

7

NATIONAL POLLING AND OTHER DISASTERS

Luke Mansillo and Simon Jackman

Perhaps one of the most important and unexpected features of the 2019 Australian federal election was the fact that public polls ‘got it wrong’. Polls pointing to a Labor win dominated pre-election prognostication, while lamentations and dissection of polling error dominated post-election analysis.

As we argue below, polling errors in the 2019 election were large, but not completely unprecedented. Following generally good performances in 2016 and 2013, few commentators or analysts recalled that the 2010 election saw large errors in the polls. What was especially noteworthy about 2019 was the consistency of the poll results indicating Labor would win, and the extent to which Australian political actors—politicians, media and the public—‘locked in’ behind these poll-shaped but wrong expectations that had shaped the content and conduct of Australian politics not just during the election campaign, but also for some years previously.

The large and extensive influence of polls on mass and elite politics is well documented in the political science literature. Polls shape how elites develop and implement public policy (Burstein 2003), how they decide whether to engage in leadership ructions (Brown 1992) and the decisions of individual politicians to resign from politics (Stone et al. 2010). We note that 13 Liberal–National Coalition parliamentarians retired ahead of the

2019 election, nine of them from the ministry. Individual and corporate donors also rely on the polls in assessing whether, and to whom, to make campaign contributions (Fuchs et al. 2000).

Polls also provide politicians with opportunities to advance issues and agendas on behalf of their constituencies or their parties or to serve more personal causes, such as leadership ambitions (Meyer and Minkoff 2004; Petersen et al. 2008). Critically, in Westminster systems—in which party leadership is in some sense under constant scrutiny, coupled with the advent of a more ‘presidentialised’ prime ministership (Hart 1992; see also Chapter 6, this volume)—polling has taken on even greater political significance. Polls provide backbenchers from marginal seats with a regular barometer of their party’s chances at the next election. Indeed, a key feature of Australian politics over the past decade has been the way poll results have driven leadership challenges and turnover, with profound consequences for who governs and how they govern.

Hence, understanding the magnitude and sources of polling error in the 2019 Australian election is not a mere technical exercise, but is fundamental to understanding Australian politics, policies and leaders.

Estimating poll bias

We begin by estimating the magnitude of the polls’ errors, or biases, in the 2019 election cycle. We rely on a methodology that one of us helped pioneer (Jackman 2005, 2009) and is used widely by polling analysts in Australia (Mark the Ballot 2017) and internationally (Ellis 2017; Pickup et al. 2011). We previously deployed this modelling strategy to analyse polls leading up to the 2016 Australian federal election (Jackman and Mansillo 2018).

We fit a statistical model to public polls fielded between the 2016 and 2019 elections. For each poll, we usually observe the following: 1) estimates of voting intentions, which are typically reported as an integer percentage; 2) the field dates; 3) the number of poll respondents; 4) the population of interest (typically, all adult Australian citizens); and 5) the identity of the polling company conducting the poll. Most media reports of polls do not contain even this much information. Sometimes reports omit information such as the number of respondents or field dates, sampling

methodologies and weighting procedures applied after data collection, which almost never appear in media reports of polls or in the polling companies' description of their methods.

Nonetheless, with this basic information from a series of polls, we set about recovering the day-by-day trajectory of voting intentions between the 2016 and 2019 elections. Let t index the $T = 1,051$ days between the 2016 and 2019 elections and let ξ_t denote true (but 'latent' or indirectly observed) voting intentions on day t . Of great utility is that ξ is revealed on election day, both in 2016 and in 2019 on days $t = 1$ and $t = T = 1,051$, once the AEC publishes final election results.¹ These two election day 'endpoint' observations serve as 'anchors' or fixed points that let us reconstruct the trajectory of voting intentions from observed polls, at the same time as we estimate and correct for biases in the polls.

Poll p fielded on day t by polling company j supplies an estimate of ξ_t , reported as proportion $y_p \in [0,1]$ (conventionally reported after rounding to two digits of precision), with sample size n_p . We assume that the variance of each poll estimate is $V_{p(t)} = y_p(1 - y_p) / n_{p(t)}$, which is a reasonable approximation absent detailed information about the polls' sampling and weighting methodologies.²

Each poll result, y_p , is assumed to be a sum of the true but unobserved state of voting intentions on the corresponding day, $\xi_{t(p)}$, and the bias of the corresponding pollster, denoted as $\delta_{j(p)}$.³ Averaging over the random error produced by sampling, we assume $E(y_p) = \xi_{t(p)} + \delta_{j(p)}$. For polls with the sample sizes considered here, we can conveniently assume a normal law for the poll results—that is, $y_p \sim N(\xi_{t(p)} + \delta_{j(p)}, V_p)$.⁴

1 Treating the published election results as errorless measures of the corresponding ξ_t assumes no meaningful errors in the national vote counts nor any sizeable electoral fraud; these assumptions seem valid in the context of both the 2016 and the 2019 Australian federal elections.

2 This variance can be interpreted as a measure of the uncertainty one possesses about true voting intentions, ξ_t , given an unbiased poll of sample size n_p acquired via simple random sampling. If 1) poll p is biased or 2) simple random sampling was not deployed, the actual variance will differ slightly from the expression in the text.

3 We use the notation $t(p)$ to refer to the day(s) in which poll p was in the field and $j(p)$ to refer to the polling house j that fielded poll p . The total sample of a poll n_p we divide by $t(p)$ days in poll p to give the incorporate information from each over each day of fielding with the correct precision. This produces a smoother estimate of voting intentions than the previous attempts that used the median date of fielding (Jackman 2005; Jackman and Mansillo 2018).

4 Most polls have a multiday field period, which we accommodate with the following procedure. A poll fielded between days t and $t + r$ is modelled as a function of ξ_t, \dots, ξ_{t+r} , spreading the polls' sample $n_{(p)}$ uniformly over the $r + 1$ field days; this 'per field day' sample size, $n_p / (r + 1)$, appears in the denominator of the expression for $V_{(p)}$ given in the previous paragraph.

The model also exploits the fact that, despite political upheavals such as leadership spills, voting intentions change relatively slowly. We assume that a priori voting intentions on day t are, on average, unchanged from the previous day. Day-to-day changes in the ξ_t follow a normal distribution, with a small but unknown variance, ω^2 —that is, $\xi_t \sim N(\xi_{t-1}, \omega^2)$. We augment the model to include a discontinuity (or ‘jump’) on 28 August 2018, the day Morrison replaced Turnbull as Liberal leader and prime minister.⁵

The inferential task is to recover estimates of: 1) the sequence of latent voting intentions, $\{\xi_t\}$; 2) the house effects, δ_j ; 3) the innovation variance parameter, ω^2 ; and 4) the discontinuity parameter, γ . Technical details on estimation and underlying computation appear in Jackman (2009). We derive estimates separately for the ALP, Liberal–National Coalition, the Australian Greens and other primary voting intentions as well as two-party-preferred voting intentions.

Data

First, we analyse a dataset of 226 polls, fielded between the 2016 and 2019 elections. These polls span six distinct polling companies and methodology sets, shown in Table 7.1. We treat three Galaxy polls fielded in 2019 as in the ‘Newspoll’ category, as they were fielded a considerable time after YouGov had acquired Galaxy but with the identical online and interactive voice response (IVR) interview modes that Galaxy Research had used for the Newspoll-branded polls commissioned by *The Australian* newspaper.

The 12 YouGov/Fifty Acres-branded polls were conducted online, administered to respondents drawn from the YouGov national panel, with quota sampling designed to produce a nationally representative sample with respect to age, gender and region. YouGov partnered with Fifty Acres, a government and public relations firm, to produce these polls. Public polling is highly concentrated in the hands of a small number of firms; weekly polls from Essential are the sole, high-frequency alternative to Newspoll/Galaxy/YouGov polls. The Newspoll/Galaxy polls use a combination of IVR phone and online panel surveys (details of the

5 On that day the model becomes $\xi_t \sim N(\xi_{t-1} + \gamma, \omega^2)$, in which γ is the jump parameter. We make no a priori assumption as to whether γ is positive or negative.

proportions of each are not public knowledge), while Essential uses only online panel surveys and Ipsos uses live-interviewer computer-assisted telephone interviewing surveys. The few Roy Morgan polls were all conducted with face-to-face interviews as part of their regular consumer and market research surveys.

To establish historical benchmarks for the errors of the 2019 polls, we also analyse a dataset of 1,624 polls, fielded between the 2004 and 2019 elections. The analysis described above is repeated for the four prior parliamentary periods.

Table 7.1 Polls by survey house

Survey house	No. of polls
Essential	108
Ipsos	18
Newspoll	61
ReachTEL	15
Roy Morgan (Face-to-Face)	12
YouGov	12

Source: Authors' calculations.

Estimated trajectories of voting intentions, 2016–19

The circular symbols in Figure 7.1 correspond to estimates from public polls of the ALP's share of first-preference vote intentions, 2016 to 2019. The two horizontal reference lines indicate the 2016 and 2019 election results, with Labor winning 34.7 per cent and 33.3 per cent of first preferences, respectively. The darker, thicker line undulating across the graph is the output of our model, the sequence of daily estimates of ξ_t between the 2016 and 2019 elections; the effect of the 'endpoint' constraints are clearly visible, with the estimates of true public opinion beginning and terminating on the observed, actual 2016 and 2019 election results.

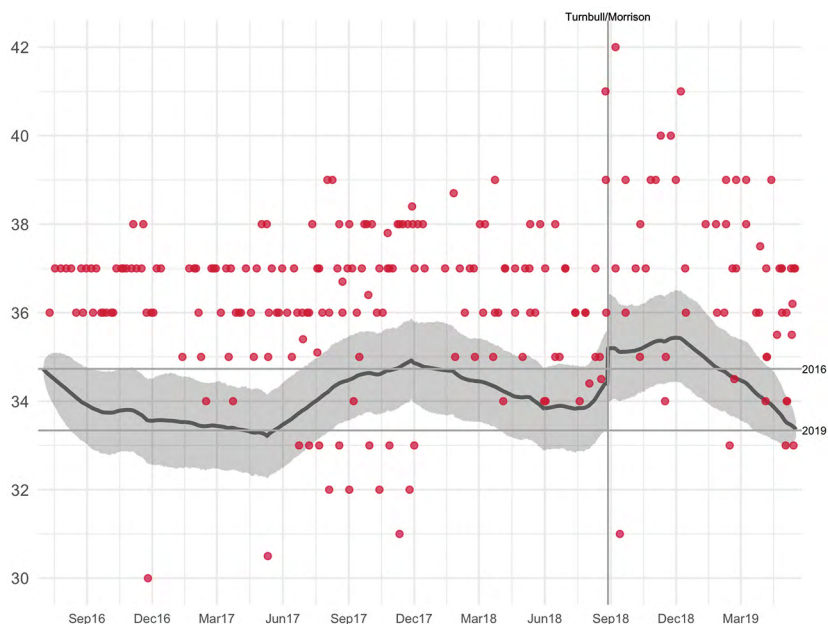


Figure 7.1 ALP first-preference voting intentions between the 2016 and 2019 federal elections

Notes: The darker, thicker line is the output of our model, connecting the daily estimates of the true state of public opinion, given our model and the observed 2016 and 2019 election results; the shaded region is a pointwise 95 per cent credibility interval around the daily estimates. Circles are poll results, used as input to our model.

Source: Authors' calculations.

Figure 7.1 also makes clear that the polls typically overestimated Labor's share of the vote; the bulk of the polls lie well above the estimated true state of public opinion. Observe also—with the benefit of hindsight—that: 1) at no point between the 2016 and 2019 elections was Labor's share of first preferences any stronger than in the 2016 election, even in the immediate aftermath of the Turnbull-to-Morrison leadership transition; and 2) there was a steep deterioration in Labor's first-preference vote share in the six months leading up to the 2019 election, of about 1.8 percentage points. This fall in support for Labor is more than twice the boost in Labor voting intentions that occurred at the time of the Turnbull–Morrison transition.

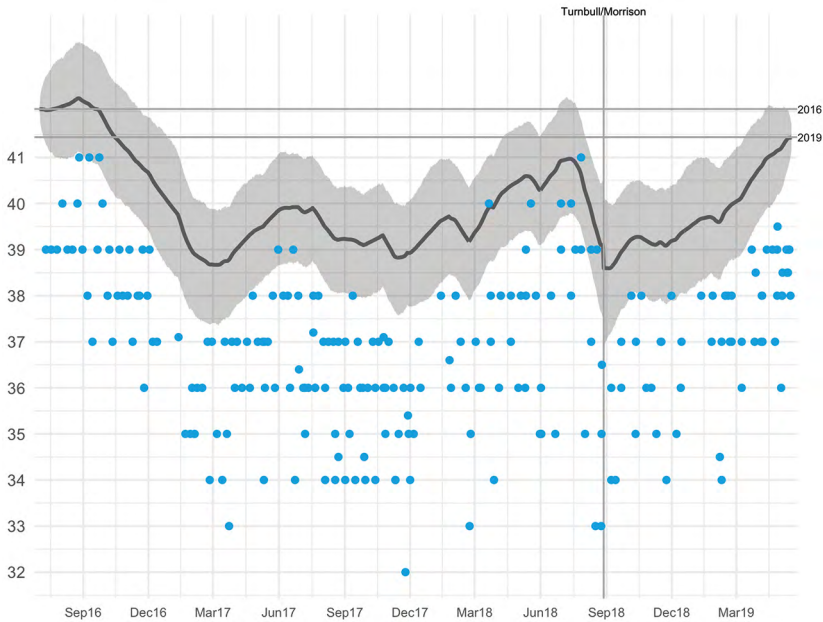


Figure 7.2 Coalition first-preference voting intentions between the 2016 and 2019 federal elections

Notes: The darker, thicker line is the output of our model, connecting the daily estimates of the true state of public opinion, given our model and the observed 2016 and 2019 election results; the shaded region is a pointwise 95 per cent credibility interval around the daily estimates. Circles are poll results, used as input to our model.

Source: Authors' calculations.

The equivalent graph for Coalition voting intentions is shown in Figure 7.2. Just as the polls tended to overstate the true level of Labor support, support for the Coalition was almost always understated, with almost all of the polls (the points in Figure 7.2) lying below the estimated true level of Coalition support.

Again, with the benefit of hindsight, we see that Coalition support did fall in the months immediately after the July 2016 election, stabilising around the 39 per cent level for most of 2017 (some 3 percentage points below the 2016 election result). A modest recovery in 2018 was reversed in the months preceding the Liberal leadership transition in August 2018. Over the course of Morrison's tenure, Coalition support recovered more than 2 percentage points, from 39 per cent to the 41.4 per cent result obtained at the May 2019 election.

Critically, not a single public poll reported Coalition first-preference support above 39.5 per cent in the period between Morrison becoming leader in August 2018 and the May 2019 election—far below the actual level of Coalition support during this period and indicative of the magnitude of poll bias in the 2016–19 cycle. It is understandable how this fed low expectations for the Coalition’s electoral prospects. In previous work, Jackman (1994) applied the Tufte (1973) log-vote share on log-seat share model to Australian federal elections between 1949 and 1993 and found the typical value for the Coalition to win 50 per cent of House of Representatives seats is 49.1 per cent of the two-party-preferred vote.⁶ Only for the briefest point after Morrison became the prime minister in August 2018 was the Coalition past this critical threshold.

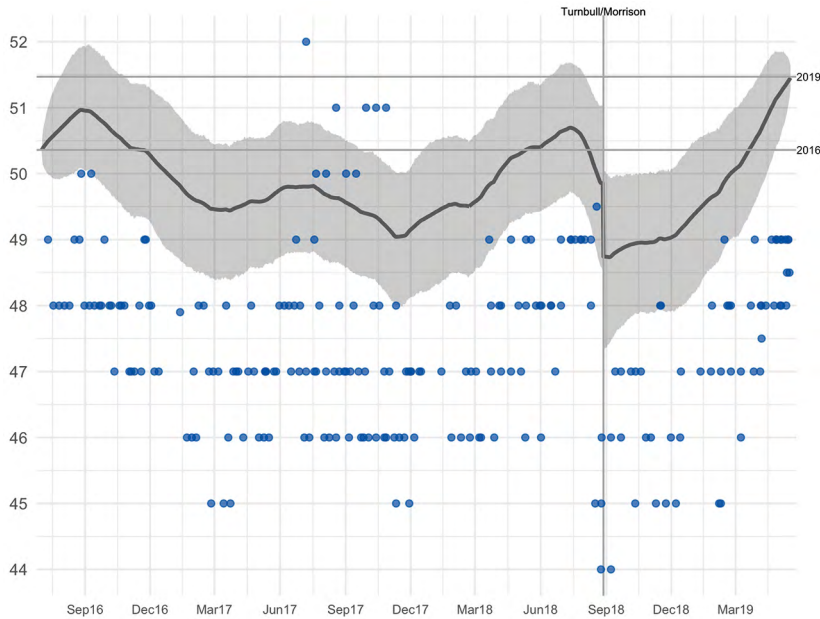


Figure 7.3 Coalition two-party-preferred voting intentions between the 2016 and 2019 federal elections

Notes: The darker, thicker line is the output of our model, connecting the daily estimates of the true state of public opinion, given our model and the observed 2016 and 2019 election results; the shaded region is a pointwise 95 per cent credibility interval around the daily estimates. Circles are poll results, used as input to our model.

Source: Authors’ calculations.

6 This was found to be invariant in Jackman (2009) and Mansillo and Evershed (2016) with the addition of more data.

A similar set of findings holds with even greater force for Coalition two-party-preferred support, shown in Figure 7.3. Few polls overestimated Coalition support; most estimated the Coalition to be underperforming relative to its 2016 two-party-preferred result. The polls did pick up a recovery in Coalition two-party-preferred support over the course of Morrison's tenure, but generally were biased by a considerable margin.

Pollster biases

Estimates of the biases specific to each polling organisation—the 'house effects' or δ parameters in our model—appear in Figures 7.4 and 7.5. Of the six polling organisations considered, four systematically overestimated Labor's first preferences: Morgan, ReachTEL, Essential and Newspoll. YouGov and Ipsos did not overestimate Labor's vote share but both organisations did underestimate Coalition first preferences, as did *every* polling organisation. The underestimates of Coalition first preferences are large in several cases, reaching almost 5 percentage points for YouGov, more than 4 percentage points for ReachTEL and 3 percentage points for Newspoll. Newspoll also overestimated support for the Australian Greens by 1 percentage point.

YouGov's large underestimate of Coalition first preferences is accompanied by a large overestimate of support for minor parties and Independents, by almost 5 percentage points. This bias is the largest polling error we observe for minor parties and Independents and is large relative to the actual level of support for minor parties and Independents in 2019 (14.8 per cent). Experimental methods may be a useful strategy for future research to identify whether such bias was a product of errors of representation or errors of measurement.

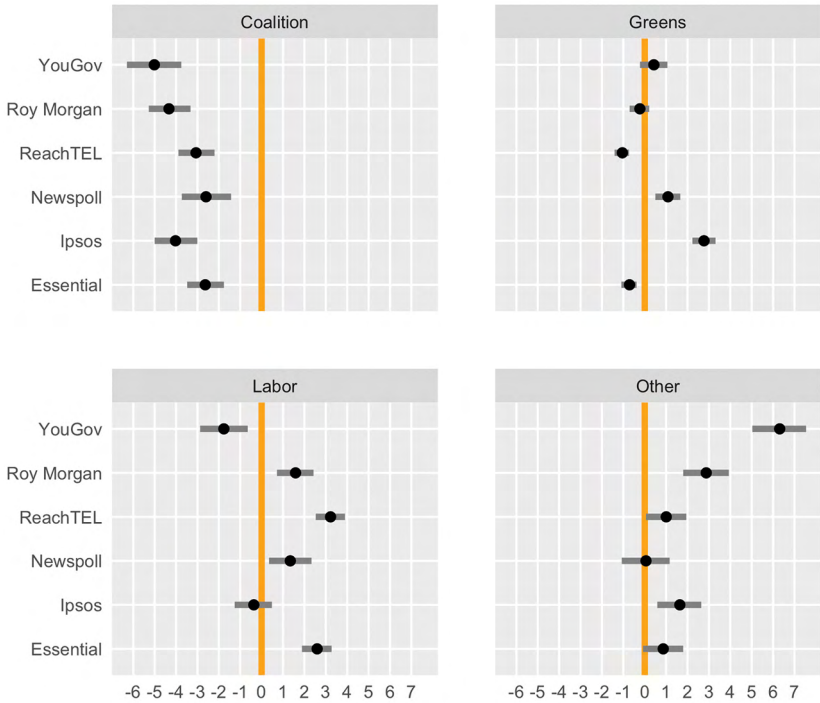


Figure 7.4 First-preference voting intentions house effects, 2016–19

Notes: Each point indicates an estimate of each polling organisation's bias with respect to the first preferences for the indicated party; horizontal bars cover 95 per cent credible intervals. Positive quantities are overestimates; negative quantities are underestimates.

Source: Authors' calculations.

Bias with respect to two-party-preferred voting intentions is shown in Figure 7.5. Consistent with Figure 7.3, all survey organisations, except YouGov, systematically underestimated the Coalition's share of the two-party-preferred vote, and by large margins: 2.5 to 3 percentage points. One possible source of Newspoll's underestimation of Coalition two-party-preferred support could be the assumption that only 60 per cent of PHON's preferences would flow to the Coalition. As it transpired, about 65 per cent of PHON preferences made their way to the Coalition from its 59 registered candidates, who secured 3.08 per cent of the national vote (see Chapters 4, 8 and 17, this volume); PHON did not run candidates in the other 92 seats. Even if PHON had fielded candidates in every seat and secured the same vote share, it would have secured 7.88 per cent of the vote (similar to where Newspoll and Essential placed the party in March 2019). This overestimate of PHON support—

and the minor underestimate of preference flows to the Coalition—explains a mere 0.4 percentage points of Newspan's two-party-preferred 2.5 percentage point underestimate.

We also note that, despite considerable variability in biases with respect to first-preference voting intentions (see Figure 7.5), there is far less variation in two-party-preferred bias across survey organisations. This finding is provocative, suggesting less dispersion across polling companies in two-party-preferred estimates relative to the variation in first-preference poll estimates. Two possible mechanisms for this are: 1) first-preference errors might fortuitously 'cancel out' when converted to two-party-preferred estimates, as appears to be the case with YouGov; or 2) pollsters use the mapping from first preferences to two-party-preferred—and assumptions about preference flows—to shade their results towards an industry consensus, which is an instance of the phenomenon known as 'herding'.

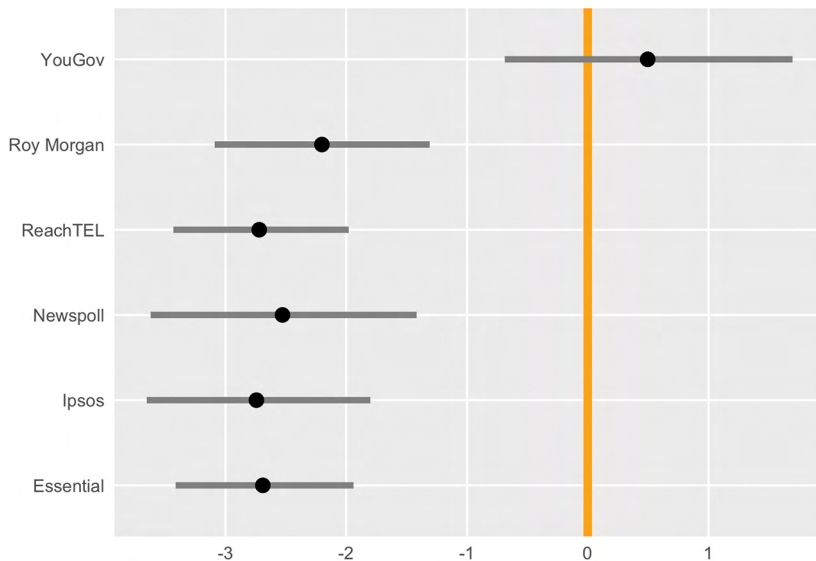


Figure 7.5 Coalition two-party-preferred voting intentions house effects, 2016–19

Notes: Each point indicates an estimate of each polling organisation's bias; horizontal bars cover 95 per cent credible intervals. Positive quantities are overestimates; negative quantities are underestimates.

Source: Authors' calculations.

Herding

Herding by pollsters has been observed at general elections in the United States (Panagopoulos et al. 2018), the United Kingdom (Sturgis et al. 2016), Canada (Whiteley 2016) and France (Evans and Ivaldi 2018). Herding is common in professions such as stock trading (Rülke et al. 2016) because of the incentive structures at play. There is safety in numbers: better to all be wrong together than risk being alone publishing a poll estimating an outcome wildly distant from the actual result. It requires a great deal of confidence for a pollster to take that sort of risk. Lonerger Research's Chris Lonerger admitted being embarrassed and jettisoning a poll he conducted because 'no one wants to release a poll that is wildly out of step ... we didn't want to be seen as having an inaccurate poll' (Koziol 2019).

Call it herding, harmonisation or looking over your shoulder; pollsters take stock of the competition and the available information each election cycle. This can be achieved by either not publishing the results of a poll conducted, such as in the Lonerger case, or 'deliberately adjust[ing] their methodology to produce certain results' (Prosser and Mellon 2018), which in effect suggests pollsters were playing the *weighting* game.

House effects in 2019 are large by historical standards

We gain historical perspective on 2019 polling errors by repeating this analysis with polls published prior to the 2007, 2010, 2013 and 2016 elections, utilising the same statistical model described above.⁷ Newspoll before July 2015 was conducted with live phone interviews and we distinguish this from the subsequent period when Newspoll was conducted with online panel and IVR interview and sampling methods. Nielsen polls, employing live phone interviews, ceased in July 2014.

⁷ The models include discontinuities for Kevin Rudd becoming the Labor leader (replacing Kim Beazley), Julia Gillard becoming prime minister (replacing Rudd), Rudd ousting Gillard to become prime minister again and the Abbott to Turnbull transition.

Comparing the 2019 two-party-preferred house effects with historical elections in Figure 7.6, it is clear that most pollsters, with the exception of YouGov, had very large misses. What is sobering is that, in 2010, every pollster managed to underestimate the Coalition's two-party-preferred vote. Newspoll was only 0.75 percentage points worse in 2019 than its 2010 performance. Essential and Roy Morgan in 2010 had estimates worse than any poll in 2019. The polls plainly 'got it wrong' across the board in 2019, as in 2010; biased estimates are not unique to 2019.

For the major parties' primary votes, we find little improvement in Figures 7.7 and 7.8. Every pollster underestimated the Coalition's primary vote and only YouGov, ReachTEL and Ipsos did so considerably more than the median house bias for the five elections. The last time Newspoll had not systematically underestimated the Coalition's primary vote was for the 2007 election. Ipsos's estimate for Labor deteriorated markedly between 2016 and 2019.

The general difficulty pollsters had in 2019 was underestimating the Coalition's primary vote—a similarity with 2010, when all pollsters systematically underestimated it, too. One avenue for future research would be to assess the available historical information on how the rate of undecided voters and their treatment affect the estimates of vote choice. Gelman et al. (2016) found during the 2012 US presidential election campaign that the swinging voter was mythical, with the swings observed in the polls related to the contacted sample's partisan enthusiasm and willingness to participate in surveys during the election campaign. Events in an election campaign can elicit survey bias as an artefact of sampling non-response, with their primes conditioning partisan response rates. It is more than plausible that a government's messy internal party turmoil could affect Australian partisans in a similar way to American partisans responding to their party candidate's performance in a presidential debate and associated media commentary. It is arguably more possible, since the breakdown of civil political norms—that is, the ousting of prime ministers—should elicit more ire than a comparatively poor television debate performance. If Labor partisan willingness to be polled was dampened after Julia Gillard became prime minister in 2010 and Coalition partisan willingness to be polled was dampened after Morrison became prime minister in 2018, there should be a relationship between the number of reported uncommitted respondents and the underestimates of the respective parties. This is not limited to leadership altercations; many other things could depress partisan spirits.

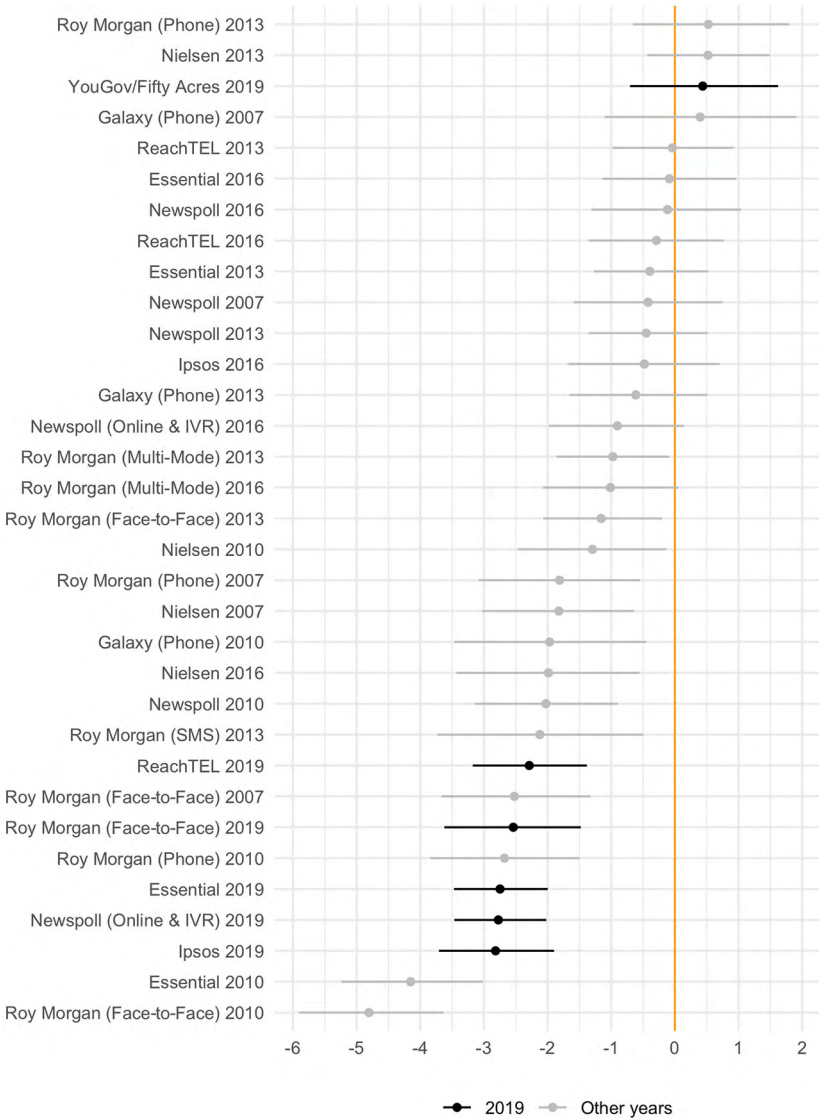


Figure 7.6 Coalition two-party-preferred voting intentions historical house effects, 2004-19

Notes: Each point indicates an estimate of each polling organisation's bias; horizontal bars cover 95 per cent credible intervals. Positive quantities are overestimates; negative quantities are underestimates.

Source: Authors' calculations.

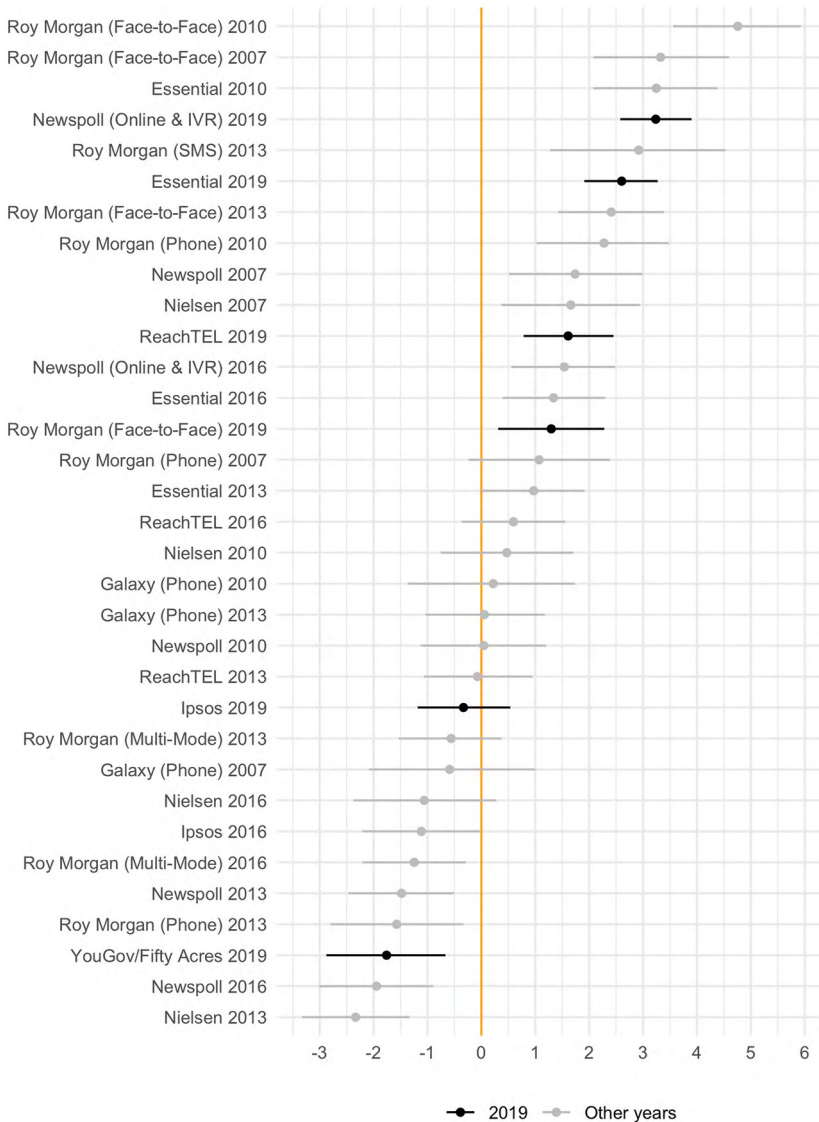


Figure 7.7 Labor first-preference voting intentions historical house effects, 2004–19

Notes: Each point indicates an estimate of each polling organisation's bias; horizontal bars cover 95 per cent credible intervals. Positive quantities are overestimates; negative quantities are underestimates.

Source: Authors' calculations.

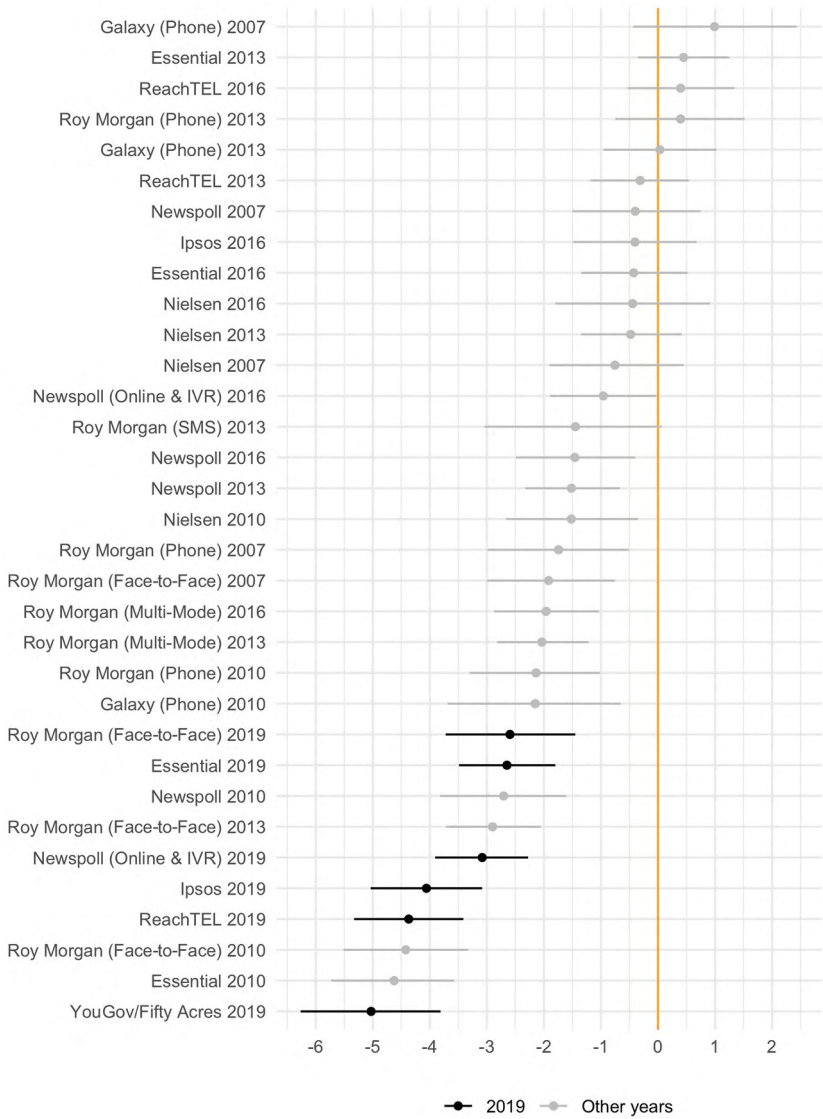


Figure 7.8 Coalition first-preference voting intentions historical house effects, 2004-19

Notes: Each point indicates an estimate of each polling organisation's bias; horizontal bars cover 95 per cent credible intervals. Positive quantities are overestimates; negative quantities are underestimates.

Source: Authors' calculations.

What is herding and how can we detect it?

As mentioned before, pollsters are susceptible to herding. In the case of election polling, pollsters who stray from the herd and meet an unexpected election result will suffer very real commercial consequences: vanishing clients. For pollster survival, it is safer to be in the herd than alone in the field.

Survey houses have a few choices about weighting once data have been collected and before publishing the results. Very few pollsters state which variables, such as age and gender, go into their weights, and none is open about how these weighting covariates are cut. For instance, age is a continuous variable that can be cut at different points on the population age structure into different groups with different sizes. Whether the data collected on these covariates are imbalanced (or what degree of imbalance exists), the public is not told. There are trade-offs that must be made when weighting. However, there comes a point when increasing weight complexity (and its variance), while reducing the estimate's bias, induces more error than the amount of bias removed. This opaque corner of polling should be made more transparent. The paucity of detail—such as what exactly is in the weighting scheme's 'secret sauce'—is a norm in Australian political polling, allowing pollsters to get away with shoddy workmanship undetected before an election.

Theory: Herding manifests as underdispersion

We leverage one of the oldest and most well-known theorems in statistics, the Central Limit Theorem (CLT), to test for herding in the polls immediately before the past five elections. Polls that herd should be underdispersed, displaying less variability than what we should see under the CLT. Suppose the true level of voting intentions is $\pi \in [0,1]$, a proportion between 0 and 1. Then via the CLT,⁸ a series of unbiased, simple random samples of size n will produce estimates of π denoted by $\hat{\pi}$ that follow a normal distribution with mean π and variance $V(\hat{\pi}) = \pi(1 - \pi) / n$.

8 Or actually, a special case of the CLT, the De Moivre–Laplace Theorem, dating to 1738 in the case of Abraham de Moivre and 1820 for Pierre-Simon Laplace's celebrated *Théorie analytique des probabilités*.

Putting to one side the bias that we have already analysed, we now compare the observed dispersion of the polls with the theoretically expected dispersion, given: 1) the polls' stated sample sizes, n_p ; and 2) an assumption about π , that voting intentions are generally unchanged from the days immediately preceding election day to the day of the election. In particular, we assume no change in voting intentions for d days prior to the election. Let $D(d)$ denote the set of polls fielded within d days of the election, with the standard deviation of those polls' estimates of π denoted as $s_d \equiv s_{D(d)}(\hat{\pi})$. We then repeat the following calculations over a range of values of d : 1) simulate poll results for each poll, $p \in D(d)$ —that is, $y_p^* \sim N(\pi, V_p)$, $V_p = \pi(1-\pi)/n_p$; 2) round each y_p to the same degree of precision as in the corresponding reported poll; and 3) compute s_d^* , the standard deviation of y_p^* . We then see—over many repetitions of the preceding steps—the rate at which $s_d^* > s_d(d)$. That is, is the variability of actual polls, D , smaller, larger or indistinguishable from what the CLT tells us we should see under simple random sampling? We apply this procedure to polls from the past five election campaigns.

Strong evidence of underdispersion in the Coalition's primary and two-party-preferred polling

The major party primary and two-party-preferred voting intentions are plotted in Figure 7.9. The dispersion of polls estimating the two-party-preferred vote in 2019 is well below what the CLT would require. This is indicated by the black line being below the shaded 95 per cent credibility intervals of where the optimum amount of dispersion would be given the estimates published and the sample sizes involved. The two-party-preferred estimates' dispersions were mostly within the optimum bands for the 2013 and 2016 elections but not for the 2010 and 2007 elections, which were overdispersed for the most part.⁹ There were large biases observed for these years (see Figure 7.6).

9 In 2013, pollsters transitioned from overdispersion to underdispersion, suggesting herding occurred as the election neared.

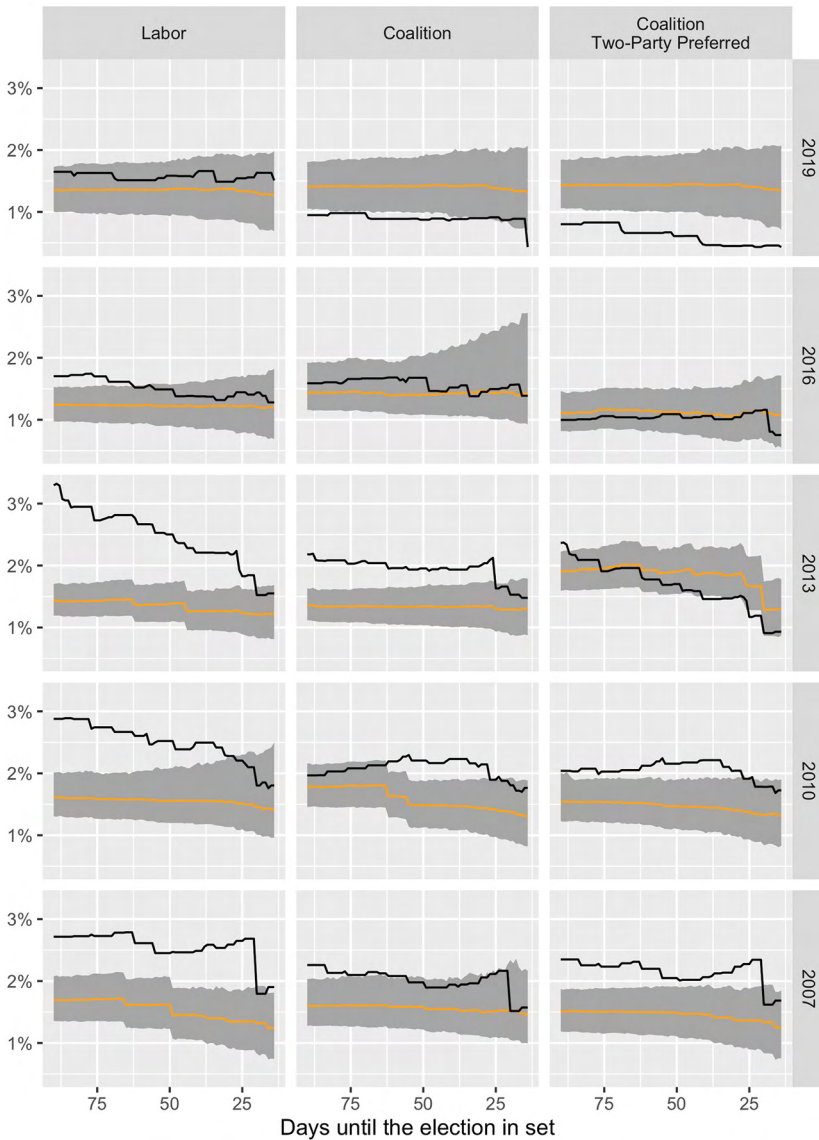


Figure 7.9 Dispersion before an election, 2004–19

Notes: For the polls conducted within each set of days until the election, the black line represents the observed dispersion in the polls and the orange line represents the theoretically expected dispersion. The poll dispersion coverage period sets take the value of 14 through to 90 days before each election.

Source: Authors' calculations.

For the 2019 major party votes, only the Coalition's primary vote was underdispersed, but less than the two-party-preferred, while estimates for Labor's support were as theoretically expected. There is a fair amount of overdispersion in historical primary votes; for the 2010 and 2013 elections, recent leadership changes made actual change to voter intentions plausible, especially for Labor partisans. The overdispersion observed might alternatively be suggestive of the non-response bias issues Gelman et al. (2016) identify, which pollsters, given their limited options, have attempted to paper over with extreme weights (Ansolabehere and Rivers 2013; Rosenbaum and Rubin 1983), inducing more variance and thereby error while reducing the observed bias in an imbalanced sample.

Conclusion: Polls and other disasters

The polls certainly missed the mark in 2019. The poll errors in 2019 were large and persistent across the 2016–19 cycle, shaping expectations about the election outcome with consequences for the way politics was conducted towards the election, spanning substantive matters such as policy debates and the parliament's agenda, through to media coverage and politicians' career decisions.

While large and enduring—and hence of immense political consequence—the magnitude of the poll errors in 2019 was not unprecedented. Our analysis of historical poll errors in Australian federal elections reveals that there were large poll errors in 2010. Any discussion of poll error in that election cycle was displaced with intense media interest in the fact that the 2010 election produced the first hung Commonwealth Parliament since 1940. Good poll performance in the 2013 and 2016 election cycles further helped to erase the prospect—and reality—of poll error from collective political memory. Hence, we have the alchemy of the 2019 election cycle, in the transformation of the uncertainty in random sampling, which produces estimates of public opinion into certainty.

Part of this alchemy was herding among the polls—a clustering of poll results reflecting the shared consensus that Labor would win the election. We provide clear evidence of underdispersion in the polls fielded in the lead-up to the 2019 election—a signature of herding. As unprofessional as herding may seem—and is—the commercial imperatives for it are clear. But the 2019 election cycle also makes clear the harm to the polity that

follows when herding and poll bias occur at once. Pollsters might not be able to do much about the latter. Bias only manifests *ex ante*, but herding happens along the way.

We also note the steady shift in Australian polling from live interviewer, random-digit dialling methods to the use of other survey modes such as pre-recruited panels of respondents taking surveys online, IVR or ‘robo polls’ and SMS or Facebook Messenger polls. Faster and less expensive ways of conducting surveys do present considerable challenges for data quality.

These newer survey modes are almost surely more susceptible than older methods to errors of coverage (are all members of the target population capable of being reached with a given sampling technology and its specific implementation?) and non-response bias (after contact, the decision about whether to participate in the survey is correlated with variables being measured by the survey). Remediating these biases usually involves weighting after data collection, with reference to a trusted data source such as the Australian census. But what variables are pollsters to weight on, and how aggressively to do this? And, until an election reveals the true state of voting intentions, survey companies are flying blind when deploying new survey modes. The temptation to look over one’s shoulder at the competition in this environment must be immense, as is the scope for doing so given the vast array of choices to be made when weighting data.

We invoke an adage applicable in many fields of science: sunlight is the best disinfectant. That is, there is a strong scientific case—and perhaps a commercial case as well—for greater transparency in how poll data are collected and weighted to ensure representativeness. We note that long-time pollster John Stirton (2019: 310–13) has called for urgent reform in this direction.

Given the errors in the 2019 election, there would seem to be a strong normative case, too, for Australia’s pollsters pulling back the curtain, even a little, to help us better understand why and how they generate the numbers they do. The quality of political discussion, the salience and content of the nation’s policy agenda—indeed, the health of Australian democracy—would gain from a commitment to this transparency around the polling of Australian public opinion.

References

- Ansolabehere, Stephen and Douglas Rivers. 2013. 'Cooperative Survey Research.' *Annual Review of Political Science* 16: 307–29. doi.org/10.1146/annurev-polisci-022811-160625.
- Brown, J. Andrew. 1992. 'The Major Effect: Changes in Party Leadership and Party Popularity'. *Parliamentary Affairs* 45(4): 545–64. doi.org/10.1093/oxfordjournals.pa.a052382.
- Burstein, Paul. 2003. 'The Impact of Public Opinion on Public Policy: A Review and An Agenda'. *Political Research Quarterly* 56(1): 29–40. doi.org/10.1177/106591290305600103.
- Ellis, Peter. 2017. 'State-space modelling of the Australian 2007 federal election'. *Free Range Statistics Blog*, 24 June. freerangestats.info/blog/2017/06/24/oz-polls-statespace.
- Evans, Jocelyn and Gilles Ivaldi. 2018. 'Forecasting and polling.' In *The 2017 French Presidential Elections*, edited by Joycelyn Evans and Gilles Ivaldi, 123–47. Cham, Switzerland: Springer. doi.org/10.1007/978-3-319-68327-0_6.
- Fuchs, Ester R., E. Scott Adler and Lincoln A. Mitchell. 2000. 'Win, Place, Show: Public Opinion Polls and Campaign Contributions in a New York City Election'. *Urban Affairs Review* 35(4): 479–501. doi.org/10.1177/10780870022184499.
- Gelman, Andrew, Sharad Goel, Douglas Rivers and David Rothschild. 2016. 'The Mythical Swing Voter'. *Quarterly Journal of Political Science* 11(1): 103–30. doi.org/10.1561/100.00015031.
- Hart, John. 1992. 'An Australian president? A comparative perspective'. In *Menzies to Keating: The Development of the Australian Prime Ministership*, edited by Patrick Weller, 183–201. Melbourne: Melbourne University Press.
- Jackman, Simon. 1994. 'Measuring Electoral Bias: Australia, 1949–93'. *British Journal of Political Science* 24(3): 319–57. doi.org/10.1017/s0007123400006888.
- Jackman, Simon. 2005. 'Pooling the Polls Over An Election Campaign'. *Australian Journal of Political Science* 40(4): 499–517. doi.org/10.1080/10361140500302472.
- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Hoboken, NJ: John Wiley & Sons. doi.org/10.1002/9780470686621.

- Jackman, Simon and Luke Mansillo. 2018. 'The campaign that wasn't: Tracking public opinion over the 44th Parliament and the 2016 election campaign'. In *Double Disillusion: The 2016 Australian Federal Election*, edited by Anika Gauja, Peter Chen, Jennifer Curtin and Juliet Pietsch, 133–56. Canberra: ANU Press. doi.org/10.22459/DD.04.2018.06.
- Koziol, Michael. 2019. "Embarrassed" pollster ripped up poll that showed Labor losing election'. *Sydney Morning Herald*, 9 June. www.smh.com.au/politics/federal/embarrassed-pollster-ripped-up-poll-that-showed-labor-losing-election-20190604-p51u9v.html.
- Mansillo, Luke and Nick Evershed. 2016. 'Pooling the polls: How we're tracking opinion polling for the Australian election'. *The Guardian*, 8 May. www.theguardian.com/australia-news/datablog/2016/may/08/pooling-the-polls-how-were-tracking-opinion-polling-for-the-australian-election.
- Mark the Ballot. 2017. 'Bayesian aggregation'. *Mark the Ballot Blog*. marktheballot.blogspot.com/p/the.html.
- Meyer, David S. and Debra C. Minkoff. 2004. 'Conceptualizing Political Opportunity'. *Social Forces* 82(4): 1457–92. doi.org/10.1353/sof.2004.0082.
- Panagopoulos, Costas, Kyle Endres and Aaron C. Weinschenk. 2018. 'Preelection Poll Accuracy and Bias in the 2016 US General Elections'. *Journal of Elections, Public Opinion and Parties* 28(2): 157–72. doi.org/10.1080/17457289.2018.1441850.
- Petersen, Jürgen, Sibylle Hardmeier and Bruno Wüest. 2008. 'Polls as Public–Politician Linkage: A Comparative Analysis of Poll Use and Roles of MPs in Parliamentary Debate'. *The Journal of Legislative Studies* 14(3): 315–38. doi.org/10.1080/13572330802259509.
- Pickup, Mark, J. Scott Matthews, Will Jennings, Robert Ford and Stephen D. Fisher. 2011. 'Why Did the Polls Overestimate Liberal Democrat Support? Sources of Polling Error in the 2010 British General Election'. *Journal of Elections, Public Opinion and Parties* 21(2): 179–209. doi.org/10.1080/17457289.2011.563309.
- Prosser, Christopher and Jonathan Mellon. 2018. 'The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses'. *Government and Opposition* 53(4): 757–90. doi.org/10.1017/gov.2018.7.
- Rosenbaum, Paul R. and Donald B. Rubin. 1983. 'The Central Role of the Propensity Score in Observational Studies for Causal Effects'. *Biometrika* 70(1): 41–55. doi.org/10.1093/biomet/70.1.41.

- Rülke, Jan-Christoph, Maria Silgoner and Julia Wörz. 2016. 'Herding Behavior of Business Cycle Forecasters'. *International Journal of Forecasting* 32(1): 23–33. doi.org/10.1016/j.ijforecast.2015.02.004.
- Stirton, John. 2019. 'Why the polls got it wrong'. In *From Turnbull to Morrison: The Trust Divide*, edited by Mark Evans, Michelle Grattan and Brendan McCaffrie, 307–13. Melbourne: Melbourne University Press.
- Stone, Walter J., Sarah A. Fulton, Cherie D. Maestas and L. Sandy Maisel. 2010. 'Incumbency Reconsidered: Prospects, Strategic Retirement, and Incumbent Quality in US House Elections'. *The Journal of Politics* 72(1): 178–90. doi.org/10.1017/s0022381609990557.
- Sturgis, Patrick, Nick Baker, Mario Callegaro, Stephen Fisher, Jane Green, Will Jennings, Jouni Kuha, Ben Lauderdale and Patten Smith. 2016. 'Report of the Inquiry into the 2015 British General Election Opinion Polls'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(3): 757–81. doi.org/10.1111/rssa.12329.
- Tufte, Edward R. 1973. 'The Relationship Between Seats and Votes in Two-Party Systems'. *American Political Science Review* 67(2): 540–54. doi.org/10.2307/1958782.
- Whiteley, Paul. 2016. 'Why Did the Polls Get It Wrong in the 2015 General Election? Evaluating the Inquiry into Pre-Election Polls'. *The Political Quarterly* 87(3): 437–42. doi.org/10.1111/1467-923X.12274.

This text is taken from *Morrison's Miracle: The 2019 Australian Federal Election*, edited by Anika Gauja, Marian Sawer and Marian Simms, published 2020 by ANU Press, The Australian National University, Canberra, Australia.

doi.org/10.22459/MM.2020.07