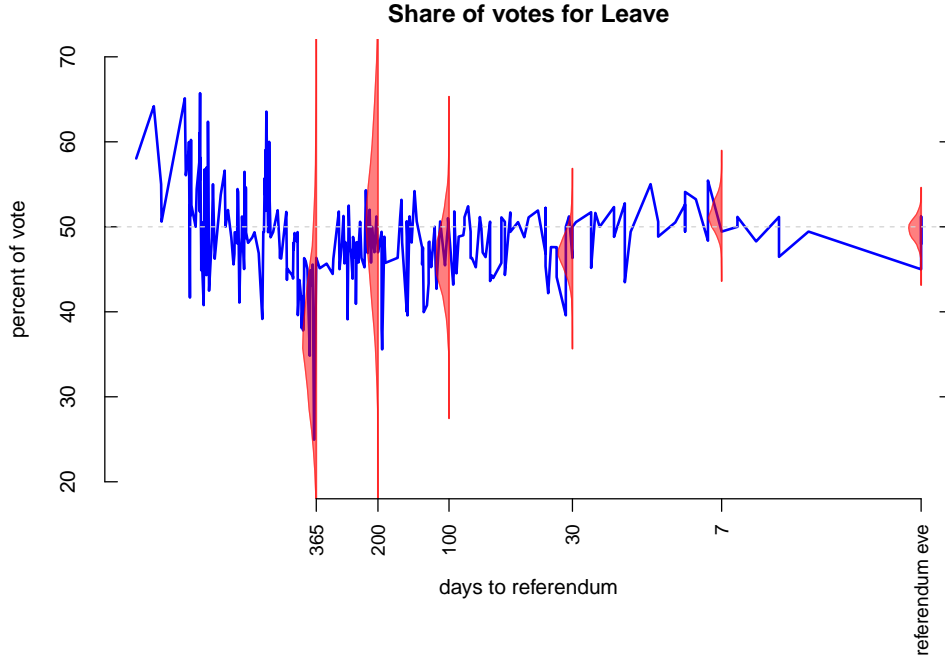
Share of votes for Leave

Figure 5: Probability distribution of the share of the votes for Leave on the referendum day (red) with the opinion polls results (blue line). Probability distributions of the share of votes relate to the day of the referendum. They are plotted on the time scale (logged) to denote the moment when the forecast was made. Shares of votes relate to responses favouring withdrawal from the EU and excluding undecided and those not voting.

Table 2: Probability of leaving the EU and quartiles of the Leave vote share

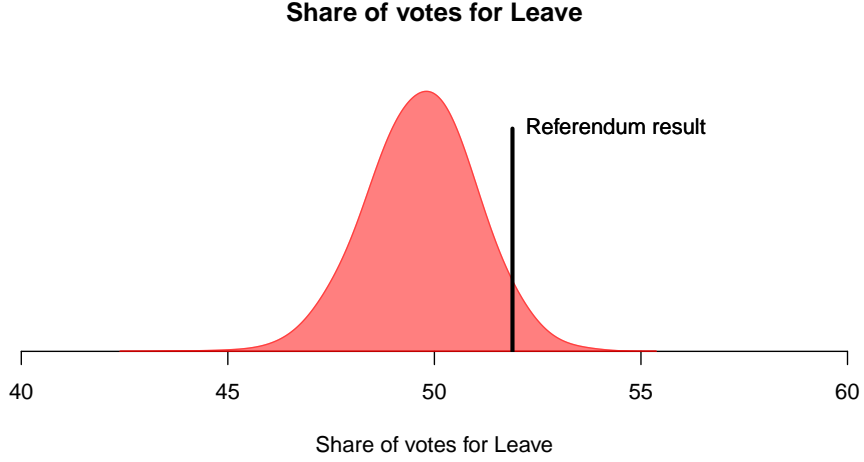| Days to referendum	Date of last poll in sample	Probability of Leave	Percentage of Leave votes		
			1st quartile	Median	3rd quartile
365	24/06/2015	6%	32.0%	36.9%	42.2%
200	06/12/2015	54%	45.0%	50.7%	56.4%
100	15/03/2016	19%	43.8%	46.4%	49.1%
30	24/05/2016	9%	45.5%	46.9%	48.4%
7	16/06/2016	78%	50.2%	51.2%	52.5%
1	22/06/2016	42%	48.9%	49.7%	50.5%

Share of votes for Leave

Figure 6: Probability distribution of the percentage of the votes for Leave (red) on the eve of the referendum with the referendum result (black). The actual outcome is the 95th percentile of the posterior distribution.

5.6. Model sensitivity: time trend specification

We analyse the use of cubic B-splines to specify the time function $f(t, \beta_k)$ in (5). This can be viewed as a more general specification of the marginal means of the logit q_{ik} in (2) that might be useful in the cases where the long-term time pattern exhibits more erratic behaviour.

In particular, we use two knots that split the data set into three subsets with a similar number of observations in each of them. The indices for the splines are created using function `bs()` from package `splines` in R (R Core Team, 2018). The results for both Scottish and EU membership referenda are mildly sensitive to the use of the B-splines. For the long-term forecasting horizons (100 days before or longer), we observe relatively large uncertainty in predictive posteriors for the vote share under study. These differences diminish for short-term horizons. On the eve of the referendum, the predicted probability of Scottish independence is estimated to be 13 percent compared to 7 percent in Table 1; for the EU membership the corresponding probability of Leave is 43 percent compared to 42 percent in Table 2. The DIC yielded by the models based on B-splines for short-term forecasts are also

25

close to the ones produced by the polynomial trends. Hence, in the cases where no informed judgement can be made about the long-term behaviour of the voting intentions, we recommend Bayesian averaging (Hoeting et al., 1999) of the results to account for the uncertainty about the trend specification for the mean.

6. Discussion

In this paper we have proposed a method of estimating the probability of the outcome of a referendum by using opinion poll data. It takes into account the variation between the polls over time, between the polls carried on the same day, and accounts for the biases of various polling companies. It smooths out the volatility of the opinion polls carried out before the referendum, yet is still able to capture swings in voting intentions. The model can also predict approximate turnout in addition to the outcome of the vote. The method has been applied to the 2014 independence referendum in Scotland and validated on the data for the 2016 UK referendum on continuing membership of the European Union. The actual outcomes of both referenda fall well within the respective predictive probability distributions obtained from our model, and in both cases are close to our predicted median values, with differences of around ±2 percentage points.

The predictions generated by the model proved quite robust with respect to the specification of the underlying time trend function, especially for the short-term horizons. For long-term horizons, B-splines tend to increase uncertainty compared to using single polynomial function. The model also yielded results comparable to those that could be obtained by using other methods and data sources, such as the prediction markets or bookmakers' bets, while retaining the advantage of much more modest data demands. In particular, for the EU referendum, according to Auld and Linton (2018), the prediction markets or currency markets alone were indicating a probability of approximately 0.1 for Leave until the polls had closed, whereas a Bayesian model estimated on their basis, using real-time, high-frequency

data, had this probability estimated as approximately 0.37 (Auld and Linton, 2018, Figure 10 on p. 36). The latter (0.37) is more in line with our main finding. In that respect, for the longer time horizons, the additional information gains provided by the prediction or currency markets can be seen as negligible. For the Scottish referendum, on the other hand, the prediction markets performed well, and similarly to the model presented in this paper, picking up the second leaders' debate as the decisive turning point in favour of Scotland remaining in the Union (Wall et al., 2017).

Further research is required for applications such as election forecasting, where additional covariates, such as fundamentals or outcomes of historical events, as well as information on past performance of the polls can be utilised. This information could be used for forecasting the spatial distribution of votes, important for the prediction of results for individual electoral districts, for example in the UK parliamentary elections, or the US presidential elections. It could be also used to strengthen the predictions of turnout, which in the current version is biased compared with the actual outcome but is only based on the share of Undecided (i.e. floating) voters whose decisions about turning out are difficult to predict until the actual voting event (Großer and Schram, 2010). Further, the bias across polling companies could be addressed by utilising, if available, information on the mode of data collection, non-response or type of the survey. This information can be incorporated in our modelling framework by including, for example, a generalised linear model for parameters γ in (4); alternatively, informative priors reflecting potential biases can be constructed based on expert assessment of the polls. Another extension could be to perform Bayesian model averaging to account for the uncertainty about the particular time trend used to make the prediction. Finally, the time trend could be replaced by other semi- or non-parametric function, such as a normal kernel or other types of splines. Still, after necessary modifications – mainly extending the two-level model hierarchy to more levels – the proposed framework can be easily

generalised to forecast voting outcomes with more than two options.

Acknowledgements

References

Auld, T. and Linton, O. (2018). The behaviour of betting and currency markets on the night of the EU referendum. Cemmap working paper CWP 01/18, available at https://www.ifs.org.uk/uploads/CWP011818.pdf (accessed on 20 July 2018).

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., and Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1(2):90–143.

BBC (1975). 1975: UK embraces Europe in referendum. British Broadcasting Corporation. Online material, available at: http://news.bbc.co.uk/onthisday/hi/dates/stories/june/6/newsid_2499000/2499297.stm (accessed on 08 September 2017).

BBC (1997a). Scottish Devolution. Briefing by Brian Taylor, Political Editor, BBC Scotland. British Broadcasting Corporation. Online material, available at: http://www.bbc.co.uk/news/special/politics97/devolution/scotland/briefing/scotbrief1.shtml (accessed on 12 February 2014).

BBC (1997b). Scottish Referendum Live – The Results. British Broadcasting Corporation. Online material, available at: http://www.bbc.co.uk/news/special/

`politics97/devolution/scotland/live/index.shtml` (accessed on 12 February 2014).

BBC (2016). EU Referendum: Did the polls all get it wrong again? British Broadcasting Corporation. Online material, available at: `http://www.bbc.co.uk/news/uk-politics-eu-referendum-36648769` (accessed on 08 September 2017).

Brown, P. J., Firth, D., and Payne, C. D. (1999). Forecasting on British election night 1997. *Journal of the Royal Statistical Society: Series A*, 162(2):211–226.

Bunker, K. and Bauchowitz, S. (2016). Electoral Forecasting and Public Opinion Tracking in Latin America: An Application to Chile. *Política*, 54(2):207–233.

Campbell, J. (1992). Forecasting the Presidential Vote in the States. *American Journal of Political Science*, 36:386–407.

Campbell, J. E. (1996). Polls and Votes The Trial-Heat Presidential Election Forecasting Model, Certainty, and Political Campaigns. *American Politics Research*, 24(4):408–433.

Campbell, J. E. and Wink, K. A. (1990). Trial-heat forecasts of the presidential vote. *American Politics Research*, 18(3):251–269.

Church, C. H. and Phinnemore, D. (1995). *European Union and European Community: a handbook and commentary on the 1992 Maastricht treaties.* London: Prentice Hall.

Curtice, J. and Firth, D. (2008). Exit polling in a cold climate: the BBC–ITV experience in Britain in 2005. *Journal of the Royal Statistical Society: Series A*, 171(3):509–539.

DeBell, M., Krosnick, J. A., Gera, K., Yeager, D. S., and McDonald, M. P. (2018). The turnout gap in surveys: Explanations and solutions. *Sociological Methods & Research*, page 0049124118769085.

Duch, R. M. and Stevenson, R. T. (2008). *The economic vote: how political and economic institutions condition election results, Political economy of institutions and decisions*. New York: Cambridge University Press.

Electoral Commission (2016). EU referendum results. Online material, available at: http://www.electoralcommission.org.uk/find-information-by-subject/elections-and-referendums/past-elections-and-referendums/eu-referendum/electorate-and-count-information (accessed on 06 November 2017).

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3):515–534.

Gelman, A. and King, G. (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science*, 23(04):409–451.

Großer, J. and Schram, A. (2010). Public opinion polls, voter turnout, and welfare: An experimental study. *American Journal of Political Science*, 54(3):700–717.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401.

Ipsos-Mori (2016). How Britain voted in the 2016 EU referendum. Ipsos MORI. Online material, available at: https://www.ipsos.com/ipsos-mori/en-uk/how-britain-voted-2016-eu-referendum (accessed on 13 August 2018).

Jäckle, A., Roberts, C., and Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, 78(1):3–20.

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

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*, volume 846. Chichester: John Wiley & Sons.

Kinder, D. R. and Kiewiet, D. R. (1981). Sociotropic politics: the American case. *British Journal of Political Science*, 11(02):129–161.

Linzer, D. A. (2013). Dynamic Bayesian forecasting of Presidential Elections in the States. *Journal of the American Statistical Association*, 108(501):124–134.

Lock, K. and Gelman, A. (2010). Bayesian combination of state polls and election forecasts. *Political Analysis*, 18(3):337–348.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067.

Lynn, P. and Jowell, R. (1996). How might opinion polls be improved?: The case for probability sampling. *Journal of the Royal Statistical Society. Series A*, 159(1):21–28.

McAllister, I. and Studlar, D. T. (1991). Bandwagon, underdog, or projection? Opinion polls and electoral choice in Britain, 1979–1987. *The Journal of Politics*, 53(03):720–741.

McCollum, D., Nowok, B., and Tindal, S. (2014). Public attitudes towards migration in Scotland: exceptionality and possible policy implications. *Scottish Affairs*, 23(1):79–102.

Nadeau, R. and Lewis-Beck, M. S. (2001). National economic voting in US presidential elections. *Journal of Politics*, 63(1):159–181.

National Research Council et al. (2013). *Nonresponse in social science surveys: A research agenda*. National Academies Press.

Nugent, N. (2017). *The government and politics of the European Union*. London: Palgrave, 3 edition.

Plummer, M. et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna, Austria.

Presser, S. and Stinson, L. (1998). Data collection mode and social desirability bias in self-reported religious attendance. *American sociological review*, pages 137–145.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Scottish Government (2013). Scotlands Future: Your Guide to an Independent Scotland. The Scottish Government: Edinburgh.

Scottish Parliament (2013). Scottish Independence Referendum Bill. Online material, available at: http://www.scottish.parliament.uk/parliamentarybusiness/Bills/61076.aspx (accessed on 17 May 2015).

Scottish Parliament (2014a). History. Online material, available at: http://www.scottish.parliament.uk/visitandlearn/9982.aspx (accessed on 10 February 2014).

Scottish Parliament (2014b). How the parliament works: evolved and reserved matters explained. Online material, available at: http://www.scottish.parliament.uk/visitandlearn/25488.aspx (accessed on 10 February 2014).

Shang, H. L., Bijak, J., and Wiśniowski, A. (2014). Directions of impact of Scottish independence on migration: A survey of experts. ESRC Centre for Population Change: University of Southampton.

Silver, N. (2012). *The signal and the noise: Why so many predictions fail-but some don't*. London: Penguin.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002).

Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64(4):583–639.

Sturgis, P., Nick, B., Mario, C., Stephen, F., Jane, G., Jennings, W., Jouni, K., Ben, L., and Patten, S. (2016). *Report of the Inquiry into the 2015 British general election opinion polls*. London: Market Research Society and British Polling Council.

The Times (2004). Boost for Kennedy as Blair and Howard slip. The Times. Online material, available at: https://www.thetimes.co.uk/article/boost-for-kennedy-as-blair-and-howard-slip-l3ksxb2fx7l (accessed on 13 August 2018).

UK Parliament (2016). Prime Minister sets out legal framework for EU withdrawal. Online material, available at: https://publications.parliament.uk/pa/cm201516/cmhansrd/cm160222/debtext/160222-0001.htm#16022210000001 (accessed on 08 September 2017).

Wall, M., Costello, R., and Lindsay, S. (2017). The miracle of the markets: Identifying key campaign events in the Scottish independence referendum using betting odds. *Electoral Studies*, 46(1):39–47.

Wells, A. (2014). Will the polls get the Scottish referendum right? Online blog material, available at http://ukpollingreport.co.uk/blog/archives/8987 (accessed on 16 September 2014).

Wiśniowski, A., Bijak, J., and Shang, H. L. (2014). Forecasting Scottish migration in the context of the 2014 constitutional change debate. *Population, Space and Place*, 20(5):455–464.

Hierarchical model for forecasting the outcomes of binary referenda
Supplementary Material

Arkadiusz Wiśniowski[a,1], Jakub Bijak[b], Jonathan J. Forster[b,c], Peter W. F. Smith[b]

[a]*Social Statistics Department, University of Manchester*
[b]*ESRC Centre for Population Change, University of Southampton*
[c]*School of Mathematics, University of Southampton*
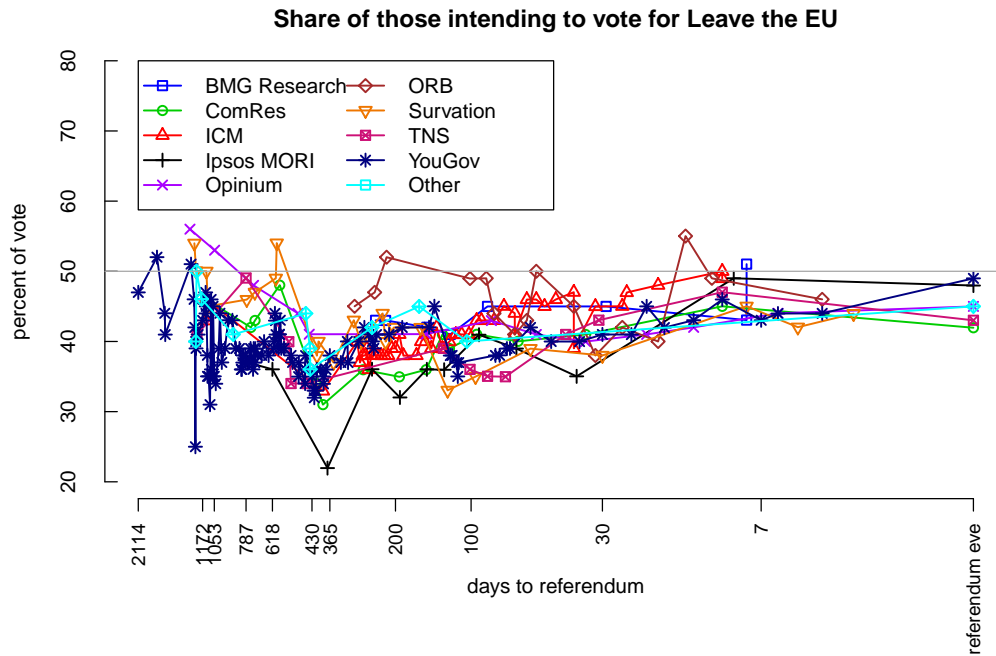
Appendix A. Plots of the data

Figure Appendix A.1: EU membership referendum - opinion polls results by polling company. We use natural logarithm of the number of days to the referendum (horizontal axis) for better visibility of the polls closer to the referendum day.

---
[1]Corresponding author. Email: a.wisniowski@manchester.ac.uk.
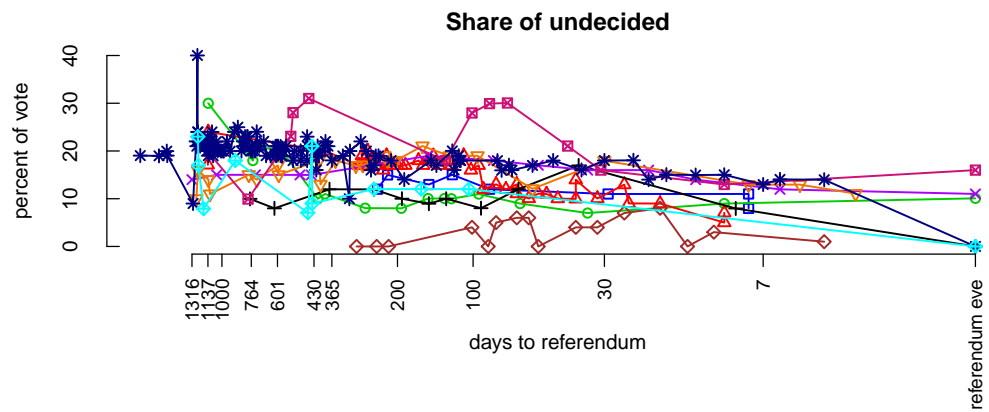
September 13, 2018

Figure Appendix A.2: EU membership referendum - opinion polls results by polling company. We use natural logarithm of the number of days to the referendum (horizontal axis) for better visibility of the polls closer to the referendum day.