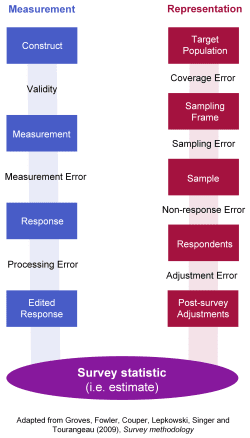
Like most people, I was surprised at yesterday’s result for the Australian federal election. But I was much *less* surprised than most people. The two-party-preferred vote for the ALP (currently 49.11%, with a lot of pre-polling still to count) came down at the bottom of my prediction interval but was still inside it; and they look to have a chance of making 68 or 69 seats, which was the lower range of my 80% prediction interval for number of seats. A 95% prediction interval would have comfortably included the result. While I thought that a narrow ALP (or ALP + Greens) win was the most likely outcome, in the last month I’d only given it a 40% to 50% chance and had always seen other outcomes as very possible.

The main reason I was less surprised than others was that my model discounted the polling companies’ reported ALP voting intention, based on their past performance.This involved the problematic assumption that the biases in these estimates didn’t change over time, an assumption I’d like to improve on if possible. But the treatment implicitly adopted by others, of assuming that there are no biases in the estimates at all, never seemed to me an improvement on assuming the same biases we’d seen before.

There’s a lot of comment out there about the “national polling failure” of “unprecedented” proportions. The perception is that Australian polls are much better for predictive purposes than in other countries; in the days leading up to the election I was in several Twitter discussions where others dismissed the large degrees of uncertainty in my model and were convinced the polls were going to be spot on. Now the results have come in, people are talking up how surprising the polling error is.

To put some perspective into things, here are the polls’ reported voting intention (two-party-preferred for the ALP) in the 28 days leading up to each of the past four elections:

The polls clearly did quite badly this time around, but not by an order of magnitude worse than previously. If like me you expected them to over-estimate the ALP vote by about the historical average, they were about 1.5 percentage points out more than expected. If you expected them to be on average right, then you were 2.5 percentage points out. For my way of interpreting surveys of people, even 2.5 percentage points out is pretty close. Certainly close enough that I think all the pundits who declared days ago that the election was over and the Coalition had lost should have a rethink not just about polling methodologies, but about how much they over-interpret survey results. After all, even if the polls had been right and the ALP gained 51% of the two-party-preferred vote, they did the same in 1998 and still failed to win due to vagaries of distribution of supporters between seats. So there was simply no excuse for pundits’ being as judgemental as they were leading up to the election.

Doubling any reported margin of error in your head is a good starting point for dealing with this. The reported margin of error for surveys is invariably based just on sampling error, because that’s easiest to quantify; other sources of uncertainty are going to be at least as important as sampling error in many situations. For example, it is clear that people who respond to surveys of voting intention, whether on internet or mobile phone, are going to be systematically different from people who don’t; and we don’t know how successful weighting is at addressing the systematic difference. In fact, generally we don’t know the details of weighting schemes used by polling firms in Australia. Anyone who wants to participate in the debate on polling methods should familiarise themselves with the concept of “total survey error”, as seen in this diagram (not my own, but my version of a standard one).

None of this lets the polling companies off the hook. Surveying is hard and in many ways getting harder, but that’s not a complete excuse. They clearly did worse than normal; they don’t publish enough of their sampling, processing and weighting methodology;proven by the visible “herding” of results (less variance than would be seen if they were genuine random samples). So I hope there will be some hard thinking and improvements to come from the polling firms. I just think that those improvements are less important than pundits understanding uncertainty better, and not pretending to more confidence than the facts deserve.

House effects

Kevin Bonham pointed out on Twitter that my forecasts assume that house effects (ie the structural under- or over-estimates of voting behaviour by individual polling firms) are the same over time. As surveying firms try to improve their methods, and as public behaviour changes, we would expect to see differences in the “biases” (speaking statistically) of each firm over time. Indeed, the same firm can publishe data simultaneously from surveys that collect data from different modes (eg phone versus online). I haven’t attempted to capture those nuances, never mind changes over time as the firms try to fix issues.

This is a definite limitation of my method, and one I’d like to fix for future elections. I also reflected on this issue in my forecasts of the New Zealand election, I concluded from the minimal effects seen there that the house effects changed surprisingly slowly so I could ignore the phenomenon and judge each firm on its overall results. That might not be the case for Australian polls, so I should at least try building in changing house effects over time. Unfortunately this will have the impact of increasing even further the uncertainty in translating polling numbers to election results.

While I don’t have time for that today, I do have time to at least explore the most likely problem candidate, Roy Morgan. My Bayesian state space model finds a strong over-estimate of the ALP vote in Roy Morgan, comparing it to the other polling firms and to actual election outcomes as far back as I have data. Bonham suggests that more recently, Roy Morgan polls if anything now under-estimate the ALP vote.

Here’s the house effects shown by my current model:

Let’s look at the voting intention reported by the four current main pollsters by election year. Just looking at these density curves certainly does suggest that this time around, Roy Morgan has reversed and is now underestimating ALP vote; the purple density is noticeably to the left of the others:

Compare to how far the yellow is to the right of the crowd in the first plot. Sorry for not matching the colours; tempus fugit.

However, I’m not going to draw firm conclusions from this. We can see that Roy Morgan have published relatively few polls this election cycle, and they have mostly been relatively late:

The ALP support now is less than six months ago, so the fact that Roy Morgan’s average reported intended vote for ALP is less than the other firms is not as convincing as it might be thought. There are just too few data points to be sure. In effect, my model judges Roy Morgan on their performance over the full period data are available, and overall that data shows an overestimate of ALP vote. It’s possible they’ve fixed the problem, but they’ll need to get quite a few runs on the board to overcome their bad history on this.

Having said that, I think there *is* enough data (and other more theoretical considerations) to be confident that house effects in general change over time, so when I can I will address it. I can also say that I think this issue is not important in my current forecasts; excluding Roy Morgan from my data altogether does not lead to a noticeably different result. The real driver of my less-certain-than-normal-pundits’ outlook for an ALP win is from the uncertainty my model exists, particularly in translating an uncertain two-party-preferred swing into an even more uncertain set of seat changes.

Here’s the R code for the above examination of Roy Morgan’s house effect:

*Post continues after code extract*

library(ozfedelect)

library(tidyverse)

library(lubridate)

*#-------------Has Roy Morgan fixed the pro-ALP overestimate----------------*

d **<-** ozpolls **%>%**

filter(firm **%in%** **c**("Election result", "Roy Morgan", "Newspoll", "Galaxy", "Ipsos", "Essential")) **%>%**

filter(preference\_type **==** "Two-party-preferred") **%>%**

filter(party **==** "ALP")

d **%>%**

filter(firm **!=** "Election result") **%>%**

ggplot(aes(x **=** intended\_vote, colour **=** firm)) **+**

geom\_density() **+**

facet\_wrap(**~**election\_year) **+**

labs(x **=** "Intended two-party-preferred vote for the ALP") **+**

ggtitle("Differences in polling firms' estimates of ALP vote by election year")

d **%>%**

filter(firm **!=** "Election result") **%>%**

ggplot(aes(x **=** mid\_date, y **=** intended\_vote, colour **=** firm)) **+**

geom\_point() **+**

facet\_wrap(**~**firm) **+**

ggtitle("Roy Morgan polls for the 2019 election are few and late",

"We can't conclude the ALP overestimate is fixed from these data points.") **+**

labs(x **=** "Survey mid date",

y **=** "Intended two-party-preferred vote for the ALP")

Herding

Another issue I and others have been pondering is the low level of variation between polls. Of the last 16 published national polls, 7 have placed the ALP two-party-preferred vote at 51%, 7 at 52%, and one each at 52.5% and 53%. As Mark the Ballot points out, that’s less variation than you’d expect if they really were based on random samples. The most likely explanation for this underdispersion is that polling firms are introducing errors in their processing that push their polls towards a consensus view. This is a known phenomenon around the world. The net effect is to reduce the usefulness of polls, by bringing them towards a pundit-based folly of the crowds consensus.

It’s not clear if the errors are being introduced accidentally (for example, by unconsciously giving more scrutiny and checks to “suspicious” points that break with the consensus view) or more deliberately through some kind of smoothing or regularisation. Either way, it means that the polling information is worth less than it would be with a more straightforward production process. But it’s something we’re going to have to live with, and just note yet another reason for being less confident in what we think is going to happen on Saturday - whatever it is that we think!

Here’s a comparison of what we’ve seen in the reported polls versus what we’d expect in actual random samples. I’ve used Mark the Ballot’s assumption of sample sizes of 1,000; in practice, some samples are higher and some smaller, but as we are ignoring non-sampling error I am happy to err a little on the small side for sampling. Also, note that the reported results are (usually) rounded in ways that impact on the distribution of overall results in complicated but unimportant ways. To derive the results below I’ve mimicked that rounding process in 100,000 simulated samples of 25 polls, all of them rounded to the nearest percent:

Here’s the code for that:

*Post continues after code extract*

*#----------How bad is the underdispersion---------*

set.seed(123)

d2 **<-** d **%>%**

filter(year(mid\_date) **==** 2019) **%>%**

select(mid\_date,

observed\_polls **=** intended\_vote)

n **<-** nrow(d2)

reps **<-** 100000

sims **<-** matrix(**round**(rbinom(n **\*** reps, 1000, mean(d2**$**observed\_polls) **/** 100) **/** 10),

ncol **=** n, nrow **=** reps)

d4 **<-** tibble(standard\_deviations **=** apply(sims, 1, sd))

d4 **%>%**

ggplot(aes(x **=** standard\_deviations)) **+**

geom\_density(fill **=** "grey", alpha **=** 0.5) **+**

geom\_vline(xintercept **=** sd(d2**$**observed\_polls), colour **=** "red") **+**

annotate("text", x **=** 0.45, y **=** 0.5, hjust **=** 0, label **=** paste0(

"Only ", **round**(mean(d4 **<** sd(d2**$**observed\_polls)), 3),

" of simulated poll\nsequences have a standard\ndeviation less than has been\nobserved in the ",

n, " polls in 2019."

)) **+**

annotate("text", x **=** 1, y **=** 1.5, colour **=** "red", label **=** "Observed", hjust **=** 1) **+**

labs(x **=** paste0("Standard deviations of simulated sequences of ",

n,

" surveys with 1,000 respondents, with survey results rounded")) **+**

ggtitle("The polls published in 2019 vary somewhat less than they should if random",

"The amount of variance is surprisingly low, but not impossibly so.")

For the record - forecasts as at 15 May 2019

I find it interesting that some people react to forecasts as though they are normative. I got a number of retweets from rightists - including Australians who proudly wear their Make America Great Again caps to early voting and tweet about, thereby donating cultural scientists all sorts of interesting material - for my most recent update on my election predictions. And a few left-leaning sceptics. Let me state my own intent to keep my own political preferences (which are strong) out of this; I want to avoid the problems that lead to herding of opinion pollsters, for one thing. While I can understand people who don’t like to see forecasts that their side is less likely to win they thought (or do like to see forecasts that their own party may have a chance after all), I think we should endeavour to avoid our preferences clouding our judgement.

Here’s my forecasts - with big chunks of uncertainty - as at 15 May 2019:

My forecast for the Australian 2019 federal election on 18 May 2019 (with the latest polling data available at 14 May 2019) is a narrow ALP win in their own right (36% chance) or with the presumed support of one or two Green members (47% chance in total). However, a range of other options are very much in play. An 80% prediction interval for the ALP in the House of Representatives is 68 to 80 seats.

The Liberal-National Coalition is likely to end up with between 67 and 78 seats. An outright win is a distinct possibility, and a win with the help of independents or minor parties more so. The Greens are most likely to end up retaining their single seat, but have a fair chance of picking up a second. In total, parties other than the ALP and Coalition are likely to end up with three to six seats.

The range of possibilities with regard to numbers of seats:

The predicted two-party-preferred vote:

The R packages that make this possible

thankr**::**shoulders() **%>%**

mutate(maintainer **=** str\_squish(gsub("<.+>", "", maintainer)),

maintainer **=** ifelse(maintainer **==** "R-core", "R Core Team", maintainer)) **%>%**

group\_by(maintainer) **%>%**

summarise(`Number packages` **=** **sum**(no\_packages),

packages **=** paste(packages, collapse **=** ", ")) **%>%**

arrange(desc(`Number packages`)) **%>%**

knitr**::**kable() **%>%**

clipr**::**write\_clip()

That’s all for today. Here’s the R code for that graphic of surveyed voting intention versus actual results. The data are all collected in ozfedelect R package.

library(ozfedelect)

library(tidyverse)

# Get the actual election dates

elections <- ozpolls %>%

filter(firm == "Election result" &

preference\_type == "Two-party-preferred" &

party == "ALP") %>%

mutate(election\_year = election\_year - 3) %>%

select(election\_year,

election\_date = mid\_date,

actual\_vote = intended\_vote) %>%

rbind(tibble(

election\_year = 2019,

election\_date = as.Date("2019-05-18"),

actual\_vote = 49.11

))

# draw the graphic

ozpolls %>%

filter(firm != "Election result" &

preference\_type == "Two-party-preferred" &

party == "ALP") %>%

left\_join(elections, by = "election\_year") %>%

mutate(days\_away\_election = as.numeric(election\_date - mid\_date),

alp\_overestimate = intended\_vote - actual\_vote) %>%

filter(days\_away\_election < 28 & days\_away\_election > 0) %>%

ggplot(aes(x = mid\_date, y = alp\_overestimate)) +

theme(panel.grid = element\_blank()) +

geom\_hline(yintercept = 0, colour = "grey") +

geom\_smooth(se = FALSE, method = "gam") +

geom\_point(aes(colour = firm)) +

facet\_wrap(~election\_year, scales = "free\_x", nrow = 1) +

ggtitle("2019 polls overestimated the ALP result somewhat more than usual",

"Showing polls in the 28 days up to the election. 2019 result is provisional, at 16 hours.") +

labs(x = "Polling date",

y = "Survey voting 2pp intention minus result on the day",

colour = "",

caption = "freerangestats.info")