**Election poll aggregation: is it biased towards the Liberal–National Coalition?**

Student’s Name: Jiaxuan Yi

Student’s Id: 33612676

Email: jyii0017@student.monash.edu

Supervisor: Damjan Vukcevic

Email: damjan.vukcevic@monash.edu

Contents

[1. Abstract 2](#_Toc1342)

[2. Introduction 2](#_Toc24295)

[2.1 Background 2](#_Toc27233)

[2.2 Motivation 3](#_Toc22790)

[2.3 Research Question 3](#_Toc2698)

[2.4 Contribution 4](#_Toc11158)

[3. Data 5](#_Toc24027)

[4. Methodology 5](#_Toc9281)

[4.1 Experimental designs: 5](#_Toc8478)

[Experiment 1: 5](#_Toc17229)

[Experiment 2: 5](#_Toc14234)

[4.3 Key variables: 6](#_Toc6807)

[4.4 Research methodology: 6](#_Toc164)

[5. Results 7](#_Toc30049)

[6. Conclusion 7](#_Toc26289)

[References 8](#_Toc6045)

# Abstract

Political opinion polls are particularly prominent during Australian federal elections. Polls are often desired to track, in real time, the support of each political party among different groups of people and the direction of the next federal election. Considering that the results of different polls often vary greatly, researchers try to integrate the data through multiple methods to obtain more accurate information about voters' intentions.

Jackman (2005) proposed a statistical model for poll aggregation, which is currently one of the dominant Australian public opinion aggregation models. It considers institutional biases within each different pollster and fluctuations in voter sentiment during the campaign to track changes in voter support over time. And the model can input actual election results as anchors to obtain corrected inferences (i.e., backcasting) for the sentiment back in time. In the application of the model to past Australian federal elections, it is interesting to note that the Liberal-National Coalition (LNP) has always shown a strong upward trend in support in the last few weeks before an election. One possible explanation is that it reflects an underlying political reality (i.e., the LNP's specific campaign tactics that put the Coalition ahead of Labor in the final stages of the campaign). An alternative explanation is that this trend is an artifact of this poll aggregation model, created during the backcasting process. No study has yet systematically explored the causes of this phenomenon.

This paper aims to investigate whether this poll aggregation model is biased towards the Liberal–National Coalition. I will analyze these possible explanations by exploring the properties and robustness of the model through simulated data, Bayesian inference, and other techniques. If the model is found to be biased, this research will improve the model to increase its robustness. This paper will provide a new literature perspective on election poll aggregation research and improve the reliability of poll aggregation. In addition, by fitting polling data from the latest 2025 Australian federal election into the model, this study will provide a timely case study for the application of poll aggregation models.

# Introduction

# **2.1 Background**

The simplest method of poll aggregation is to average estimates from recent polls, another is to use locally weighted scatterplot smoothing (LOESS) to generate poll estimates (Jackson, 2018). However, simply averaging or weighting polling data can sometimes increase the overall error, as such approaches often fail to account for systematic effects introduced by the methodological choices of different polling organisations—commonly referred to as “house effects” (Jackman, 2005).

Thus, researchers began to adopt more sophisticated Dynamic Linear Models (DLMs) to handle noisy polling data (Fisher et al., 2011; Jackman, 2005; Linzer, 2013). DLMs rely on random-walk Markov processes and Kalman filtering techniques (West & Harrison, 1997) to estimate the latent public support for each political party. In particular, Jackman (2005) proposed a state-space model based on Kalman filtering. The model adjusts for“house effects”and accounts for fluctuations in voter sentiment during the campaign by modelling vote intention as a simple random walk. In addition, it incorporates actual historical election outcomes as fixed anchors to retrospectively calibrate the model, thereby improving the estimation of historical trends in public opinion. Subsequent analyses of political poll aggregation in Australia have largely been based on adaptations and refinements of this model (Jackman & Mansillo, 2018; Mansillo & Evershed, 2016; Nicholas et al., 2025). In addition, Linzer (2013) applied a similar dynamic linear model to pre-election polling data from the 2008 U.S. presidential election and produced accurate estimates of the final outcome.

However, despite various methodological extensions and refinements to these poll aggregation models—such as the incorporation of seasonal components (Walther, 2015)—there remains a lack of interpretive research on the outputs they produce. Specifically, in past Australian election polling cycles, Jackman’s (2005) model has repeatedly shown a sustained increase in support for the Liberal–National Coalition (LNP) during the final stages of the campaign (Jackman, 2023), yet the underlying causes of this pattern have not been systematically examined in the existing literature. This paper will systematically explore possible explanations for this phenomenon.

# **2.2 Motivation**

This research was motivated by a desire to enhance the reliability of election poll aggregation. As the 2025 Australian federal election gets underway, political polls are coming thick and fast. Throughout this period, The Guardian (Nicholas et al., 2025) applied Jackman's (2005) dynamic linear model to continually update the results of opinion polls for the Australian federal election. These real-world incentives drew my attention to the model. One interesting feature of the model in past election results is that there has traditionally always been a strong trend in favor of the Liberal-National Party (LNP) coalition in the final weeks before an election (Jackman, 2023). One interesting feature of the model in past election results is that there is always a strong trend in favor of the Liberal-National Party (LNP) coalition in the final weeks before an election (Jackman, 2023). Thus, there is an interesting key point here: does the phenomenon reflect actual LNP trends in voter intentions, or is it merely an artifact generated by the model in backcasting? This phenomenon has not yet been fully explored.

It is important to examine possible explanations for this phenomenon. As the 2025 Australian federal election unfolds, the influence of polls on public opinion and political behavior becomes more apparent. Looking back at the 2019 federal election, there had been significant bias in the polls that severely eroded public trust (Goot, 2021). In addition, considering that Jackman's model is widely adopted as the mainstream structure of poll aggregation, its poll averages affect public perceptions of the electorate and political parties' campaign strategies. Therefore, it is particularly important to re-examine the bias and reliability of the poll aggregation model in the current context and to explore poll aggregation systematically.

2.3 Research Question

The research question is to analyze whether this Jackman's model of election poll aggregation is biased towards the Liberal–National Coalition. The research question will be for four steps:

* Step 1: Explore the properties and robustness of the model.

Use polling data from the 2007-2022 Australian federal elections to replicate the trend in corrected Labor (ALP) polling averages estimated by Jackman (2023) from the model. First, starting with the simplest model, I implemented a basic Normal-Inverse-Gamma (NIG) model to analyze polling data. Subsequently, I extended the model by incorporating the time trend and the house effect separately. Then, I included both the time trend and the house effect simultaneously. Finally, I replaced the basic NIG framework with Jackman’s full specification, thereby completing the fitting of the Jackman’s model.

This process serves as a preliminary “warm-up” step in the study, with the core objective of verifying whether the theoretical framework, model implementation, and code execution used in this research are consistent with the original study. Only after confirming that the model can accurately reproduce existing results can the subsequent steps be considered theoretically and methodologically valid.

* Step 2: Examine two possible explanations for the trend in poll averages from the poll aggregation model.

Based on the confirmation that the model is able to function correctly, the second step is to further explore through simulation the central question of this study: whether the model is biased. To this end, I specify two distinct distributional models based on two competing interpretations of the model’s results, and generate simulated data accordingly to serve as observed inputs for estimation within the model. By comparing the fit between the two sets of estimates output by the model and the preset “true parameter values”, it is possible to test whether the model accurately recovers the respective trends, and thus to assess which interpretation is more plausible.

* Step 3: If the model is biased, further refine the model to improve its robustness.
* Step 4: Fitting the model with new data from the 2025 Australian federal election to explain this election.

2.4 Contribution

First, this research responds to a critical gap in the current literature that lacks an explanatory analysis of the “trend phenomenon” in the output of poll aggregation models. I hope to provide new theoretical perspectives on the study of election poll aggregation through an in-depth exploration of Jackman's (2005) model. Second, by assessing the robustness of the model through simulation and Bayesian inference systematically, this research will identify whether the model has structural biases and improves the reliability of the model itself and the accuracy of future polling estimates. Third, by improving the test model, this study helps voters, the media, and political experts to correctly interpret the pre-election data, and enhances the public's trust in Australian polls as well as the science of political parties' campaign strategy development. Finally, by bringing the recent polling data from the 2025 federal election into the model, this paper will provide a timely empirical analytical perspective to further analyze the applicability of the model in the current electoral context.

# Data

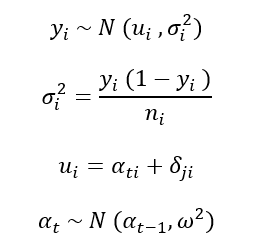
The primary data source for this paper will be a continuation of the existing data used by Jackman (2023) in his model to ensure that the results are comparable to existing research. The polling data he used from the 2007 to 2022 Australian federal elections are available on his GitHub page (Jackman, 2022/2024). Building on this, and in order to extend the application of the model to the most recent election, this paper will also collect polling data released in the run-up to the 2025 Australian Federal Election from the following major organizations: Essential, Newspoll, Resolve, YouGov Australia, and Roy Morgan. Detailed data content and code can be found in the appendix.

# Methodology

* 1. Experimental designs

The experimental designs correspond to the steps in the research question:

Experiment 1: I apply polling data for the period from 2007 to the 2022 Australian federal election to the state space model described by Jackman (2005). Define as the proportion of the vote received by a party in poll i:



The result of the poll is equal to plus random measurement error with variance . The value of is equal to the true voting intention on the day of the poll, , plus the pollster's systematic bias, . True voting intention is modeled as a function of past voting intention (random wandering) with a variance of for daily perturbations. is observed, the support level of the corresponding political party in the last election. The variance of the random measurement error term is specified as , is the poll’s sample size. In addition, we assume that the polls are collectively unbiased, i.e., . See Jackman (2005) for additional details. Comparison of the model output with the results of Jackman (2023) verifies the accuracy and consistency of the model implementation in this research.

Experiment 2: After confirming that the model runs correctly, experiment 2 examines whether the model exhibits bias using simulated data.

First, I set up two sets of representative distributional models from which I generate simulated data corresponding to two interpretations of polling trends: one is that the models reflect real political trends. Thus, the first distribution is set such that the fluctuations in support y (our simulated ALP support level) converge with time t (indexed to the campaign day) in the weeks leading up to the election, i.e., the magnitude of the fluctuations continues to decline. The second is that the model is biased. Thus, the second distribution set up has relatively uniform fluctuations in y over time t with no significant trend change. Subsequently, the two sets of simulated data are input into the model, and the posterior distributions of the parameters are estimated using the Markov Chain Monte Carlo (MCMC) method. Finally, by comparing the estimated results with the known 'true parameters,' the model's ability to accurately recover each predefined scenario is assessed.

In these two experiments, my election poll aggregation method allows for dependence across polls.

* 1. Key variables

The key variables of interest in this research include real and simulated data for model validation and bias testing, respectively. Experiment 1 uses real polling data, which is used to test whether the model implemented in this study is consistent with the original model (Jackman, 2005), thus validating the validity of the model run. Building on this, Experiment 2 introduces simulated data to evaluate the model’s performance under a known data-generating process, in order to assess whether it exhibits systematic bias and to address the core research question.

* 1. Research methodology

The methodology of this research will include Bayesian inference, simulation, and statistical modeling.

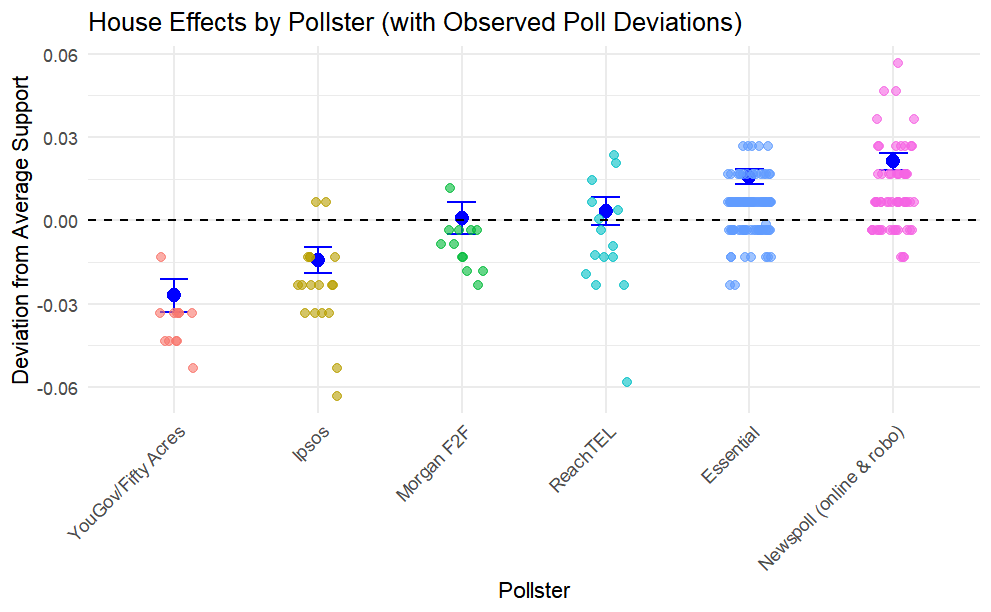
Bayesian methods allow for the incorporation of prior knowledge, e.g., experience of past election results as a priori information, thus enhancing the stability of the estimates in the presence of limited sample size or noisy data. Given that election polls are usually cost-constrained and have small sample sizes, and that frequency theory methods rely on large-sample characteristics, Bayesian inference is more suitable for poll aggregation analysis. In Experiment 1, I employ a Bayesian approach to estimate the unknown parameters in the model and generate their posterior distributions using the MCMC method (Jackman, 2005).

As Gelman et al. (n.d.) point out, simulation studies are one of the fundamental tools for evaluating Bayesian models. Since voters' true intentions at each point in time are unobservable, it is difficult to directly assess model bias. Simulation offers a controlled and transparent environment with known parameters, making it essential for identifying potential bias in this model. In Experiment 2, computer simulation is used to generate virtual data with known parameters corresponding to two possible real-life scenarios, so as to test the performance of the model's inference under different data structures at a lower cost. Experiment 2 uses a computer to simulate the generation of virtual data with known parameters corresponding to two possible real-life scenarios, so as to test the performance of the model's inference under different data structures at a lower cost.

Statistical modeling is used throughout this study: in Experiment 1, it is essential for correctly understanding and implementing the structure and inferential process of the Bayesian dynamic linear model (Jackman, 2005); in Experiment 2, the design of the simulation study also relies on modeling techniques to specify representative parameter distributions and generate controlled synthetic data for systematically evaluating the model’s inferential performance and robustness.

# Results

I have successfully fitted the time trend and house effects based on the NIG model.



Next, I will replace the NIG model with Jackman's model.

# Conclusion

**Appendices**

[Yijiaxuan/Trying\_an\_NIG\_model](https://github.com/Yijiaxuan/Trying_an_NIG_model)

# References

Fisher, S. D., Ford, R., Jennings, W., Pickup, M., & Wlezien, C. (2011). From polls to votes to seats: Forecasting the 2010 British general election. Electoral Studies, 30(2), 250–257. https://doi.org/10.1016/j.electstud.2010.09.005

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (n.d.). Bayesian Data Analysis Third edition (with errors ﬁxed as of 20 February 2025).

Jackman, S. (2005). Pooling the polls over an election campaign. Australian Journal of Political Science, 40(4), 499–517. https://doi.org/10.1080/10361140500302472

Jackman, S. (2023, August 25). index—Estimating mass sentiment from noisy, biased signals:recent Australian elections and the Voice referendum. https://simonjackman.github.io/poll\_averaging\_voice\_2023/presentation/Monash\_2023\_08\_25.html#/title-slide

Jackman, S. (2024). Simonjackman/poll\_averaging\_aus\_2022 [R]. https://github.com/simonjackman/poll\_averaging\_aus\_2022 (Original work published 2022)

Jackman, S., & Mansillo, L. (2018). The Campaign that Wasn’t: Tracking Public Opinion over the 44th Parliament and the 2016 Election Campaign. In A. Gauja, P. Chen, J. Curtin, & J. Pietsch (Eds.), Double Disillusion: The 2016 Australian Federal Election (1st ed., pp. 133–156). ANU Press. https://doi.org/10.22459/DD.04.2018.06

Jackson, N. (2018). The Rise of Poll Aggregation and Election Forecasting. In L. R. Atkeson & R. M. Alvarez (Eds.), The Oxford Handbook of Polling and Survey Methods (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190213299.013.28

Linzer, D. A. (2013). Dynamic Bayesian Forecasting of Presidential Elections in the States. Journal of the American Statistical Association, 108(501), 124–134. https://doi.org/10.1080/01621459.2012.737735

Mansillo, L., & Evershed, N. (2016, May 8). Pooling the polls: How we’re tracking opinion polling for the Australian election. The Guardian. https://www.theguardian.com/australia-news/datablog/2016/may/08/pooling-the-polls-how-were-tracking-opinion-polling-for-the-australian-election

Nicholas, J., Mansillo, L., & Evershed, N. (2025, April 25). Australian election 2025 poll tracker: Labor v Coalition latest opinion polls results. The Guardian. https://www.theguardian.com/australia-news/ng-interactive/2025/apr/25/australia-election-polls-latest-aus-opinion-poll-tracker-results-current-polling-survey-labor-vs-liberal-dutton-albanese

Walther, D. (2015). Picking the winner(s): Forecasting elections in multiparty systems. Electoral Studies, 40, 1–13. https://doi.org/10.1016/j.electstud.2015.06.003

West, M., & Harrison, J. (Eds.). (1997). The Dynamic Linear Model. In Bayesian Forecasting and Dynamic Models (pp. 97–142). Springer. https://doi.org/10.1007/0-387-22777-6\_4