Forecasting Multi-District Elections *

Rohan Alexander   *University of Toronto*
Patrick Dumont   *Australian National University*
Lauren Kennedy   *Columbia University*
Patrick Leslie   *Australian National University*

The overall outcome of an election in a multi-district system turns on the number of electoral divisions won by each party. Despite this, traditional political opinion polls tend to focus on broad national-level estimates of support for major parties. In this paper we develop a framework to characterise a useful forecasting model of multi-district elections. We then implement an approach to forecast the number of electoral divisions won by each of the larger political parties in Australia at the 2019 Federal Election. Our approach first uses survey data and multi-level regression with post-stratification (MRP) to estimate electoral-division-level first-preference shares. To account for Australia's use of preferential voting, we then use a model that we trained to estimate electoral-division-level two-party-preferred shares. We find that our approach performs well out-of-sample, and it has additional advantages over traditional polling including improved interpretability, transparency, and better communication of statistical uncertainty. Our framework allows the consistent evaluation of forecasting models of elections in a multi-district system. Finally, our paper also improves our understanding of the circumstances in which MRP is appropriate, as we find that despite the bias of voter advice applications, and despite the relatively small number of respondents per electoral division for usually political opinion polls, both can still provide informative results on an electoral division basis.

Keywords: Multi-district elections, political opinion polls, forecasting, multi-level regression with post-stratification, Australia.

1   Introduction

Political opinion polling is an integral aspect of modern politics. For instance, campaigns use it to inform strategic decisions (Pereira, 2019), and the media draw on it for content (Mills and Tiffen, 2012; Toff, 2019). But, not only are political opinion polls often just wrong, they are rarely transparent (Young, 2010, p. 188), and statistical uncertainty is commonly downplayed (Oleskog Tryggvason and Strömbäck, 2018; Reavy and Pavlick, 2018). Most fundamentally, in a multi-district system they are usually only somewhat

related to the actual outcome of interest: the number of electoral divisions won by each party.

In this paper we show that it is possible to use the individual-level responses already collected by political opinion polling firms, along with publicly available datasets and accessible quantitative methods to produce estimates of the number of electoral divisions won by each party. We first establish a framework for evaluating a forecasting model for multi-district elections. It consists of five aspects. These are that a model should: be interpretable and provide reasonable estimates of the number of electoral divisions that each party will win; account for preferences and/or turnout; operate in a reproducible and transparent way; produce results that account for the fact that it is a probability model; and finally not be worse than traditional approaches in terms of speed or cost. Within this framework, the method that we implement involves first estimating the first-preference share of each party in each division based on survey data using multi-level regression with post-stratification. We then use train a model to estimate division-level two-party-preferred shares.

Multi-level regression with post-stratification (MRP) (Gelman and Little, 1997) is a popular way to analyse opinion by way of estimating good-quality estimates of subsets of the population. It uses a regression model to consider individual-level survey responses as being determined by various sub-groups and then rebuilds the sample to better match the population. In this way MRP can not only allow a better understanding of opinion, but also allow us to analyse data that may otherwise be unusable and consider outcomes within the context of distributions and uncertainty. MRP has been used in a variety of settings and one of its advantages is that a survey analysed in this way can be biased to a larger extent than normal because the approach should go some way to account for this. For instance, in the US Wang et al. (2015) used data collected from the XBox platform, which tends to be both younger and more male than the US electorate, and were able to draw meaningful insights about the 2012 US Presidential Election.

In this paper we construct a Bayesian MRP model to generate division-level two-party-preferred estimates for the Australian 2019 Federal Election. In our regression model an individual's first-preference is a function of the demographic and geographic sub-groups to which they belong. That first-preferences model is then post-stratified to better reflect the electoral divisions used at the 2019 Australian Federal Election. Finally, these post-stratified first-preference estimates are used as an input to a trained model that accounts for preferences distribution to estimate a two-party-preferred vote share for every electoral division.

We compare the effect of two political opinion datasets as inputs to our model. The first is a dataset from the Life in Australia survey, and the second is a dataset from the Smartvote Australia voter advice application. To be clear, we analyse these political opinion inputs entirely separately, but with the same approach and model. This allows us to compare the effect of these different datasets. In this way our paper is analogous to, albeit the inverse of, Cohn (2016) who gave the same data to a variety of pollsters to see how, given the same data, the different methods affected their estimates.

The central results of our model relate to seat counts expected by party for the 2019 Australian Federal Election. Using the Smartvote Australia dataset we expected the Australian Labor Party to lose 19 seats, win 93 seats, and for there to be 39 that were too close

to call. Using the Life in Australia dataset, we expected the Australian Labor Party to lose 60 seats, win 29 seats, and for there to be 62 that were too close to call. In actual fact, the Australian Labor Party lost 83 seats and won 68 seats.

Our paper contributes in both academic and practical ways. Academically, we establish a framework for evaluating forecasting models of multi-district elections. We then develop a model for forecasting elections in Australia. Our model is electoral-division-based, but it is less influenced by outlying results than, say, electoral-division-specific polling as it shares information between electorates. Our model goes some way to accounting for preferences distribution, but it has enough flexibility to allow for differences between elections. Training our model required considerable data wrangling, including spatial and time series issues and the matching of disparate datasets. Our datasets are publicly available to make it easier for other researchers to improve on our method. Practically, we show a way in which existing political polling could be analysed in a way that is more meaningful. Our evaluation framework and the model that we put together is relevant to any multi-district system, such as Australia, Canada, and the UK, or even systems such as the US Electoral College. Our model does not require any additional information, hence there would be no additional cost imposed to adopt our approach. Furthermore, the results obtained by our approach may help enhance the credibility of polling forecasts, as they better reflect both the realities of multi-district systems and the uncertainty inherent in any survey approach.

2 Framework

Here we discuss the framework that we developed to characterise a useful forecasting model of parliamentary elections and how the model that we sketched out and implemented relates to each aspect of that framework.

Our framework to characterise the usefulness of a forecasting model of multi-district elections has five aspects. Firstly, a model should be interpretable and provide reasonable estimates of the number of electoral divisions that each party will win. Secondly, a model should account for preferences or turnout, depending on the specifics of the election being considered. Thirdly, a model should operate in as reproducible and transparent way as is possible, and would ideally be replicable. Fourthly, a model should produce results that account for the fact that it is a probability model and not rely on the audience for those results to be statistically fluent. And finally, a model should not be worse than traditional approaches in terms of speed or cost.

2.1 Reasonable estimates of electoral division counts

The first aspect of our framework is that a useful forecasting model of multi-district elections provides probabilistic output from an interpretable model that is ex anti consistent with the available evidence and ex post only inaccurate for discernable reasons.

A reasonable model must be interpretable because it is unavoidably partially assessed on its explanatory value, as elections occur only every few years. Dowding and Miller (2019, p. 1,001) distinguish between 'pragmatic prediction' which is what we have engaged in here, and 'scientific prediction', and argue that '[a]ssessing how good [prag-

matic] predictions are is based on how closely the assigned probabilities match what actually happens.' But given how irregular and idiosyncratic each election is, a useful forecasting model would also engage in 'scientific prediction', in which the explanatory value of the model is also important. Although substantial improvements to a useful forecasting model of multi-district elections would occur after each election, an interpretable model allows continuous smaller-scale improvements. Additionally, one reason that political opinion polling occurs is so that campaigns and the media can use it. An uninterpretable model would be limiting in this regard.

Probabilistic outputs mean that uncertainty is an inseparable aspect of the results. This is important when the occurrence of anything other than the central estimate would have significant effects; hence critical in a multi-district elections setting. The shift to probabilistic results has occurred in a variety of fields and scenarios, for instance, the United Nations moved toward probabilistic population projections in recent years (Raftery et al., 2014).

To be consistent with the available, relevant, evidence ex ante, in this context means that in practice a model is likely to be Bayesian. While it does not have to be fully Bayesian, the ability to borrow information across nested sub-groups and geographies, and take advantage of peripheral information, means there are benefits from this approach. If a model was not drawing on all available relevant evidence, then it would be outperformed by one that was.

While no model in this area will ever be entirely accurate ex post, a reasonable model would only be inaccurate for discernable reasons. For instance, perhaps the assumptions around preferences turned out to be inappropriate. A model that was wrong for reasons that cannot be determined is difficult to update in an appropriate way.

If a model does not explicitly output electoral division counts by party then it is being done implicitly, and this would firstly, be unlikely to appropriately propagate uncertainty, and secondly, not be reproducible or transparent, which are concerns touched on in the third aspect of our framework.

2.2 Account for preferences or turnout

The second aspect of our framework for evaluating the reasonableness of a model for forecasting multi-district elections is that is accounts for preferences, or turnout, as appropriate, in a way that propagates uncertainty.

Preferences are crucial in Australian elections, and much effort is spent when conducting political opinion polling to decide how best to distribute them for their estimates. It is easy to do this in a deterministic way, for instance, by assuming that what happened last election happens again, or to assume some other reasonable sounding but similar approach. We argue that a useful forecasting model would account for preferences in a probabilistic way.

The most important reason for this is the propagation of uncertainty. If preferences are not probabilistically modelled in some way, then the resulting forecasts are likely to be too confident in the central estimates. One way to see whether this is likely to be done is whether the estimates of the two-party-preferred shares are more distributed than the first-preferences shares.

The issues around preferences in Australia are relevant to turnout in many other multi-district elections. Turnout is less of an issue in Australia, due to compulsory voting, and many of the studies of actual changes, as opposed to survey responses, are necessarily focused on elections from over a century ago, such as Fowler (2013). Using survey data, Dassonneville et al. (2019) find that around 60 per cent of survey respondents would have voted even in the absence of compulsory voting. Mackerras and McAllister (1999, p. 217) find compulsory voting reduces the Coalition vote by around five per cent, although Jackman (1999) argues it is not this large. In any case, our model is trained on datasets for which compulsory voting has consistently applied.

2.3 *Transparency, reproducibility, and replicability*

The third aspect of a useful forecasting model of multi-district elections is that it is reproducible when, and in as similar manner as possible, it is made public, and that, further, ideally it would be replicable.

Reproducibility refers to the ability of a third-party to repeat a process and get essentially the same outcome. This is the minimum that is required of credible modern quantitative approaches, and in its absence credibility requires 'trust in the confidence or authority of the originator' (Munafò et al., 2017, p. 5). Given how easy it is for the incentives of those forecasting multi-district elections to be questioned we contend reproducibility also ought to be the minimum required of a useful forecasting model for multi-district elections. But given general practise, we would suggest that greater transparency is needed before reproducibility is possible. To a certain extent, recommendations from industry associations such as the American Association for Public Opinion Research (AAORP) or, in Australia, the Association of Market and Social Research Organisations (AMSRO) are also concerned with this.

Jamieson et al. (2019, p. 19,233) describe transparency '[a]s a prerequisite to scrutiny' and argue that it allows 'others to examine study design, execution, and data and, as a result, replicate results.' However, we go further than Jamieson et al. (2019) and argue that a useful forecasting model would not just require '[a]rchiving data and analysis plans in publicly available repositories', but would also impose a timing and access constraint. Specifically, the data and analysis of a useful forecasting model of multi-district elections must be available at the same time as any public release relating to the forecasting model, be that an academic research paper or press coverage, and it must be just as easily accessed. For instance, it is not enough that datasets and analysis be available on request, if the coverage of the forecasting model is in the media; they must be able to be accessed as easily as the media, for instance in a publicly available repository.

2.4 *Appropriate communication of results*

The fourth aspect of a useful forecasting model of multi-district elections is that its central results should be inextricably linked to their distribution, and these results should be communicated in a way that does not rely on an audience being statistically fluent. Although political opinion polls regularly include some measure of confidence, this tends to be separable from the central estimate, and in practice, regularly is (Larsen and Fazekas,

2019). While this is understandable given the constraints faced by journalists (Ricketson et al., 2019) and the difficulty that even experts have appreciating the nuances of statistical concepts (Hoekstra et al., 2018; Greenland et al., 2016), a useful model would account not allow this to happen.

In another context Vehtari et al. (2019, p. 22) argue that '[t]he most pernicious idea in statistics is the idea that we can produce a single-number summary of any data set and this will be enough to make a decision.' And yet, this is essentially what much of political opinion polling does. Even when the confidence intervals are reported, the results are usually able to be read without these. We contend that in a useful forecasting model the central estimates would be inseparable from their distributional aspects.

Further, as poll-of-poll models increased in popularity, analysts increasingly began to report overall estimates in terms of probabilities. But firstly, in an Australian context, the overall result of an election is essentially a Bernoulli random variable, and so claiming success when a party wins where the model gave that party a probability of 0.2 is disingenuous. Even if this wasn't the case, there is considerable evidence that it is extremely difficult to correctly interpret probabilities (Handmer and Proudley, 2007), and that this can have an effect on electoral behaviour (Westwood et al., 2019).

Finally, the results of a useful forecasting model should not require statistical fluency to understand. Misunderstandings on the part of consumers of estimates are, to a large extent, the responsibility of those who produce estimates.

2.5 Relevancy

Our fifth and final aspect of a useful forecasting model of multi-district elections relates to its relevancy. In order to be useful the model should not be worse than traditional approaches in terms of speed or cost. We see this aspect as critical if models of the type that we introduce here are to be more widely adopted.

3 Data

Our main data requirement is individual-level political opinion polling data that elicits an individual's first-preference, along with demographic and geographic information about that individual such as their electoral division. We use datasets from the Life in Australia survey and the Smartvote Australia voter advice application. In order to post-stratify our estimates we require nested electoral-division-level data for those same demographic and geographic variables. We obtain those data from the 2016 Australian Census. Finally, in order to train the preferences model we require data from past elections on preference flows, which we obtain from the Australian Electoral Commission. We detail these datasets in this section. Our data wrangling was completed in the statistical language R (R Core Team, 2019b), using the broom (Robinson and Hayes, 2019), haven (Wickham and Miller, 2019), foreign (R Core Team, 2019a), rgdal (Bivand et al., 2019), sp (Pebesma and Bivand, 2005; Bivand et al., 2013), and tidyverse (Wickham, 2017) packages.

3.1 Preferences data

The Australian Federal Parliament is divided into two chambers and in this paper we focus on the 'House of Representatives', which is the lower house and from where government is formed. In 2019 for the purposes of lower house representation Australia was divided into 151 electoral divisions each of which with one representative. The boundaries of the electoral divisions are determined by an independent commission that is meant to ensure that each electoral division contains roughly the same amount of people, given various historical constraints such as not crossing state boundaries, and reasonableness. For the 2019 election the number of electors in each electoral division ranged from 69,332 in Solomon which is in the Northern Territory, to 124,507 in Cowper which is in New South Wales. The median was 109,430 electors, while the mean was 108,770 electors, and the standard deviation was 9,686. The level of gerrymandering in Australia is generally considered to be low (i Coma and Lago, 2018). Although there have been infamous episodes historically, these tended to be at a state level, rather than the federal level which is the focus of this paper.

Figure 1: Example completed ballot paper for the House of Representatives.

There are typically five to ten candidates in each electoral division, and voters express preferences over candidates (see Figure 1 for an example of a completed ballot paper). A voter's most-preferred candidate is their 'first-preference', and they rank every candidate

through to their least-preferred candidate. The initial step to decide the electoral division's representative is to count first-preferences. The candidate with the least number of first-preference votes is eliminated, and the second-preferences of the voters whose first-preference was eliminated are distributed to the remaining candidates. This preferences distribution continues until there are only two candidates. The vote proportions at this point are 'two-candidate-preferred' proportions, and the winner is the candidate with a two-candidate-preferred percentage that is at least 50 per cent.[1]

Typical political opinion polling data in Australia focuses on a respondent's first-preference. As such, meaningful political opinion polling forecasts require converting these into two-party-preferred proportions for each of the 151 electoral divisions. One option would be to assume that what happened in the most recent relevant election, in our case 2016, happens again. This approach is easy to understand and implement, but it comes with significant weaknesses. For instance, the conditions that were present in 2016 may not be present in 2019. Additionally, using this approach would require manual adjustment for many electoral divisions as the minor parties that stand in any particular election tend to be different.

Our preferred approach is to account for this using a trained regression model of the preference flows from first-preferences through to two-party-preferred, for each electoral division, from the past four federal elections: 2007, 2010, 2013, and 2016.[2] For each electoral division and election, this model, detailed in the next section, relates the Australian Labor Party's two-party-preferred proportion to the Australian Labor Party's first-preferences proportion. We train this model on a dataset obtained from the Australian Electoral Commission (AEC).

The main difficulty with training this model is that the boundaries of the electoral divisions can change between elections. This means that, for instance, as shown in Figure 2, constructed using Kahle and Wickham (2013), the Division of Canberra as it was in the 2016 election is not geographically identical to the Division of Canberra as it was for the 2019 election. In order to account for this we obtain both first preferences and two-party-preferred historical counts on a voting-booth basis. Using the latitude and longitude of each booth, and maps of the electoral divisions as they were at the 2019 election, we assign each booth to the appropriate 2019 electoral division. This allows us to construct a geographically consistent dataset that we use to train this level of our model.

In Australia, the Australian Electoral Commission sends teams to prisons, hospitals, and other remote locations in an effort to provide an opportunity to vote. As part of our process we will not be counting these votes. The exact number that we remove, each year,

---

[1]The winner's 'margin' is the extent to which their two-candidate-preferred percentage is over 50 per cent. The main two parties in Australia are the Australian Labor Party (ALP) and the Coalition (made up of the Liberal and National parties). In contrast to the two-candidate-preferred measure, the two-party-preferred measure focuses on these two parties. If the final two parties are not the Australian Labor Party and a Coalition party then the two-candidate-preferred and two-party-preferred measures will differ.

[2]Data on preferences are available since preferential voting was introduced in 1918, for instance via the AustralianElections R package (Alexander, 2019c). Nonetheless we train the model on four elections' worth of data. This is a balance between over-fitting by using too few elections, and ensuring the relevance of the preferences that we are training on. Additionally, because of the need to convert the results into the electoral divisions used in the 2019 Federal Election every additional election that we train the preferences model on would introduce more error.
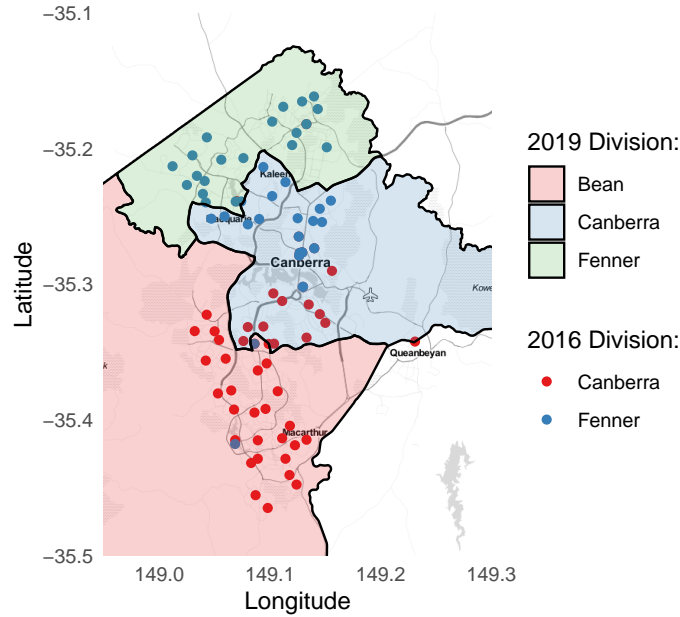
Figure 2: Electoral division boundaries for the Australian Capital Territory in 2019, compared with 2016.

Table 1: Number of first preference votes dropped when the electoral divisions are made consistent

Year	Hospital	Remote	Other/Prison
2007	56,964	20,430	205
2010	53,983	18,727	445
2013	85,859	21,598	616
2016	88,863	21,636	875

is summarised in Table 1.

The three main parties over the four elections, 2007 to 2016 inclusive, are the Australian Labor Party, the Liberal and National Parties, and the Greens, but there are many smaller parties and pivotal independents. Figure 3 shows first-preference votes on a 2019-electoral-division-basis for these four party groups. To be clear, these results are after we convert the electoral divisions used for, say, the 2007 election, into those used for the 2019 election. The specifics of the results in Figure 3 will be similar to, but different from, the usual results for those elections.

Over the elections from 2007 to 2016 the median share of first-preferences for the Greens was between six and ten per cent to nine per cent. However, in 2004 their maximum share of first-preferences was 20 per cent, and that maximum increased each election to be competitive in 2010 when they first won a seat in the lower house, and had reached 43 per cent by 2016.

There has consistently been a few independent and other parties that receive a competitive share of first-preferences. This has included Tony Windsor, and Bob Katter in
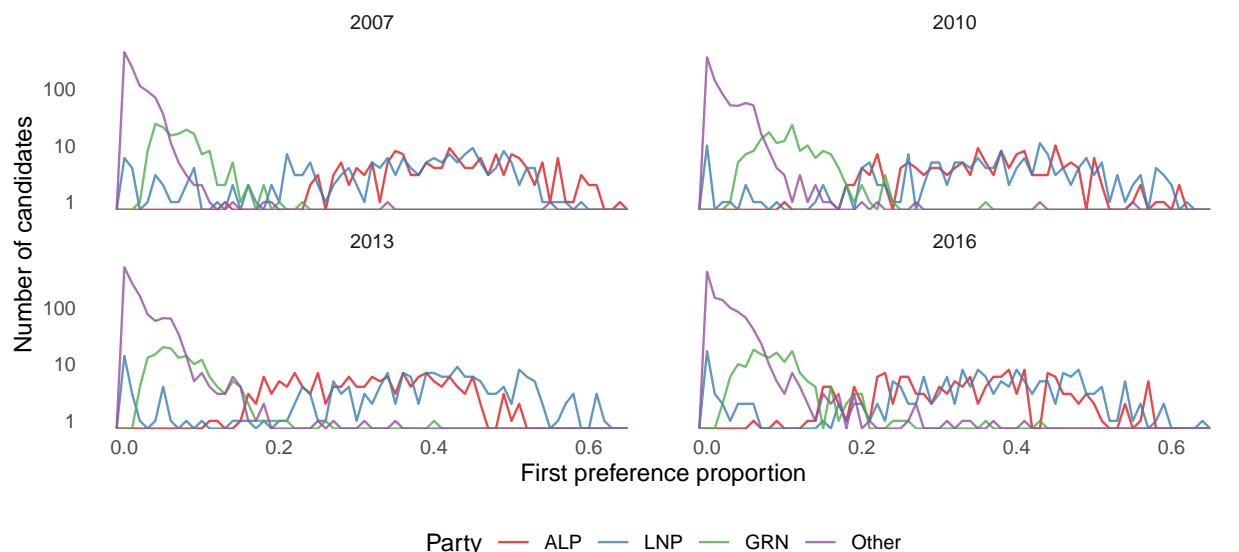
Figure 3: Distribution of first-preferences, by election. Note the log scale on the y-axis.

2007; Robert Oakeshott in 2010; Andrew Wilkie, Cathy McGowen and Clive Palmer in 2013; and most recently, Rebekha Sharkie in 2016. But in general, independent and other parties do not get a large share of first-preference votes.

The lowest average percentage of first-preference votes that the Labor party received over this time was 31 per cent in 2013, and the highest was 42 per cent in 2007. Over this time the Liberal party received a higher average in four of the five elections with a low of 39 per cent in 2016 and a high of 43 in 2004. The only election where the Labor party had a higher share of the first-preference votes was 2007 when Kevin Rudd replaced John Howard as prime minister. The largest difference was in 2013 when the Liberal party had an average share that was larger than the average share for the Labor party by almost 11 percentage points. The National party has a competitive median share of first-preference votes at between 29 and 44 per cent. However, they have a large spread, possibly given the nature of the electoral divisions in which they run

The candidate that wins the vote is determined after preferences are allocated. It is the candidate that gets over 50 per cent on a post-preferences basis. Figure 4 shows the two-party-preferred proportions for the Australian Labor Party and the Liberal National Parties, compared with their first-preferences proportion. It is this relationship that informs Equation 3, detailed in the next section.

Preferences tend to flow toward the major parties for the simple reason that the smaller parties and independents are usually eliminated first. Labor tends to be above a linear best-fit line and the Liberals tend to be below it. To be elected requires a post-preferences proportion of more than 0.5, and so the implication is that for a given first-preference proportion, all else being equal, it is more likely that a Labor candidate will be elected than a Liberal candidate.
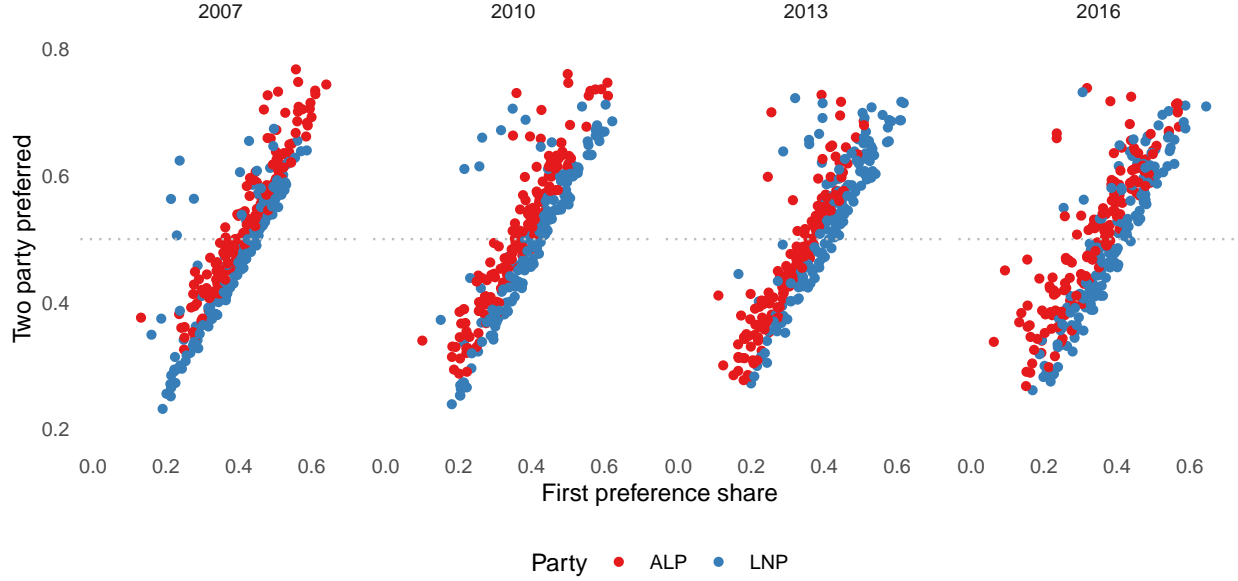
Figure 4: First-preferences compared with two-party-preferred, by election

3.2   Life in Australia Survey

The Life in Australia (LinA) Survey panel is a survey conducted online and by phone by the Social Research Centre. It is a panel survey that roughly matches the Australian adult population. The survey that we use here was conducted between 8 April 2019 and 26 April 2019, with 2,054 people completing the survey, although after cleaning we are left with 1,548 responses. Dove (2019) provides further details about survey specifics.

Table 2 shows the distribution of support by major party grouping. LNP refers to any of the Liberal, National, or Liberal National (in the case of Queensland) parties. ALP is the Australian Labor Party, GRN is the Greens, Other accounts for any other party.

The Life in Australia survey also provides the necessary explanatory variables for an MRP approach (Table 3). Even before re-weighting the Life in Australia survey has

Table 2: First preferences from the Smartvote Australia and Life in Australia datasets, by state

Dataset	Party	NSW	VIC	QLD	SA	WA	TAS	NT	ACT	All
LinA	ALP	152	139	92	44	40	14	3	23	507
	GRN	58	56	37	11	18	5	0	6	191
	LNP	176	131	127	65	88	11	4	5	607
	Other	70	50	51	39	19	9	3	6	247
Smartvote	ALP	9,903	8,657	2,162	677	917	320	80	1,220	23,936
	GRN	5,666	6,297	1,746	511	662	334	57	1,014	16,287
	LNP	5,735	3,999	1,431	327	679	109	33	375	12,688
	Other	2,682	1,938	616	205	284	87	11	238	6,061

11

Table 3: Descriptive statistics of the datasets.

Type	Class	Life in Australia		Smartvote	
		Number	Proportion	Number	Proportion
State	NSW	456	0.29	23,986	0.41
	VIC	376	0.24	20,891	0.35
	QLD	307	0.20	5,955	0.10
	WA	165	0.11	2,847	0.05
	SA	159	0.10	2,542	0.04
	ACT	40	0.03	1,720	0.03
	TAS	39	0.03	850	0.01
	NT	10	0.01	181	0.00
Gender	Female	805	0.52	19,344	0.33
	Male	743	0.48	39,628	0.67
Age group	18 to 29	473	0.30	11,232	0.19
	30 to 44	451	0.29	20,261	0.34
	45 to 59	513	0.33	16,595	0.28
	60+	115	0.07	10,884	0.18
Education	High school	584	0.38	NA	NA
	Grad. dip.	371	0.24	NA	NA
	Bachelor	335	0.22	NA	NA
	Postgraduate	232	0.15	NA	NA
	Unknown	30	0.02	NA	NA

been conducted so that the sample should resemble the Australian population. For instance, the distributions of respondents by state and gender are fairly similar to values that are obtained by the census. The survey has slightly over-sampled respondents with a post-graduate degree; and substantially under-sampled respondents in the 60+ age-group which only make-up 8 per cent of the sample, but account for around 27 per cent of the population (ABS, 2018a).

The main issue with the Life in Australia survey for the purposes of forecasting the number of electoral districts won by each party is its size. Although it is of a typical and appropriate size for national-level estimates, even just disaggregating to the states, let alone the electoral division level, highlights the issue in the Australian Capital Territory, Tasmania, and the Northern Territory, with 50, 48, and 15 respondents respectively. For the purposes of forecasting election outcomes in Australian federal elections with a Bayesian MRP model a small sample in the Australian Capital Territory and the Northern Territory is not overly problematic. There are only two electoral divisions in the Northern Territory – Solomon which is focused on Darwin, and Lingiari which accounts for the rest of the territory – and they are fairly stable. Similarly, while there are three electoral divisions in the Australian Capital Territory – Bean, Canberra, and Fenner – they tend to be reliably Australian Labor Party. The main issue is Tasmania. As one of the six constituent states of Australia, Tasmania is guaranteed five electoral divisions in the lower

house (despite having a similar number of voters, the Australian Capital Territory, as it is a 'territory' only has three). Furthermore, the politicians representing these electoral divisions regularly change. For instance, in 2016, the electoral divisions of Bass, Braddon, and Lyons unexpectedly changed. Then in 2019 they changed back. This weakness motivates the use of a larger, but more-biased, sample. It will also motivate the use of a Bayesian model.

3.3 Smartvote Australia

The Smartvote Australia voter advice application is a survey that attempts to elucidate the policy preferences of a voter and then advise them which party best matches these preferences (Dumond, 2019). It is a biased, self-selected, sample.

The data from voter advice applications have been used in the context of MRP, including by Popp et al. (2016). While the use of voter advice applications has been found to increase participation in elections (Germann and Gemenis, 2019), compulsory voting in Australia should negate much of that effect. There are some concerns around endogeneity as it has been found that voter advice applications may cause a small number of voters to change the party they vote for based on the advice of the application (Kleinnijenhuis et al., 2019). There are also concerns around the nature of the questions asked. Specifically, Smartvote Australia asked respondents to rank how likely they were to vote for a particular party. To a certain extent, therefore, this is a utility measure rather than a singular choice. Additionally, there are some respondents who have all of the parties ranked equally. van der Eijk et al. (2006) expands on some of these issues.

After cleaning and adjustment we have 58,972 responses.

Table 2 shows the distribution of support by major party grouping. There is a consistent bias toward the Australian Labor Party.

Smartvote Australia provides the necessary explanatory variables for an MRP model (Table 3). However, in contrast to the Life in Australia sample, the Smartvote Australia sample is quite biased. For instance, New South Wales and Victoria account for 41 and 35 per cent of the sample, even though they only make up 32 and 26 per cent of the Australian population, respectively. Females are also considerably over-sampled.

Notwithstanding the biases in the Smartvote Australia sample, the main benefit of it is its larger size. This provides substantial benefits to our model given that it needs to generate estimates for all 151 electoral divisions.

3.4 Post-stratification data

The data requirements of an MRP model can be onerous. For every level of every geographic and demographic feature that is to be used to characterise a survey respondent, we need to know the relative proportion of every combination in the population. For instance, if we are interested in results on some geographic basis, while age is being grouped into four categories (18-29, 30-44, 45-59, and 60+) and gender is being grouped into two categories (male and female), then we need to know the proportion of males aged 18-29, the proportion of males aged 30-44, etc, for a total of eight sub-groups, for each of the geographic areas. Additional features compound, so if we were to also char-

acterise by highest level of education completed, say grouped into four categories (didn't complete high school; high school; undergraduate; post-graduate), then the sub-groups would be the proportion of males aged 18-29 who didn't complete high school, etc, for a total for 32 sub-groups for every geographic area.
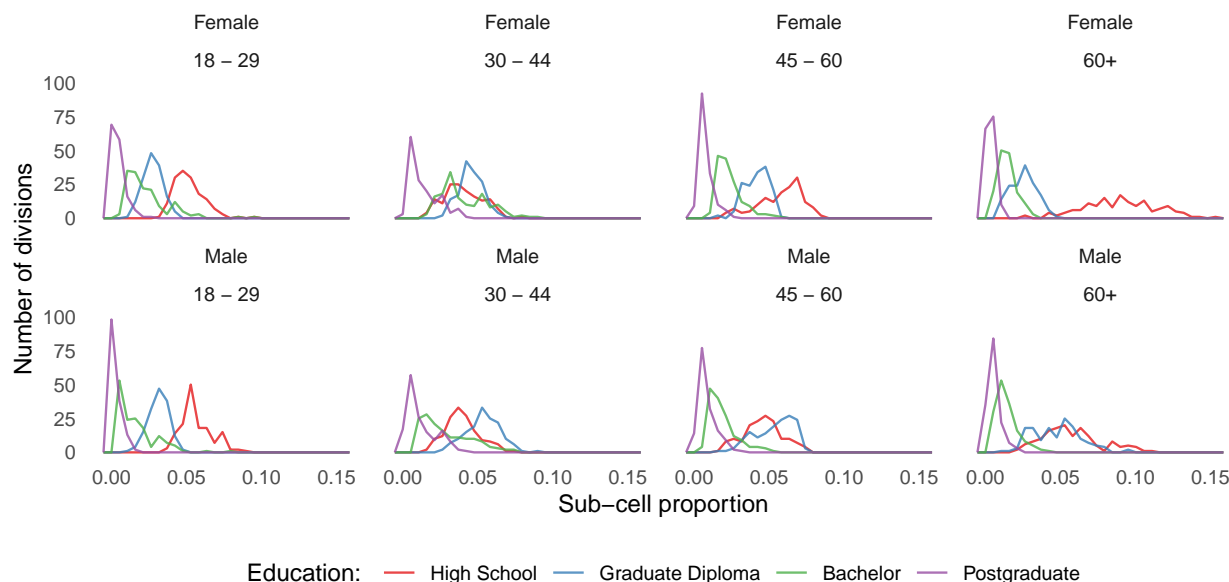

Figure 5: Density of electoral division sub-group proportions, by age-group and gender

We use data from the Australian Bureau of Statistics 2016 Census accessed using Table-builder (ABS, 2016). This provides us with the nested sub-group counts that we require on the basis of the electoral divisions used in the 2019 Federal Election, after we upload custom geographies from ABS (2018b). If it were not possible to use custom geographies then a statistical re-weighting approach such as that of Forbes et al. (2019) could be used to construct this correspondence.

There are 151 electoral divisions and the number of people aged at least 18 ranges from 64,472 in the electoral division of Lingiari, Northern Territory, to 153,033 in the electoral division of Sydney, New South Wales. Note the difference in this dataset, compared with the actual number of electors. These are divided into 32 sub-groups, and the composition of each electoral division tends to change across electoral divisions.

The extent of the change across electoral divisions can be seen in Figure 5, which show the density plots. The major difference is in the 60+ age-group, the proportion of which varies considerably across the electoral divisions.

4 Model

We use a Bayesian MRP model that is based on Wang et al. (2015) and Alexander and Alexander (2016).[3] The output that we focus on is the Australian Labor Party's two-

---

[3]To aid understanding and to help encourage the use of our approach Appendix A provides R code for a simplified model in an Australian context. Similar 'starter models' are available for the US (Kastellec et al.,

party-preferred (2PP) share in each of the 151 electoral divisions in the 2019 Australian Federal Election for the House of Representatives. The political opinion polling inputs to our model are the individual-level preferences, along with a respondent's demographic characteristics (gender, age-group, and highest education), and their electoral division. We post-stratify these first-preference share estimates using data from the 2016 Australian Census. Finally, we provide estimates of the Australian Labor Party's two-party-preferred share for each electoral division on the basis of these post-stratified first-preference share estimates, using a model trained on data from the past four elections.

Our model focuses on whether the Australian Labor Party is likely to win a particular electoral division, but differs from many other electoral-division-specific-polls through the use of MRP. Our model is also quite different to poll-of-polls approaches such as Jackman (2005) which have become popular. Even though Australian elections are decided on the basis of seat counts, many poll-of-polls models tend to focus on providing Australia-wide first-preference or two-party-preferred estimates.

In our model, the electoral divisions used at the 2019 Federal Election are denoted by $d \in [1, 2, \ldots, 151]$; individual survey respondents 'individuals' are denoted by $i \in [1, 2, \ldots, I]$; political parties are denoted by $p \in [1, 2, 3, 4]$, where party 1 is the Australian Labor Party, party 2 comprises the Liberal and National parties, party 3 is the Greens, and party 4 comprises all other parties and independents, so that $P = 4$; and the demographic and other 'characteristics' are denoted by $c \in [1, 2, 3]$, where characteristic 1 is gender, characteristic 2 is age-group, and characteristic 3 is highest education obtained.

The combination of the levels of the characteristics characterise the sub-groups that are used for post-stratification. In this case, there are two categories for gender, four categories for age-group, and four categories for highest education obtained, so there are thirty-two sub-groups, denoted by $c \in [1, 2, \ldots, C]$, where $C = 32$. The number of people in each sub-group for each electoral division is similarly denoted N_c. We discard informal votes. In this set-up, $2PP_{d,p=1}$ is the Australian Labor Party's two-party-preferred share in electoral division d.

We will now discuss each component of our model, starting with individual responses and building up toward electoral-division-level estimates of two-party-preferred.

Equation 1 is at an individual-respondent level:

$$\Pr(\hat{FP}_{i,p=1}) = \text{logit}^{-1}\left(\delta_0 + \delta_1 \text{gender}_i + \delta_2 \text{age}_i + \delta_3 \text{education}_i + \delta_4 \text{division}_{d[i]}\right) \quad (1)$$

It says that the probability that some person i, in electoral division d, will vote for the Australian Labor Party depends on their gender, their age-group, their highest level of education, and their division. Each of the demographic variables has a fixed effect on the intercept and enters separately, while division enters as a random intercept. The important aspect of this estimation is the coefficients and the uncertainty around them. In a sense, all that the survey does is train a model that in the next step is applied to a new dataset that is more geographically- and demographically-correct.

2016), and the UK (Hanretty, 2019).

Equation 2 is the post-stratification equation:

$$\Pr(\hat{\text{FP}}_{d,p=1}^{PS}) = \frac{\left(\sum_{c=1}^{C} N_{c,d} \times \text{logit}^{-1}\left(\hat{\delta}_0 + \hat{\delta}_1\text{gender}_c + \hat{\delta}_2\text{age}_c + \hat{\delta}_3\text{education}_c\right)\right)}{\sum_{c=1}^{C} N_{c,d}} \quad (2)$$

This equation considers each of the geographic- and demographic-characteristics of the individual respondents as compounded to place the individual into a particular sub-group. For instance, a 32-year-old, high-school-educated, male would be placed in the 'male-high-school-aged-30-44' sub-group. The number of sub-groups depends on the number of characteristics that are considered and the number of groups within each sub-group. For instance, for our model gender has two groups: male and female (we drop 'other' and 'non-responses') and age-group has four groups: 18-29, 30-44, 45-60, and 60+ (we drop those <17, 'other', and 'non-responses'); and education has four groups: high school, graduate diploma, bachelor, and post-graduate, leaving us with 32 sub-groups. This equation determines the proportion of the population, not the sample, that each sub-group represents. Many MRP papers introduce a new variable here, often θ. Instead we explicitly include the estimates from Equation 1 in an effort to make the process more clear. The output of this equation is post-stratified estimates for the Australian Labor Party's first-preference share for every electoral division.

Australia has preferential voting which means that those first-preference shares need to be converted into two-party-preferred shares. To do this we train Equation 3 on data from the past four elections (2007, 2010, 2013, and 2016) to take the electoral-division-level estimates for the Australian Labor Party's first-preferences shares to estimate the Australian Labor Party's subsequent two-party-preferred share in each electoral division.

$$2\text{PP}_{d,p=1} = \hat{\beta}_0 + \hat{\beta}_1\hat{\text{FP}}_{d,p=1}^{PS} \quad (3)$$

Essentially these equations are attempting to account for preference-flows.

We propagate uncertainty between these two stages by applying Equation 3 to the posterior distribution from Equation 2, not just the central estimate.

We estimate our model using Stan (Carpenter et al., 2017) via the brms (Bürkner, 2017) and tidybayes (Kay, 2019) packages in the statistical language R (R Core Team, 2019b).

5 Results

5.1 Individual-level model

The first-step of our approach, Equation 1, trains an individual-level model to express the likelihood of voting for the Australian Labor Party as depending on various individual-level characteristics, using data from the surveys. Table 4 contains population-level (as opposed to electoral-division-specific) coefficients estimated by our model for the Life in Australia and Smartvote Australia datasets, and Figure 6 shows the distribution of these estimates.

At this stage of our approach, the model is being trained on the raw datasets, and these results are not yet post-stratified. We have more to say about this in Section 6, but

Table 4: Model coefficients when using each dataset.

| Term	Life in Australia				Smartvote			
	Estimate	Std. Error	Lower	Upper	Estimate	Std. Error	Lower	Upper
Intercept	-0.65	0.14	-0.89	-0.42	-0.60	0.03	-0.65	-0.55
Is male	-0.10	0.11	-0.28	0.09	-0.13	0.02	-0.17	-0.10
Age 30-44	-0.14	0.14	-0.37	0.10	0.02	0.03	-0.02	0.06
Age 45-59	0.03	0.14	-0.19	0.26	0.28	0.03	0.24	0.32
Age 60+	-0.48	0.25	-0.89	-0.09	0.82	0.03	0.78	0.87
High school	-0.01	0.15	-0.25	0.24	NA	NA	NA	NA
Grad dip.	0.11	0.17	-0.16	0.37	NA	NA	NA	NA
Post grad.	0.14	0.18	-0.16	0.43	NA	NA	NA	NA
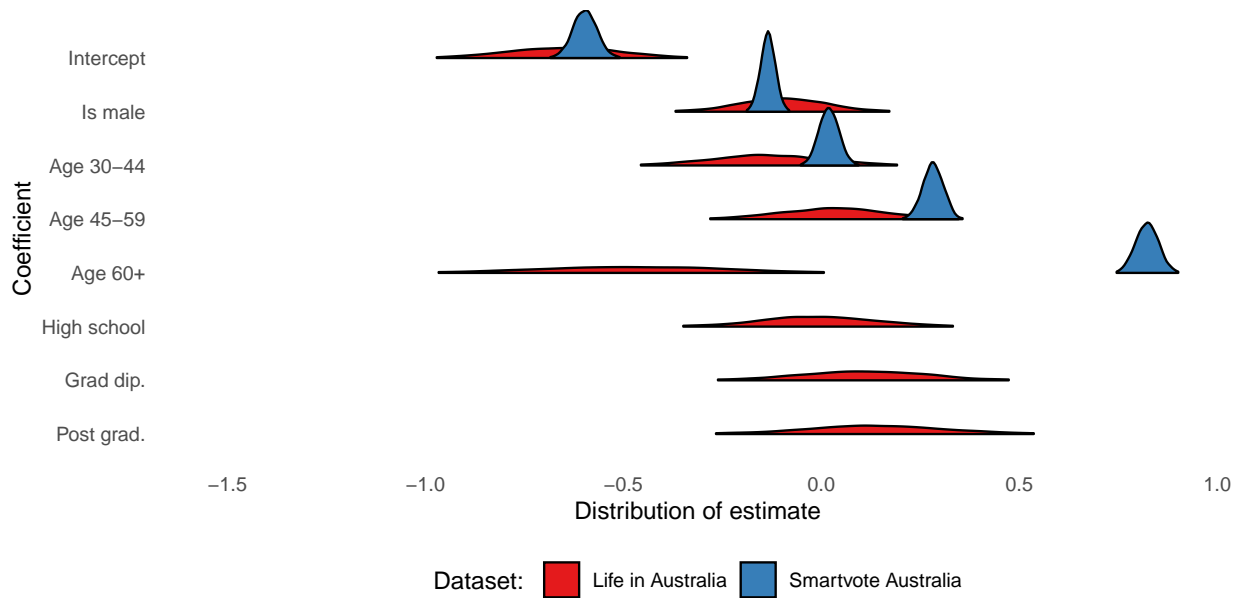
Figure 6: Distribution of coefficient estimates

ours is first and foremost a forecasting rather than explanatory model, nonetheless there is value in examining the results for reasonableness. The estimates are in relation to how likely a respondent is to support the Australian Labor Party. In the case of both datasets, males are expected to be less likely to support the Australian Labor Party. Monotonically, respondents in older age-groups are more likely to support the Australian Labor Party in the Smartvote Australia dataset, but there is no monotonic trend in the Life in Australia dataset.

Our datasets were unrepresentative in various ways. The next step in our approach is to use our trained model with a dataset that is more representative of the voting population.

5.2   With post-stratification

To post-stratify our estimates, Equation 2, we take the models trained on the raw datasets and apply them to a less-biased dataset, albeit one that does not have information about voting intention.

One important aspect of post-stratification is to understand how the raw survey results are being changed. We consider this on the basis of electoral division in Figure 7. The estimate of the first preference share for each electoral division is made by making many forecasts of the outcome, and then examining this distribution. Figure 7, which follows Alexander (2019a), shows the mean of the distribution and 95 per cent prediction intervals (as a reminder, prediction intervals tend to be wider than confidence intervals). These can be compared with the raw data from the survey. In general, it can be seen that the raw estimate of the Australian Labor Party's share of first preferences in both surveys is being reduced by the model. Figure 7 is on the basis of electoral divisions, however it is also possible to do a similar exercise on the basis of, say, age-group, gender, or education.

5.3   Adding preferences

One critical aspect of the Australian voting system is the use of preferences. However, political opinion polling survey results are typically on the basis of first-preferences. In this step of our results we apply a model trained on the past four elections' worth of first-preferences and two-party-preferred results to our post-stratified results. As we are focussed on the Australian Labor Party, in general their two-party-preferred estimate will be higher than their first-preferences estimate for an electoral division, as a majority of Greens votes will flow to the Australian Labor Party.

Figure 8 compares the mean estimates of first-preferences and two-party-preferred shares for each electoral division, with the actual result.

If the central estimate from our model aligned with the actual outcome then the point would lie on the 45 degree line, indicated by the dotted line in Figure 8. Points that are above that line suggest our model over-estimates the Australian Labor Party's share of first preferences and two party-preferred votes, respectively.
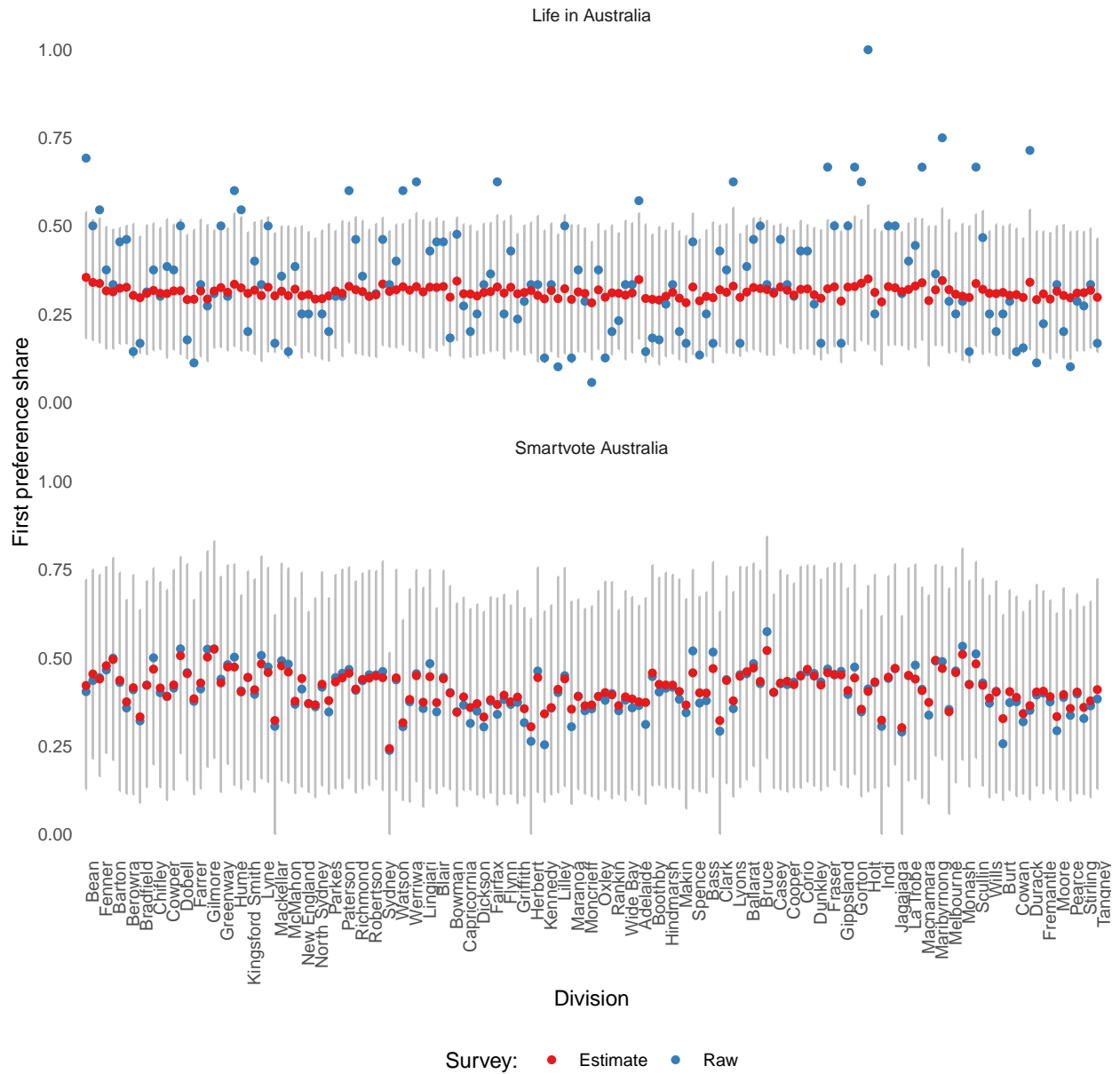
18

Figure 7: Comparison of raw and post-stratified estimates of the Australian Labor Party share of first preferences for each electoral division.
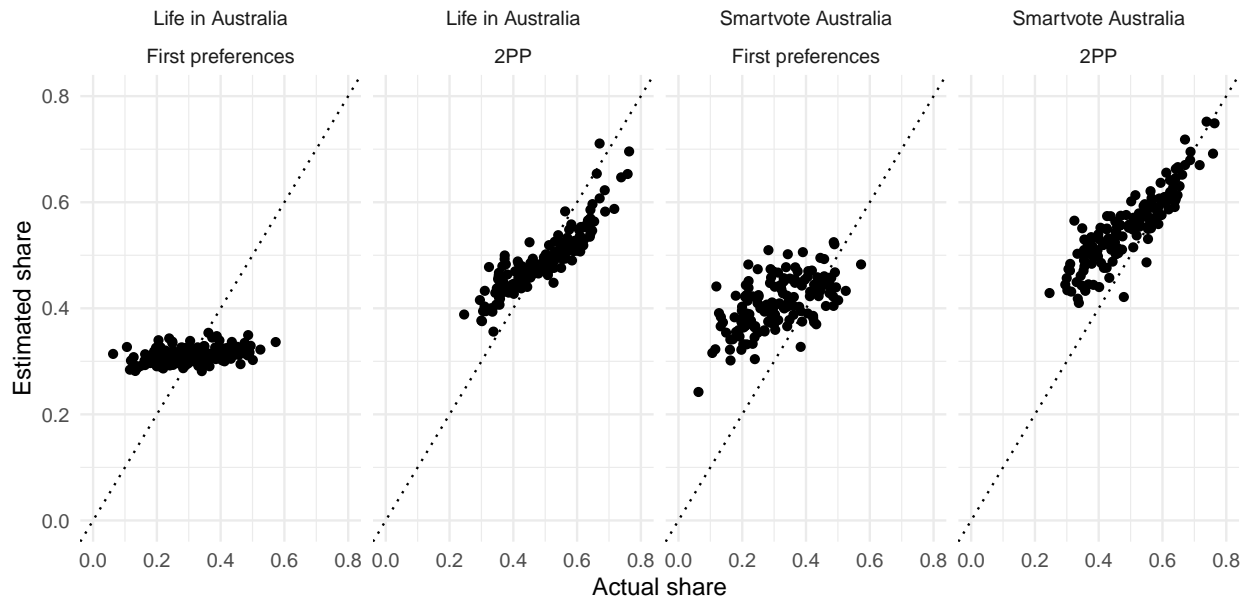
Figure 8: Comparison of actual with estimated Australian Labor Party share of first preference and two-party-preferred votes

5.4 Overall results

Probabilistic forecasts are notoriously difficult to interpret (for instance, see Morss et al. (2010), Morss et al. (2008), Patt and Schrag (2003), or Mislavsky and Gaertig (2019)). However, in Australia there is a commonly used framework for describing how 'safe' an electoral division is. One way to better communicate our results is to translate our central estimates into this framework. Table 5 classifies the two-party-preferred share for both our model estimates and the actual outcomes into commonly used descriptive tranches.

The classifications used by the Australian Electoral Commission are usually used only on actual results, and so we adjust the central classification in a way that is more explicit about uncertainty. Specifically: 'safe loss' means a two-party-preferred share of less than 0.4; 'fairly safe loss' means between 0.4 and 0.44; 'marginal loss' means between 0.44 and 0.47; 'too close to call' means between 0.47 and 0.53; 'marginal win' means between 0.53 and 0.56; 'fairly safe win' means between 0.56 and 0.6; and finally, 'safe win' means a two-party-preferred share greater than 0.6. Table 6 shows this disaggregation.

Using this classification, it can be seen that estimates based on the Life in Australia dataset tend to be more accurate than those based on the Smartvote Australia dataset.

6 Discussion

6.1 Evaluation within the context of the framework

In terms of providing electoral division counts, our results highlight the benefits of combining raw individual-level political opinion polling data, with large public datasets and a quantitative model. Our model is able to provide forecasts that are directly related to the

Table 5: Both outcomes, by broad classification

Classification	Share	Actual	LinA-based	SmartVote-based
Safe loss	<0.4	42	9	NA
Fairly safe loss	0.4-0.44	18	20	9
Marginal loss	0.44-0.47	15	31	10
Too close to call	0.47-0.53	19	62	39
Marginal win	0.53-0.57	12	14	25
Fairly safe win	0.57-0.6	17	8	33
Safe win	>0.6	28	7	35

Table 6: Outcomes, by broad classification

Estimate	Actual						
	Safe loss	Fairly safe loss	Marginal loss	Too close to call	Marginal win	Fairly safe win	Safe win
Smartvote Australia							
Fairly safe loss	7	0	0	1	0	0	0
Marginal loss	7	2	0	0	0	0	0
Too close to call	24	9	7	2	1	0	0
Marginal win	3	5	6	8	3	3	0
Fairly safe win	1	2	2	6	8	9	3
Safe win	0	0	0	2	0	5	25
Life in Australia							
Safe loss	8	0	0	0	0	0	0
Fairly safe loss	18	3	0	0	0	0	0
Marginal loss	11	9	7	4	0	0	0
Too close to call	5	6	8	15	11	13	3
Marginal win	0	0	0	0	1	3	11
Fairly safe win	0	0	0	0	0	1	7
Safe win	0	0	0	0	0	0	7

actual outcome of interest in elections for multi-district systems: the number of electoral divisions won by each party. Furthermore, similar approaches to ours have also found that they were able to obtain reasonable electoral-division-level results. For instance, in Australia, Jackman et al. (2019) use a dataset of opinions about same-sex marriage from a 2016 voter advice application with an MRP approach to construct estimates of support in the, then, 150 Australian electoral divisions, and compare their results favourably to the 2017 plebiscite on same-sex marriage. Of more relevance in terms of trying to forecasting election outcomes is Lauderdale et al. (2017) who apply a similar model to ours to the UK and the US; and Munzert (2017) who analyse constituency level results in Germany. Our model is interpretable and provides probabilistic outputs. However our model does not draw on all available evidence, as it is deliberately parsimonious.

In our model we account for preferences in a way that propagates uncertainty by applying a model that relates first-preferences shares to two-party-preferred shares which has been trained on the past four elections. This results in increased variance, compared with a deterministic approach.

The datasets that we use required a large amount of preparation, cleaning, and wrangling before they were in a state that we were able to use them. Nonetheless, we have made all of them publicly available. The quantitative nature of our approach means that it is transparent and reproducible because bias adjustments, of the sort that may be usually done under the table, is an inherent part of our approach.

Replicability is a more onerous requirement than reproducibility, and that is for a third-party to be able to carry 'out a new study based on the description of the data and method provided in the original publication, and obtaining results that are similar enough' (Alexander, 2019b). To this end, a third-party with individual-level responses could use the datasets and models that we have made publicly available to analyse their own dataset. However, we argue additionally, that replicability has different requirements depending on the setting. Given the use case of political opinion polling we do not believe that a reasonable level of reproducibility is achieved by, say, solely making the code and datasets available on GitHub. There are many users of political opinion polling who would not have the specific technical background required to be able to replicate our approach on their own dataset. To this end, we have additionally, we have constructed a Shiny app that allows people to upload their own political opinion polling data and immediately see our model's forecast. This is based on Alkema and New (2015), who implement a web application to provide country-specific estimates of contraceptive prevalence and unmet need for family planning, broadening the replicability of a model that otherwise requires extensive experience and training to use. Our Shiny app could be easily inputted with the results of political opinion polling to generate meaningful results. While there could be some issues in terms of matching the exact groupings that we have used, they are reasonably broad and so many existing surveys would be able to be re-aggregated toward it if necessary.

The approach that we implemented addresses this aspect by adopting the commonly used classifier of how safe an electoral division is. The advantage of this approach is that there is a natural 'too close to call' category that is inseparable from the rest of the results.

For reasons that are not clear, much of the political opinion polling literature and practice is focused on trying to forecast results that are only correlated with who wins the elec-

tion rather than who wins the election. For instance, forecasting the popular vote in the US, rather than the Electoral College, or nation-wide preferences in Australia, Canada, and the UK, rather than seat counts. Our results show that there is no need to do this and that widespread adoption of political opinion polling that focused on seat counts would bring substantial improvements. Nonetheless the reality of the situation is that an approach that was either considerably slower or more costly than the traditional alternatives would not be widely accepted.

Rothschild (2015), Goel et al. (2015) and Konitzer and Rothschild (2016) are similar to us in terms of approach, motivation, and philosophy. They look at political opinion polling in a US context and find significant room for improvement. Significantly they argue that an approach that is similar to ours is neither more costly, nor slower.

Our use of MRP allows us to conduct meaningful analysis of incomplete and non-representative data. This is in comparison with traditional political opinion polling where there are greater constraints on the samples. This serves to reduce the cost of collecting survey responses. Buttice and Highton (2013) examine the outputs of MRP models given surveys and simulations with different characteristics. They find that it is important to optimise the model toward the question of interest, and to increase the sample size when possible.

The Bayesian aspect of our approach allows us to take advantage of data that are available, but may be too small for classical methods. This is because of partial pooling. Specifically, when there are small sub-groups then these tend toward the overall average, rather than needing to be thrown out. For instance, in a survey of 1,500, given 151 electoral divisions, we would only expect around 10 respondents per electoral division, and so given the number of sub-groups that we consider, it is inevitable that many will be empty. In this cases, the model tends toward the average.

Additionally, our use of a model-based pipeline means that once the model has been used for one sample, it is almost costless to use it for another, similar, sample, and doing this takes very little time. Indeed our Shiny app produces estimates within minutes.

6.2   Comparison of survey inputs and the appropriateness of MRP

Our paper is also about examining how MRP performs given surveys with different characteristics. Comparing the results of our model when we consider two different inputs. The first was a small sample designed to roughly match the Australian adult population, and the second was a self-selected voter advice application. The results of our model suggest that a larger sample than is typically used would be beneficial. In this way, our results are similar to Buttice and Highton (2013) who perform this exercise in a more focused manner.

One interesting aspect of our approach is the consideration of preferences. Preferential, but compulsory, voting is one way in which the Australian system is different to that of the Canada, the UK and the US. It removes considerable complexity to just consider the first-preferences results. Similarly, many surveys only have information on the state of the respondent, instead of their electoral division. Considering these two variants is worthwhile as they expand the potential use cases of the approach. They are implemented in Appendix B.

6.3 Limitations

There are a variety of limitations in our approach. Here we consider them in terms of data limitations, methodological weaknesses, and communication failings.

In terms of the data that we use, one of the most notable issues with our post-stratification dataset is that it comes from the census. While we adjust this population to only include individuals that could be a voter, this is not the same as the population of those who actually voted, or even those who are enrolled to vote. There is evidence that the difference between the census and those who actually voted is systematic. For instance Hannan-Morrow and Roden (2014) finds that age and gender have an effect on turnout; specifically, being younger or male implies a person is slightly less likely to vote. Similarly, Sheppard et al. (2019) finds differences, in terms of political participation, between migrants and native-born in Australia.

Another data-related issue is that we do not distinguish between those who vote early (that is before election day) and those who vote on election day. The rate of early voting is increasing in Australia, and there is evidence that older voters are more likely to vote early, and may be more motivated 'against' the leader of the party that they do not support (Sheppard and Beauregard, 2018). If there were a compelling reason, it would be possible to complicate the model to account for early voting, however other complications, discussed below, are likely to be more beneficial.

From a methodological perspective, our model is anchored to past results. Given the nature of the Australian electorate, where few voters change their mind, this is generally a feature. However, it does mean that our model struggles with out-of-the-blue changes. For instance, the electoral division of Warringah in the Northern Beaches of Sydney has been reliably conservative, for instance most recently being the division of Tony Abbott, a former conservative prime minister. But in 2019, the independent Zali Steggall harnessed anti-Abbott sentiment and ended up winning by seven percentage points. It would have required considerable survey data from Warringah for our model to forecast an Abbott loss.

Our model is set-up on the assumption that the parliament is dominated by two larger parties. If coalition-formation were an important aspect to the multi-district election being forecast, then there would be value in adding an additional layer to the model, along the lines of Stoetzer and Orlowski (2019).

Finally, in terms of communication, we have provided some variations designed to ensure that the probability nature of our model cannot be separated from central estimates. However, this is just a first step and there is much more that could be done, including survey work and field testing to examine how this could be done in a better way.

6.4 Next steps

While for the purposes of this paper we wanted a parsimonious model, adding several complicating factors would be beneficial. Firstly, there is room to add another two or three co-variates in the individual-level model, especially if these co-variates were able to speak to the sample's bias. For instance, place of birth—specifically distinguishing between whether a person was born in Australia—is suggested by Leigh (2005), and re-

ligion is suggested by Jackman et al. (2019). The constraint here is that the covariant needs to be available to be computed as a nested sub-group and so needs to be present in the census or another post-stratification dataset. Secondly, adding another layer, say, to inform the model of the public results from traditional political opinion polling would be beneficial. In Germany and New Zealand settings Stoetzer et al. (2019) sketch a potentially fruitful path forward in this regard. And finally, adding electoral-division- and state-specific complications would be beneficial in terms of better using related information that is available, and the electoral-division-specific political opinion polling that can be produced. In a UK setting Hanretty et al. (2016) show one way in which this could be accomplished. As an example of this, while the level of aggregation that we use in this paper is that which is commonly used for national-level political opinion polling, it could be improved by accounting for state-specific parties. For instance, while it is appropriate to group the Jacki Lambie Party or Katter's Australia Party, with 'Other' at a national level, they attract considerable support in Tasmania and Queensland, respectively.

One weakness of our approach is that it is not electoral-division-specific, which means that there are nuances of specific electoral divisions that it misses. Better preferences model - take account of preferences deals, and also not lump all non-major parties in as other - separate them based on ideology so that there is at least some better flow.

In this paper we have used Australia-wide surveys as our input. However, one interesting aspect of our approach is that for the purposes of forecasting Australian elections this is a misallocation of resources, as, ceteris paribus, there is little need to survey safe electoral divisions. Our results suggest that combining our model with a few surveys, each just of a usual size of say 1,500 to 2,000, conducted over the course of a few weeks, but between them targeted at the seats that are 'too close to call', 'marginal wins' or 'marginal losses', could be an efficient use of resources, that is yield valuable results without considerable expenditure.

Our results also point to the need for better quality, transparent, survey data in Australia. In the same way that the Australian Bureau of Statistics provides GDP estimates for anyone to use, there is a clear need for political surveys in the range of 1,000 to 10,000 respondents. Importantly the raw data from these surveys needs to be publicly released, so that anyone is able to analyse it, and from which a pool of comparable political opinion polling analysis could be conducted. These datasets need to be released in close to real-time and be freely available, unlike much of the research conducted presently, even that paid for by government grants.

Many of the current forecasting models of multi-district elections, especially in Australia, fall short when characterised by our framework. Political opinion polling is inextricably linked to modern politics, and the way in which it is done leaves much to be desired. While leaving plenty of room for further development, here we have sketched out and implemented an approach that points to a way forward, and we look forward to others building on and improving it.

A Example MRP model

This appendix provides a small worked example with R code so that readers can more easily get started with multi-level regression with post-stratification themselves. It is similar in spirit to (Kastellec et al., 2016) and (Hanretty, 2019).

A.1 Set up the workspace

First load the packages.

```r
library(brms) # Needed for the Bayesian section
library(tidyverse)
```

Then generate some sample polling data to analyse. The advantage of this is that we have some idea of what to expect from the model. We'll just use two variables here:

- is_male, which is 0 for females and 1 for males; and
- in_age_group, which is one of four groups: 1, 2, 3, 4.

```r
example_poll <-
  tibble(
    is_male = sample(c(0, 1),
                     5000,
                     replace = TRUE,
                     prob = c(0.75, 0.25)
                     ), # deliberately over-sample females
    in_age_group = sample(c(1:4), 5000, replace = TRUE)
  )
```

We'll make males and older people less likely to vote for the Australian Labor Party in our sample; and females and younger people more likely to vote for the Australian Labor Party.

```r
example_poll <- example_poll %>%
  mutate(
    binomial_param = case_when(
      is_male == 0 & in_age_group ==  1 ~ 0.9,
      is_male == 0 & in_age_group ==  2 ~ 0.7,
      is_male == 0 & in_age_group ==  3 ~ 0.5,
      is_male == 0 & in_age_group ==  4 ~ 0.3,
      is_male == 1 & in_age_group ==  1 ~ 0.7,
      is_male == 1 & in_age_group ==  2 ~ 0.5,
      is_male == 1 & in_age_group ==  3 ~ 0.3,
      is_male == 1 & in_age_group ==  4 ~ 0.1,
      TRUE ~ 0.5
```

```
  )
) %>%
rowwise() %>%
mutate(supports_ALP =
          rbinom(n = 1, size = 1, prob = binomial_param))
```

Now we'd like to see if we can get our results back (we should find females less likely than males to vote for Australian Labor Party and that people are less likely to vote Australian Labor Party as they get older). Our model is:

$$\text{ALP support}_j = \text{gender}_j + \text{age-group}_j.$$

This model says that the probability that some person, j, will vote for the Australian Labor Party depends on their gender and their age-group. Based on our simulated data, we would like older age-groups to be less likely to vote for the Australian Labor Party and for males to be less likely to vote for the Australian Labor Party.

```
#### Individual-level model ####
model <- brm(ALP_supporter ~ gender + age_group + education + (1|division),
              data = LINA_regression_data,
              family = bernoulli(),
              cores = number_of_cores,
              file = "outputs/models/LinA" # If running this again, either
              # need to change this file name to something new or delete
              # the saved model from the folder.
              )
pairs(model)
ranef(model)
summary(model)
```

The results can be a little unintuitive, but essentially we've got our inputs back. In practice, it's usually fine to start with an OLS model and then iterate toward an approach that may be more appropriate, if that is easier for you.

Now we'd like to see if we can use what we found in the poll to get an estimate for each of the seats based on their deomgraphic features.

There are packages that make cleaned Australian Bureau of Statistics datasets available, but it can be worthwhile going through the process yourself at least once.

First read in some real demographic data, on a seat basis, from the ABS.

```
females <- read_csv("inputs/data/census/table-female.csv",
                    skip = 8,
                    col_names = FALSE)
males <- read_csv("inputs/data/census/table-male.csv",
                  skip = 8,
                  col_names = FALSE)
```

```
both <- cbind(females, males)
rm(females, males)
```

Now we need to do some minor tidying and aggregating: clean the names; fill the rows; and construct the groups and proportions that we need.

```
# First transpose it - use this rather than t() so that it's still a tibble
both <- as_tibble(cbind(nms = names(both), t(both)), .name_repair = "minimal")
names(both) <- both[1,]
both <- both[2:nrow(both),]

# Here we get rid of some of the debris that the ABS includes
both <-
  both %>%
  select(-X1,
         -'MB by Commonwealth Electoral Divisions (UR)',
         -'Data Source: Census of Population and Housing, 2016, TableBuilder',
         -'INFO',
         -'Copyright Commonwealth of Australia, 2018, see abs.gov.au/copyright',
         -'ABS data licensed under Creative Commons, see abs.gov.au/ccby'
         ) %>%
  rename(age_group = 'AGEP Age',
         heap = 'HEAP - 1 Digit Level',
         gender = 'SEXP Sex')

both <- rbind(both[1:20,], both[23:42,])

# Now we push the age-groups into each cell.
both <-
  fill(both, age_group)

census_data <- gather(both,
                      key = "division",
                      value = "number",
                      Adelaide:Total)

rm(both)
```

Finally, we need to construct proportions for each of our sub-groups.

```
# Check the type before we convert to integer
# Here we're looking for anything that may not nicely convert to a number.
census_data$check_number <- str_detect(census_data$number, "[:digit:]")
census_data$check_alpha <- str_detect(census_data$number, "[:alpha:]")
```

```r
census_data$check_punct <- str_detect(census_data$number, "[:punct:]")
census_data$check_blank <- str_detect(census_data$number, "[:blank:]")
table(census_data$check_number)
table(census_data$check_alpha)
table(census_data$check_punct)
table(census_data$check_blank)

census_data$number <- as.integer(census_data$number)

census_data <-
  census_data %>%
  select(-check_number,
         -check_alpha,
         -check_punct,
         -check_blank)

census_data$age_group[census_data$age_group == "age18to29"] <- "ages18to29"

census_data <-
  census_data %>%
  select(division, gender, age_group, heap, number) %>%
  arrange(division, gender, age_group, heap)

census_data <-
  census_data %>%
  filter(age_group != "Total") %>%
  filter(division != "Total")



#### Make proportions ####
census_data <-
  census_data %>%
  group_by(division) %>%
  mutate(total = sum(number)) %>%
  mutate(cell_proportion_of_division_total = number / total) %>%
  ungroup() %>%
  select(-total)

census_data <-
  census_data %>%
  rename(education = heap)
```

We're just going to do some rough forecasts that won't have appropriate errors. For

each is_male and in_age_group we want the relevant coefficient in the example_data.

```
#### Individual-level model ####
model <- brm(ALP_supporter ~ gender + age_group + education + (1|division),
                data = LINA_regression_data,
                family = bernoulli(),
                cores = number_of_cores,
                file = "outputs/models/LinA" # If running this again, either
              # need to change this file name to something new or delete
              # the saved model from the folder.
                )
pairs(model)
ranef(model)
summary(model)
```

Now we can construct the estimates

```
post_stratified_posterior_estimates_LinA <-
  LinA_model %>%
  tidybayes::add_predicted_draws(newdata = poststratification_data, allow_new_levels = T
  rename(alp_predict = .prediction) %>%
  mutate(alp_predict_prop = alp_predict*cell_proportion_of_division_total) %>%
  group_by(division, .draw) %>%
  summarise(alp_predict = sum(alp_predict_prop)) %>%
  ungroup()
```

So we find that the areas that skew young are more likely to vote Australian Labor Party, than the areas that skew old.

B Variants

B.1 *State-based-geography*

TBD

Redo with similar model, but only take state inputs, not electoral-division. Likely would increase the number of survey inputs available.

B.2 *First-preferences only*

TBD

Redo just considering first-preferences. Substantial decrease in complexity and uncertainty, but what is the tradeoff in terms of accuracy?

References

ABS (2016). *TableBuilder*. Australian Bureau of Statistics, Canberra, Australia. Findings based on use of ABS TableBuilder data, last accessed 10 March 2019.

ABS (2018a). *Australian Demographic Statistics*. Australian Bureau of Statistics, Canberra, Australia. Cat no 3101.0, Table 59, Estimated Resident Population By Single Year Of Age, Australia, released 20/12/2018, as downloaded on 21 September 2019.

ABS (2018b). *Australian Statistical Geography Standard (ASGS): Volume 3 - Non ABS Structures, July 2018, 'Correspondence CED 2017 to CED 2018'*. Australian Bureau of Statistics, Canberra, Australia. Cat no 1270.0.55.003, released 28/08/2018, as downloaded on 21 September 2019.

Alexander, M. (2019a). *Analyzing Name Changes After Marriage Using A Non-Representative Survey*. Available at: https://www.monicaalexander.com/posts/2019-08-07-mrp/, last accessed 30 September 2019.

Alexander, M. (2019b). Reproducibility in Demographic Research. Keynote address at the Open Science Workshop, Max Planck Institute for Demographic Research, Rostock, Germany, October 10.

Alexander, M. and Alexander, R. (2016). *Small Scale Australian Political Polling*. Petit Poll.

Alexander, R. (2019c). *AustralianElections: Datasets on Australian Elections*. Findings based on use of version 0.08, as accessed 19 September 2019.

Alkema, L. and New, J. R. (2015). Family Planning Estimation Tool (FPET). Available at http://fpet.track20.org/.

Bivand, R., Keitt, T., and Rowlingson, B. (2019). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.4-4.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28.

Buttice, M. K. and Highton, B. (2013). How Does Multilevel Regression and Poststratification Perform with Conventional National Surveys? *Political Analysis*, 21(4):449–467.

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

Cohn, N. (2016). We Gave Four Good Pollsters the Same Raw Data They Had Four Different Results. *New York Times*.

Dassonneville, R., Feitosa, F., Hooghe, M., Lau, R. R., and Stiers, D. (2019). Compulsory Voting Rules, Reluctant Voters and Ideological Proximity Voting. *Political Behavior*, 41(1):209–230.

Dove, C. (2019). *Life in Australia - Wave 26, Technical Report, May 2019*. The Social Research Centre, Melbourne, Australia. Version: 1.0 - May 2019.

Dowding, K. and Miller, C. (2019). On prediction in political science. *European Journal of Political Research*, 58(3):1001–1018.

Dumond, P. (2019). *Methodology for the Voting Advice and Matching*. Available at: https://australia.smartvote.org/assets/pdf/Positional_matching_methodology.pdf.

Forbes, J., Cook, D., and Hyndman, R. J. (2019). Spatial modelling of the two-party preferred vote in Australian federal elections.

Fowler, A. (2013). Electoral and policy consequences of voter turnout: Evidence from compulsory voting in Australia. *Quarterly Journal of Political Science*, 8(2):159–182.

Gelman, A. and Little, T. C. (1997). Poststratification into Many Categories Using Hierarchical Logistic Regression. *Survey Methodology*, 23(2):127–135.

Germann, M. and Gemenis, K. (2019). Getting Out the Vote With Voting Advice Applications. *Political Communication*, 36(1):149–170.

Goel, S., Obeng, A., and Rothschild, D. (2015). Non-representative surveys: Fast, cheap, and mostly accurate. In *Working Paper*.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., and Altman, D. G. (2016). Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, 31(4):337–350.

Handmer, J. and Proudley, B. (2007). Communicating uncertainty via probabilities: The case of weather forecasts. *Environmental Hazards*, 7(2):79–87.

Hannan-Morrow, S. and Roden, M. (2014). Gender, Age and Generational Effects on Turnout in Australian Federal Elections. In *The Australian Political Studies Association Annual Conference, University of Sydney Paper*.

Hanretty, C. (2019). An introduction to multilevel regression and post-stratification for estimating constituency opinion. *Political Studies Review*, pages 1–16.

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

Hoekstra, R., Morey, R. D., and Wagenmakers, E.-J. (2018). Improving the interpretation of confidence and credible intervals. In Sorto, M. A., White, A., and Guyo, L., editors, *Looking back, looking forward. Proceedings of theTenth International Conference on Teaching Statistics*.

i Coma, F. M. and Lago, I. (2018). Gerrymandering in comparative perspective. *Party Politics*, 24(2):99–104.

Jackman, S. (1999). Non-compulsory voting in Australia?: What surveys can (and can't) tell us. *Electoral Studies*, 18(1):29–48.

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

Jackman, S., Ratcliff, S., and Mansillo, L. (2019). *Small area estimates of public opinion: Model-assisted post-stratification of data from voter advice applications*. Paper presented at 2019 Annual Asian Political Methodology Meeting. Available at: https://www.cambridge.org/core/membership/services/aop-file-manager/file/5c2f6ebb7cf9ee1118d11c0a/APMM-2019-Simon-Jackman.pdf.

Jamieson, K. H., McNutt, M., Kiermer, V., and Sever, R. (2019). Signaling the trustworthiness of science. *Proceedings of the National Academy of Sciences*, 116(39):19231–19236.

Kahle, D. and Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161.

Kastellec, J. P., Lax, J. R., and Phillips, J. (2016). Estimating State Public Opinion With Multi-Level Regression and Poststratification using R. *Working Paper*.

Kay, M. (2019). *tidybayes: Tidy Data and Geoms for Bayesian Models*. R package version 1.1.0.

Kleinnijenhuis, J., van de Pol, J., van Hoof, A. M., and Krouwel, A. P. (2019). Genuine effects of vote advice applications on party choice: Filtering out factors that affect both the advice obtained and the vote. *Party Politics*, 25(3):291–302.

Konitzer, T. B. and Rothschild, D. M. (2016). Why Political Polling is not Dead – A Plea for Non-Probability Polling, Algorithms, and Big Data.

Larsen, E. G. and Fazekas, Z. (2019). Transforming stability into change: How the media select and report opinion polls. *The International Journal of Press/Politics*.

Lauderdale, B. E., Bailey, D., Blumenau, Y., and Rivers, D. (2017). Model-Based Pre-Election Polling for National and Sub-National Outcomes in the US and UK. *Unpublished Work). Available online: https://www. jackblumenau. com/papers/mrp_polling. pdf (accessed on 5 December 2018)*.

Leigh, A. (2005). Economic Voting and Electoral Behavior: How Do Individual, Local, and National Factors Affect the Partisan Choice? *Economics & Politics*, 17(2):265–296.

Mackerras, M. and McAllister, I. (1999). Compulsory voting, party stability and electoral advantage in Australia. *Electoral Studies*, 18(2):217–233.

Mills, S. and Tiffen, R. (2012). Opinion Polls and the Media in Australia. In *Opinion Polls and the Media*, pages 155–174. Springer.

Mislavsky, R. and Gaertig, C. (2019). Combining probability forecasts: 60% and 60% is 60%, but likely and likely is very likely. *Available at SSRN 3454796*.

Morss, R. E., Demuth, J. L., and Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: A survey of the US public. *Weather and forecasting*, 23(5):974–991.

Morss, R. E., Lazo, J. K., and Demuth, J. L. (2010). Examining the use of weather forecasts in decision scenarios: Results from a US survey with implications for uncertainty communication. *Meteorological Applications*, 17(2):149–162.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., and Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, 1(1):1–9.

Munzert, S. (2017). Forecasting elections at the constituency level: A correction–combination procedure. *International journal of forecasting*, 33(2):467–481.

Oleskog Tryggvason, P. and Strömbäck, J. (2018). Fact or Fiction? Investigating the quality of opinion poll coverage and its antecedents. *Journalism Studies*, 19(14):2148–2167.

Patt, A. G. and Schrag, D. P. (2003). Using specific language to describe risk and probability. *Climatic change*, 61(1-2):17–30.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

Pereira, M. M. (2019). Do parties respond strategically to opinion polls? Evidence from campaign statements. *Electoral Studies*, 59:78–86.

Popp, R., Horvath, L., Banducci, S., Coan, T., and Krouwel, A. (2016). *What Voting Advice Applications can teach us about Voters and Elections*. Paper presented at the 2016 ECPR General Conference in Prague.

R Core Team (2019a). *foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...* R package version 0.8-72.

R Core Team (2019b). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A. E., Alkema, L., and Gerland, P. (2014). Bayesian population projections for the United Nations. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29(1):58.

Reavy, M. M. and Pavlick, K. (2018). The de-emphasis of sampling error in coverage of presidential approval ratings. *Newspaper Research Journal*, 39(2):180–190.

Ricketson, M., Dodd, A., Zion, L., and Winarnita, M. (2019). "like being shot in the face" or "i'm glad i'm out": Journalists' experiences of job loss in the australian media industry 2012–2014. *Journalism Studies*, pages 1–18.

Robinson, D. and Hayes, A. (2019). *broom: Convert Statistical Analysis Objects into Tidy Tibbles*. R package version 0.5.2.

Rothschild, D. (2015). Combining forecasts for elections: Accurate, relevant, and timely. *International Journal of Forecasting*, 31(3):952–964.

Sheppard, J. and Beauregard, K. (2018). Early voting in Australia: The costs and benefits of convenience. *Political Science*, 70(2):117–134.

Sheppard, J., Taflaga, M., and Jiang, L. (2019). Explaining high rates of political participation among Chinese migrants to Australia. *International Political Science Review*.

Stoetzer, L. F., Neunhoeffer, M., Gschwend, T., Munzert, S., and Sternberg, S. (2019). Forecasting Elections in Multiparty Systems: A Bayesian Approach Combining Polls and Fundamentals. *Political Analysis*, 27(2):255–262.

Stoetzer, L. F. and Orlowski, M. (2019). Estimating coalition majorities during political campaigns based on pre-election polls. *Journal of Elections, Public Opinion and Parties*, pages 1–12.

Toff, B. (2019). The 'Nate Silver effect' on political journalism: Gatecrashers, gatekeepers, and changing newsroom practices around coverage of public opinion polls. *Journalism*, 20(7):873–889.

van der Eijk, C., van der Brug, W., Kroh, M., and Franklin, M. (2006). Rethinking the dependent variable in voting behavior: On the measurement and analysis of electoral utilities. *Electoral Studies*, 25:424–447.

Vehtari, A., Simpson, D. P., Yao, Y., and Gelman, A. (2019). Limitations of "Limitations of Bayesian leave-one-out cross-validation for model selection". *Computational Brain & Behavior*, 2(1):22–27.

Wang, W., Rothschild, D., Goel, S., and Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3):980–991.

Westwood, S., Messing, S., and Lelkes, Y. (2019). Projecting confidence: How the probabilistic horse race confuses and demobilizes the public. *Available at SSRN 3117054*.

Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1.

Wickham, H. and Miller, E. (2019). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. R package version 2.1.1.

Young, S. (2010). *How Australia decides: Election reporting and the media*. Cambridge University Press.