As the analysis I’ve seen used nzelect R package, this motivated me to update it for recent polls.

**The nzelect update**

nzelect hasn’t been updated on CRAN for some time, because about a year ago I made some major changes to the data model for the historical election results by voting place and I haven’t been able to complete testing and stabilisation of the result. I do hope to do this some time in the next few months. In the meantime, the version on GitHub has the current polling data, and I intend to keep it current. Political polls are very thin on the ground these days for New Zealand, so that’s not too big an ask! I’ve now spent a bit of time tidying it up and adding the three most recent polls, which I’d previously neglected.

Let’s start with the basics. One of the reasons I first put the package together several years ago was to help facilitate analysis of the relatively long run in political opinion. I wanted to lift up analysis from over-interpretation of the last few noisy data points. Here’s the expressed voting intention of New Zealanders for the four currently largest parties in Parliament over time.

It’s striking, but unsurprising, how the Greens and New Zealand First support has collapsed since coming into government with Labour (or just before, in the case of the Greens and their disappointing 2017 election campaign). The junior party in a coalition often suffers, as attention and kudos goes to the leaders of the large party (in this case, Prime Minister Jacinda Ardern) and the smaller parties’ own base deals with the realities and compromises of being in government.

The other interesting (and disappointing, for statisticians and political scientists) observation from this chart is how obviously the number of polls has decreased. Topic for another day (or more likely, someone else to write about).

Here’s the R code for that chart:

# load packages needed including for rest of session

library(nzelect)

library(tidyverse)

library(scales)

library(lubridate)

library(gridExtra)

library(broom)

library(mgcv)

#-----------------------------polling overview---------------------

p1 <- polls %>%

filter(MidDate > as.Date("2014-11-20") & !is.na(VotingIntention)) %>%

filter(Party %in% c("National", "Labour")) %>%

mutate(Party = fct\_reorder(Party, VotingIntention, .desc = TRUE),

Party = fct\_drop(Party)) %>%

ggplot(aes(x = MidDate, y = VotingIntention, colour = Party, linetype = Pollster)) +

geom\_line(alpha = 0.5) +

geom\_point(aes(shape = Pollster)) +

geom\_smooth(aes(group = Party), se = FALSE, colour = "grey15", span = .4) +

scale\_colour\_manual(values = parties\_v, guide = "none") +

scale\_y\_continuous("Voting intention", label = percent) +

scale\_x\_date("") +

facet\_wrap(~Party, scales = "fixed") +

theme(panel.grid.minor = element\_blank(),

legend.position = "none")

p2 <- polls %>%

filter(MidDate > as.Date("2014-11-20") & !is.na(VotingIntention)) %>%

filter(Party %in% c("Green", "NZ First")) %>%

mutate(Party = fct\_reorder(Party, VotingIntention, .desc = TRUE),

Party = fct\_drop(Party)) %>%

ggplot(aes(x = MidDate, y = VotingIntention, colour = Party, linetype = Pollster)) +

geom\_line(alpha = 0.5) +

geom\_point(aes(shape = Pollster)) +

geom\_smooth(aes(group = Party), se = FALSE, colour = "grey15", span = .4) +

scale\_colour\_manual(values = parties\_v, guide = "none") +

scale\_y\_continuous("Voting intention", label = percent) +

scale\_x\_date("") +

facet\_wrap(~Party, scales = "fixed") +

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

grid.arrange(p1, p2)

There’s obvious interest for supporters of the left-of-centre parties in the combined vote for Labour and the Greens. That suggests the importance of this chart:

(Apologies for red-green colour blind people in the use of these party colours; the Greens are the lowest of the three lines and Labour the middle.)

It’s clear that to a significant degree electoral support for the two is substitutable, with the green and red lines moving in scissors-like counter directions at several key times since 2010, with the last 24 months just the most dramatic example.

It’s also clear that the combined support for the centre-left in New Zealand is pretty strong, recovered to a point it hasn’t been since several years before the end of the government led by Helen Clarke in the 2000s.

Here’s the code for that chart. Note how I use the dates of elections to make a simple data frame of who is in power when, for the background rectangles; and leverage the parties\_v vector of colours in nzelect to allocate the official party colours to both the parties’ lines and to the background fill.

#----------------long run-------------

elections <- polls %>%

filter(Pollster == "Election result") %>%

distinct(MidDate) %>%

rename(start\_date = MidDate) %>%

mutate(end\_date = lead(start\_date, default = as.Date("2020-10-01")),

pm = c("Labour", "Labour", "National", "National", "National", "Labour"))

p3 <- polls %>%

as\_tibble %>%

filter(!is.na(VotingIntention)) %>%

filter(Party %in% c("Green", "Labour")) %>%

select(Party, MidDate, VotingIntention, Pollster) %>%

spread(Party, VotingIntention) %>%

mutate(Combined = Labour + Green) %>%

gather(Party, VotingIntention, -MidDate, -Pollster) %>%

mutate(Party = fct\_relevel(Party, c("Combined", "Labour"))) %>%

ggplot() +

geom\_rect(data = elections, ymin = -Inf, ymax = Inf, alpha = 0.1,

aes(xmin = start\_date, xmax = end\_date, fill = pm)) +

geom\_line(aes(x = MidDate, y = VotingIntention, colour = Party)) +

scale\_colour\_manual(values = c(parties\_v, "Combined" = "black")) +

scale\_fill\_manual(values = parties\_v) +

labs(x = "Survey date", y = "Voting intention",

colour = "Surveyed voting intention:", fill = "Prime Minister's party:") +

scale\_y\_continuous(label = percent\_format(accuracy = 2)) +

ggtitle("Voting intention in New Zealand over a longer period than usually presented",

"Labour, Greens, and combined")

**Seasonality**

Now, on to the question of seasonality. I don’t have much to add to the analysis of David Hood on Twitter and Thomas Lumley on StatsChat; I basically agree with their conclusions. I have the advantage of a few more polls because of the update to nzelect this morning.

Here is the expressed intended vote for the party of the Prime Minister over time:

The blue line for the National Party is higher than the equivalent for Labour Prime Ministers because National has tended to form a larger proportion of its governing coalitions than Labour in this time period. I could (and probably should) have added the intended vote for all parties currently in coalition government, but this is actually a pretty complicated thing to do so I’ve gone for the simpler approach for now. So long as we don’t make simplistic comparisons forgetting that New Zealand has a proportional representation system, that is ok for our purposes.

We obviously can’t tell anything from this chart about the seasonality; there are two many data points and too much noise. Actually, one thing we can say for sure is that seasonality isn’t strong. For very seasonal data such as tourist numbers, the seasonality would be obvious even in a chart like this.

To see if there is a subtle seasonality effect, I tried modelling voting intention for the Prime Minister’s party on the month of the year, controlling for the party in power, a smooth trend over time, and whether or not it is an election month (otherwise September and November, with five of the six elections in this period, would certainly cloud the data). I used (as I nearly always do in this situation) Simon Woods excellent mgcv R package.

Having done that, we can approximate confidence intervals for the impact on voting preference of the month of the year. The next chart shows those estimates:

As the title says, it’s weak evidence of a weak effect, which might be around half a percentage point more positive for the Prime Minister’s party in the summer months than it is in June. Or it might be more than that, or even negative.

Here’s the code for that model and the last two charts:

#------------------------seasonality of govt support------------------

# data frame of voting intention for the lead party of the government

d <- polls %>%

as\_tibble() %>%

left\_join(elections, by = c("MidDate" = "start\_date")) %>%

select(-(WikipediaDates:EndDate)) %>%

fill(pm, end\_date) %>%

# limit to the party that has the Prime Minister:

filter(Party == pm) %>%

mutate(survey\_month = as.character(month(MidDate, label = TRUE)),

survey\_month = fct\_relevel(survey\_month, "Jun"),

election\_month = (month(MidDate) == month(end\_date) & year(MidDate) == year(end\_date)))

ggplot(d) +

geom\_rect(data = elections, ymin = -Inf, ymax = Inf, alpha = 0.1,

aes(xmin = start\_date, xmax = end\_date, fill = pm)) +

geom\_line(aes(x = MidDate, y = VotingIntention, colour = Party, group = end\_date)) +

scale\_colour\_manual(values = c(parties\_v, "Combined" = "black")) +

scale\_fill\_manual(values = parties\_v) +

labs(x = "Survey date", y = "Voting intention for PM's party",

colour = "Surveyed voting intention:", fill = "Prime Minister's party:") +

scale\_y\_continuous(label = percent\_format(accuracy = 2)) +

ggtitle("Voting intention for the Prime Minister's party")

# model of vote for govt's lead party, controlling for trend (the smooth term) and for which

# party is in government, looking for effects from month of year.

mod <- gam(VotingIntention ~ election\_month + survey\_month + Party + s(as.numeric(MidDate)), data = d)

anova(mod) # yes, the survey\_month is 'statistically significant'

tibble(variable = names(coef(mod)),

estimate = coef(mod),

se = summary(mod)$se) %>%

# very approximate confidence intervals:

mutate(lower = estimate - se \* 1.96,

upper = estimate + se \* 1.96) %>%

filter(grepl("month", variable)) %>%

mutate(survey\_month = gsub("survey\_month", "", variable)) %>%

mutate(survey\_month = factor(survey\_month,

levels = c("election\_monthTRUE", as.character(month(c(7:12, 1:6), label = TRUE))))) %>%

ggplot(aes(x = lower, xend = upper, y = survey\_month, yend = survey\_month)) +

geom\_vline(xintercept = 0, colour = "red") +

geom\_segment(size = 3, colour = "steelblue", alpha = 0.7) +

geom\_point(aes(x = estimate)) +

scale\_x\_continuous(label = percent) +

labs(x = "Estimated impact (in percentage points) on intended vote for PM's party, relative to June",

y = "Survey month") +

ggtitle("Seasonality of voting preference for the New Zealand PM's party?",

"Weak evidence of a very weak pro-government effect from late Spring to early Autumn")

It’s very motivating to see others using the nzelect package. Please tag me in Twitter, or let me know some other way, if you use this and it will encourage me for further enhancements, and to get the new version with better historical data onto CRAN!

**Thanksgiving**

I’m going to try to get into the habit of this at the end of each blog post. Without the hard work, innovation and sheer smarts of the open source community my blog (and many much more important things!) wouldn’t be possible. Here are just those in the R world whose code I used in this session (not all of it made it into the excerpt above, but that’s all the more reason to give thanks below).

thankr::shoulders() %>% knitr::kable() %>% clipr::write\_clip()