

Who we are affects how we vote

Rob J Hyndman and Di Cook

1 Introduction

We often hear commentary about voting patterns — people in their 20s without kids are more likely to be left wing, migrants are more conservative, wealthier people tend to favour the conservative parties, and so on. These might be built on myth and stereotypes. There is open data available from each election and from regular national censuses; this data can be used to examine just what are the indicators for voting tendency.

Voting tendencies can change over time. For example, if wealth was a good predictor of conservative voting in the 2001 election, is it still a good predictor in 2019? Australia has changed in many ways over the last two decades. Rising house prices, country-wide improvements in education, an ageing population, and a decline in religious affiliation, are just some of the ways we have changed. At the same time, political power has moved back and forth between the two major parties. How much can we attribute changes in political power to changes in who we are?

2 Census and electoral data

The Census provides data on electoral socio-demographics, and vote counts in each electorate can be obtained from Australian federal elections. However, joining these two data sources is difficult because the Censuses are not held at the same time as the elections. Between 2001 and 2016 there were six elections and four Censuses, as shown in the timeline below.

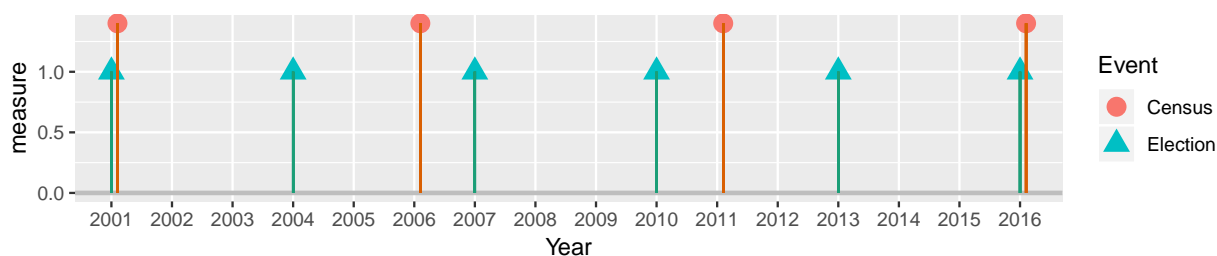


Figure 1: *Timeline of Australian elections and Censuses. Censuses happen regularly, elections not quite regularly, which means data needs to be interpolated by time.*

Not only can an electorate change between the last Census and an election, but even the electorate boundaries can change. Some electorates can disappear altogether and new electorates can arise. Electoral boundaries are redistributed regularly by the AEC, meaning that only in the years where both a Census and an election occur are all boundaries likely to match — the case for the 2001 and 2016 elections. So we first had to estimate what the socio-demographic characteristics of an electorate would have been at the time of each election using a complicated method of interpolation over time and geography. This method uses Census information from

both before and after the election of interest, and information from neighbouring electorates when boundaries have changed.

3 2PP Modelling

A simple way to measure voting patterns is to consider the two-party preferred (2PP) vote, which is based on the tally of preferences for the Labor and Liberal candidates, ignoring all other candidates. By convention, this is recorded as a percentage preference in favour of the Liberal party — for example, a 2PP value of 45% indicates that 45% of voters ranked the Liberal candidate higher than the Labor candidate, while the remaining 55% ranked them in the reverse order.

We consider how various socio-demographic variables obtained from Census data can be used to explain the 2PP values for each of the 150 electorates in each of the federal elections between 2001 and 2016.

Many of the socio-demographic variables have changing scales over the years. For example, inflation-adjusted median rental prices increased across almost all electorates, with median rent of 200 dollars per week placing an electorate in the 90th percentile in 2001, but only the 30th percentile in 2016. In order for socio-demographic effects to be comparable across years, all socio-demographic variables were standardized.

There are dozens of socio-demographic variables available in the Censuses, with many variables representing similar information about an electorate. So we combined some variables to avoid redundant information. For example, our “Incomes” variable is a combination of median personal income, household income and family income.

We are using data at the electorate level. The results do not directly address the voting intentions of individuals.

Each election was modelled separately, to allow us to see any changes over time, and to account for changing electorate boundaries. In this article, we highlight the variables with the strongest relationship to the two-party preferred vote, or which have had substantial changes over time. The full analysis is available at <https://robjhyndman.com/publications/elections/>.

Country-wide trend

First, we show the estimated two-party preferred vote for an “average” electorate.

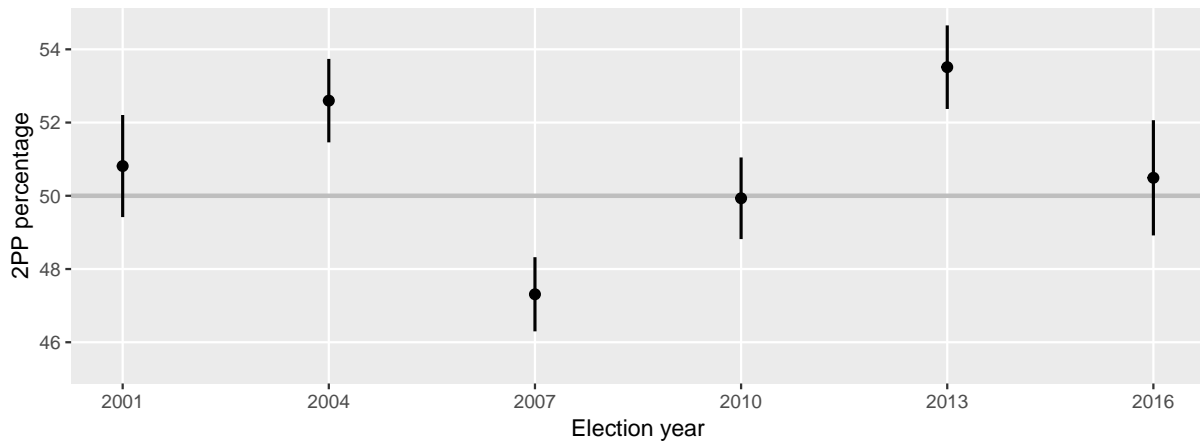


Figure 2: *Estimated two-party preferred vote for an electorate with average characteristics at each election.*

This shows that the baseline of party preference has varied over the elections, with the biggest swing occurring in the 2007 election where the mean electorate shifted more than five percentage points in favour of the Labor party. The dots represent the estimated 2PP value, and the lines indicate a 95% confidence interval providing a guide to the uncertainty in the estimate. Where the vertical lines corresponding to each election cross the 50% horizontal line, the 2PP vote for an average electorate is not statistically distinguishable from 50% for that election.

Age

Regions comprising more older people are often believed to be more conservative, and indeed we found that electorates with a higher median age are more likely to support the Liberal party — although this effect is significant only in 2004 and 2007.

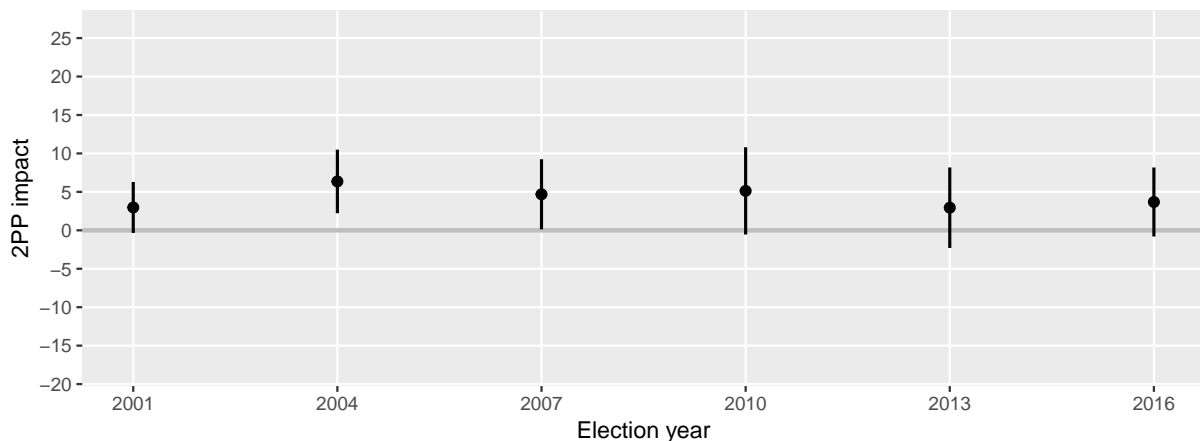


Figure 3: *Estimated impact of electorate median age on the two-party preferred vote.*

This graph shows the effect of median age on the 2PP vote. The vertical scale can be interpreted as a score measuring how much the 2PP vote differs between an average electorate, and an electorate with median age in the top 10% of all electorates (with all other socio-demographic variables unchanged). The reverse is true for young electorates — the score can be interpreted

as how much the 2PP vote would be reduced if an average electorate kept all socio-demographic variables fixed but median age was reduced to be in the bottom 10% of all electorates.

Income

Typically the Labor party campaigns on more progressive policies, which often include tax reform that adversely affects higher income earners, and more generous social assistance programs. Perhaps it is due to these policies that higher income electorates appear more likely to support the Liberal party, as the Incomes factor has a positive effect on Liberal preference. The effect also seems to have been increasing over time.

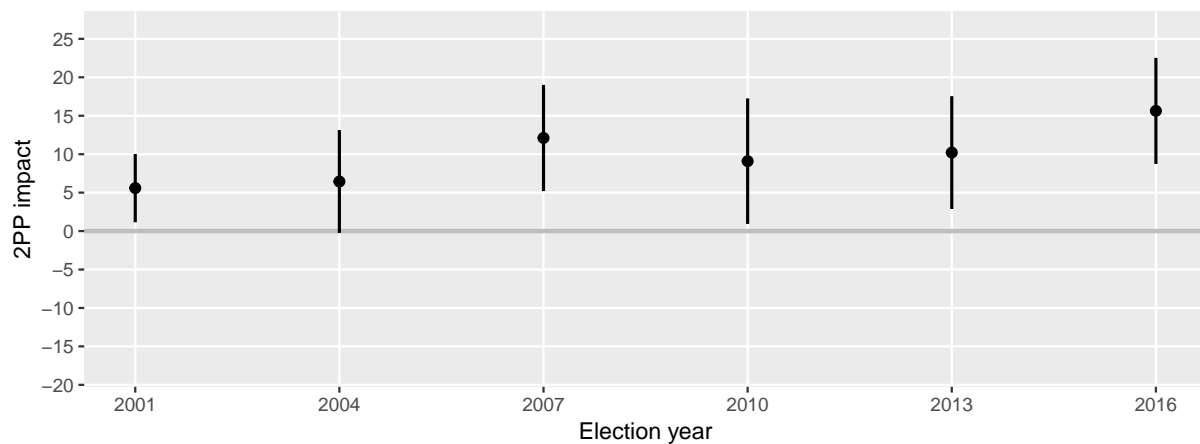


Figure 4: *Estimated impact of electorate income on the two-party preferred vote.*

As before, the vertical scale can be interpreted as a score measuring how much the 2PP vote differs between an average electorate, and an electorate with income in the top 10% of all electorates (with all other socio-demographic variables unchanged).

Unemployment

Unemployment however, is not as influential. In 2001 and 2004, electorates with higher unemployment align with Labor, but over time this shifts towards support for the Liberal party, culminating in a significantly positive (but small) effect in 2016.

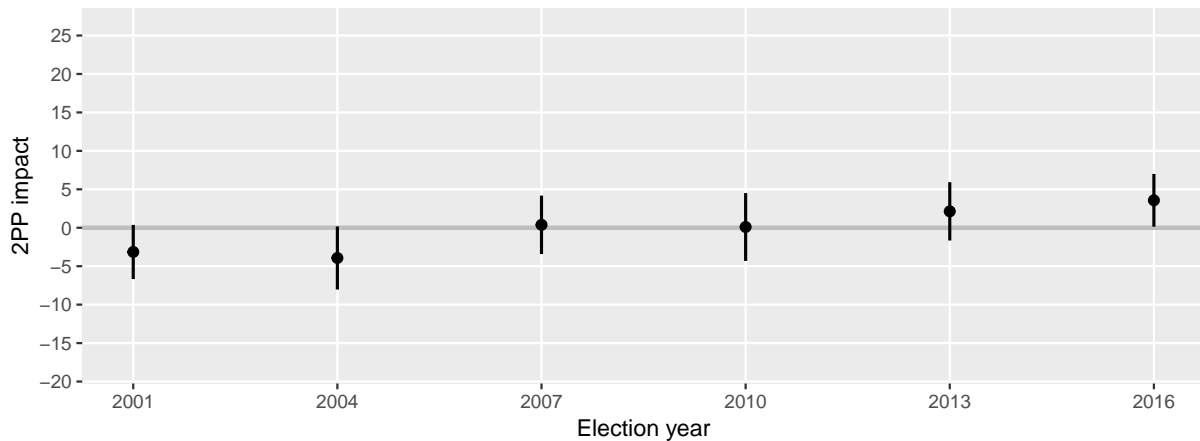


Figure 5: *Estimated impact of electorate unemployment on the two-party preferred vote.*

The vertical scale can be interpreted as a score measuring how much the 2PP vote differs between an average electorate, and an electorate with unemployment in the top 10% of all electorates (with all other socio-demographic variables unchanged)

Industry and type of work

Electorates with higher proportions of workers in extractive industries (mining, gas, water, agriculture, waste and electricity) are consistently linked with higher support for the Liberal party, with the magnitude of this effect slightly increasing over the years.

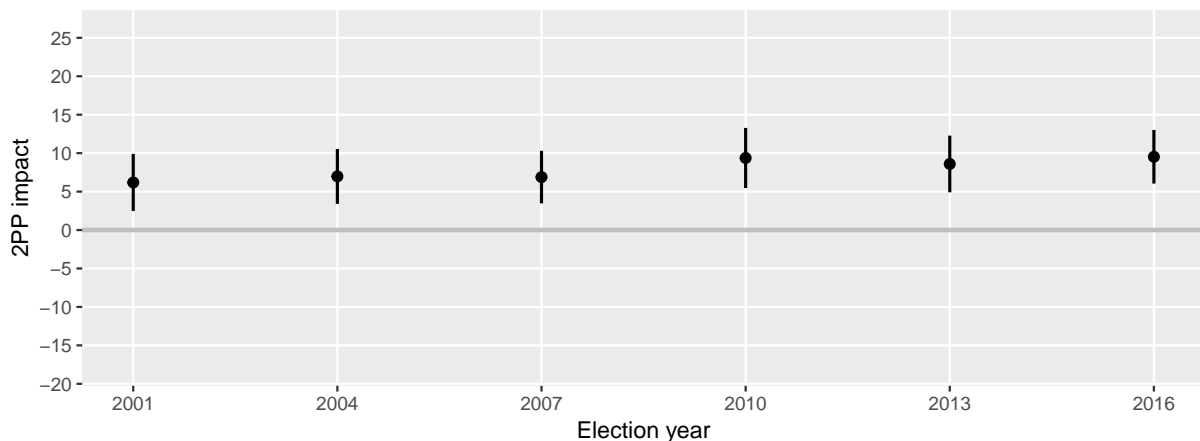


Figure 6: *Estimated impact of employment in the extractive industries on the two-party preferred vote.*

This is unsurprising, as the Liberal party has close ties with these traditional energy industries, and typically present policies to reduce taxation on energy production.

Electorates with more workers in transformative industries (construction or manufacturing) are also more likely to support the Liberal party.

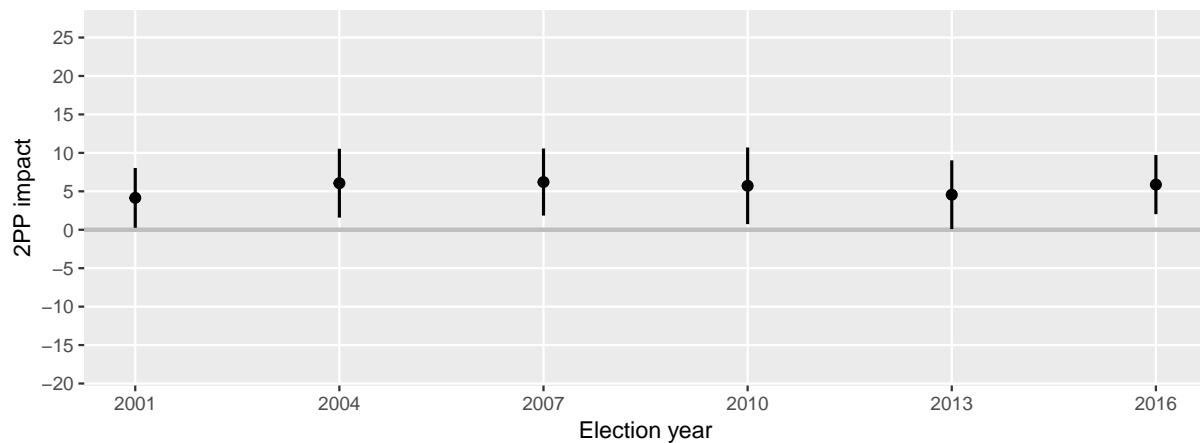


Figure 7: *Estimated impact of employment in the transformative industries on the two-party preferred vote.*

The proportion of workers in managerial, administrative, clerical and sales roles is also a significant predictor of two-party preference vote across all six elections, with a higher proportion of people working these jobs increasing Liberal support.

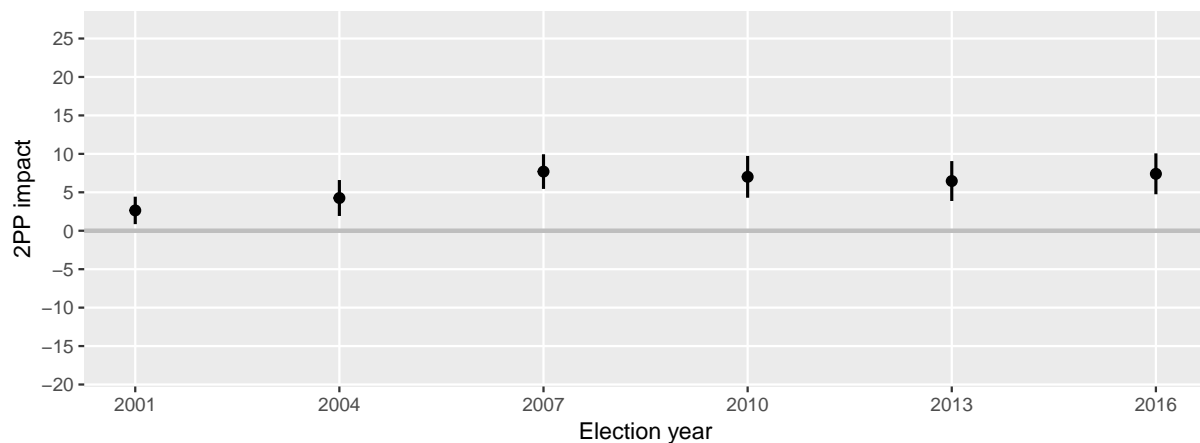


Figure 8: *Estimated impact of employment in management or administration on the two-party preferred vote.*

Household mobility

In each of the six elections, electorates with a higher proportion of people that have recently (in the past five years) moved house are more likely to support the Liberal party.

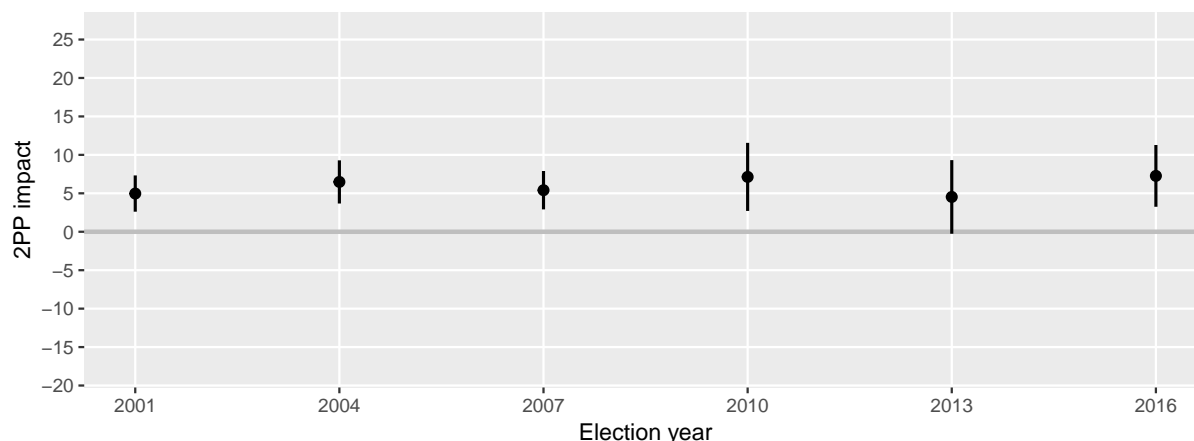


Figure 9: *Estimated impact of household mobility on the two-party preferred vote.*

Our analysis controls for characteristics of home ownership and rental prices, so this effect is not simply due to electorates having low rates of home ownership, or due to electorates having high rental prices. Instead, it suggests that people who are more transient are also more likely to be conservative voters, regardless of their home ownership or rental status. (This would need further study, as we do not have individual level voting data.)

Relationships

De facto relationships, but not marriages, are found to be an important (and significant) predictor of the two-party preferred vote in all six elections, with more de facto relationships associated with higher support for the Labor party. The proportion of individuals who are married however, is insignificant (not shown).

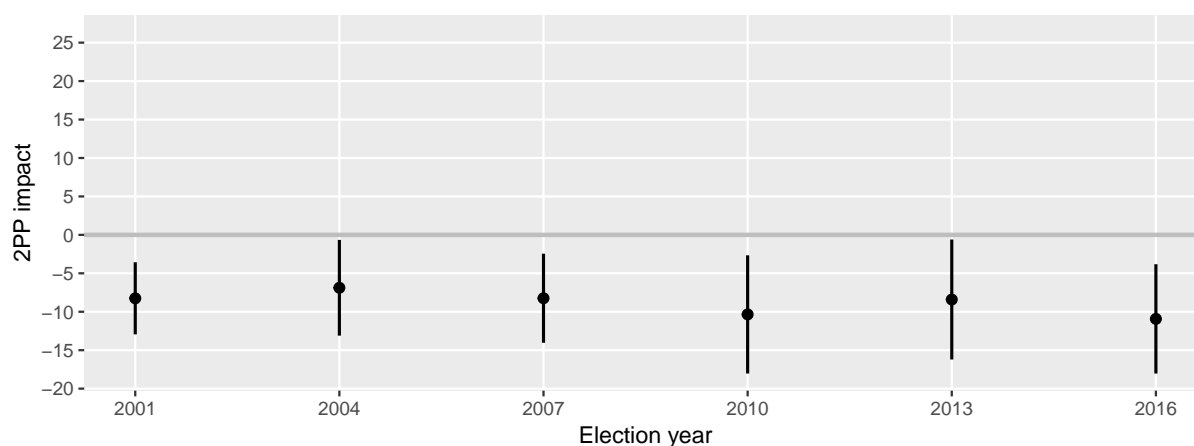


Figure 10: *Estimated impact of de facto relationships on the two-party preferred vote.*

Education

Since 2007, electorates with higher education levels are associated with supporting the Labor party, although this effect is significant only in 2016. Before 2007, education had a negligible effect.

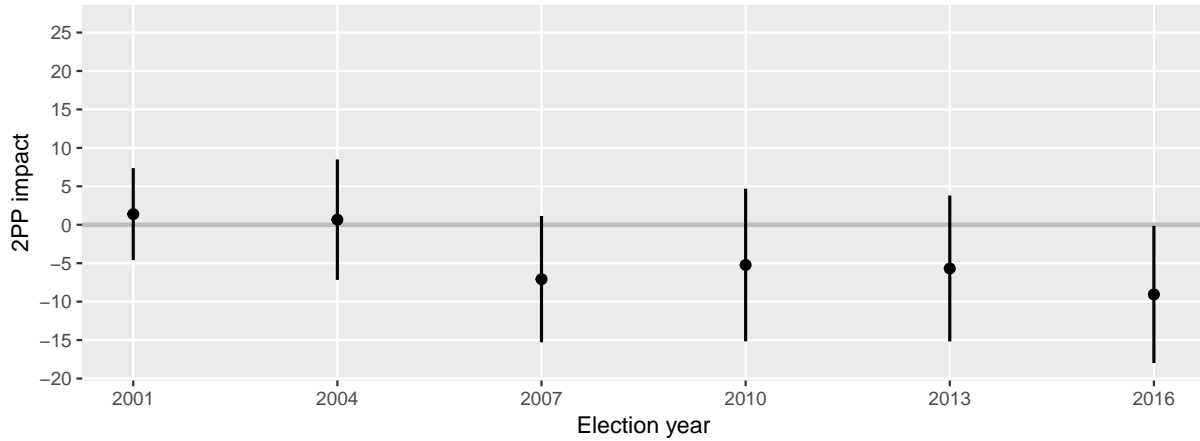


Figure 11: *Estimated impact of education on the two-party preferred vote.*

Diversity

Larger migrant populations from Asia, the Middle East, South-Eastern Europe, the United Kingdom and elsewhere, are either associated with Labor support, or have no effect. Of these areas, only differences in South-Eastern European populations appear to have had a significant impact in each election, with the proportion of Asian migrants also being significant in 2010.

Speaking languages other than English however, appears to have a far stronger effect. Electorates with more diverse languages are associated with higher support for the Liberal party from 2004 onwards, with this effect being significant in 2007, 2010 and 2016.

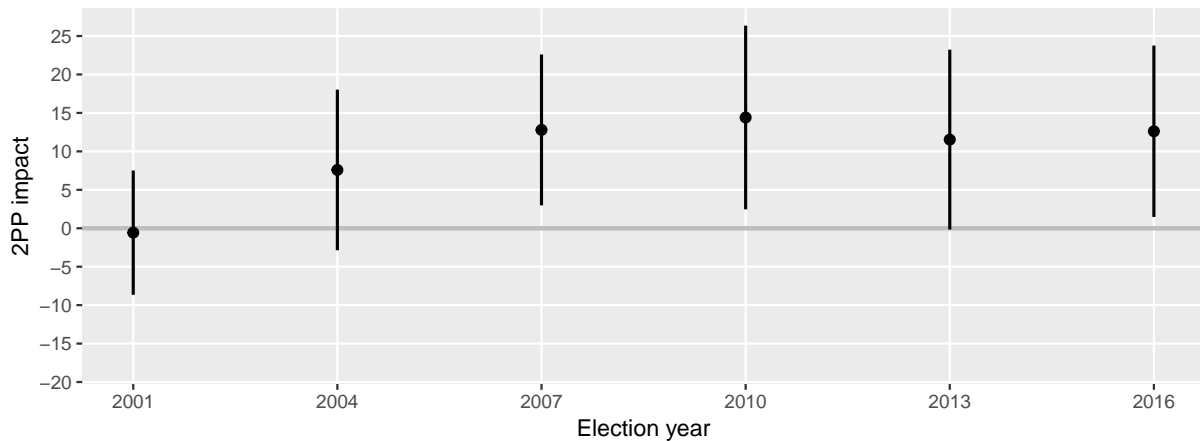


Figure 12: *Estimated impact of ethnic diversity on the two-party preferred vote.*

Of the variables relating to religion, only Judaism shows a consistent effect, with electorates with relatively large Jewish populations more likely to vote Liberal.

4 State effects

It is often suggested that states have systematic differences that cause their electorates to vote differently. We can explore this effect by looking at the difference between the actual 2PP vote in

each electorate and what our modelling would predict based only on electorate characteristics. We found a bias toward Labor in the ACT, and to a smaller extent in Tasmania, and a bias toward the Liberals in the Northern Territory. No other states showed a significantly different voting pattern from what you would expect using only socio-demographic information. That suggests that any differences between Queensland and Victoria (for example) are due to different voter socio-demographics.

5 Against the tide

As well as looking at whether some states have results different from what is predicted, we can also look at individual electorates. That is, when does an electorate vote very differently from what their socio-demographics would suggest? This suggests something is going on beyond the effect of socio-demographics. For example, a very good (or bad) local member could lead to voting patterns that are very different from what is predicted.

Based on the distribution of the Cook's distance values, a Cook's distance greater than 0.1 is considered to be influential and a potential outlier. The electorate of Sydney (NSW) has a large Cook's distance from 2001 to 2013, due to its diverse population (language, birthplace and religion), high number of defacto relationships, high income, high household mobility and small amount of workers in extractive and transformative jobs. It has remained a strong supporter of the Labor party and the Liberal vote is severely overpredicted by the model, making it an outlier. Nearby in metropolitan NSW, the electorate of Wentworth is found to be an outlier in all but the 2007 election. Although historically Liberal, its two-party vote jumped by over 10 percentage points in 2010 without experiencing any notable changes in its socio-demographic makeup — implying that this may be the direct effect of its Liberal member, Malcolm Turnbull, becoming the leader of the Liberal party. Liberal support in Wentworth is underpredicted by the model in each year, and more so with Turnbull as Liberal leader.

Lingiari, an electorate taking up almost all of the Northern Territory, is an outlier in the 2001–2007 elections due to its large Indigenous population, young age profile and low rates of property ownership. Fowler (NSW) has a diverse population with a high proportion of migrants, many Buddhists and Muslims, and has strong Labor support, making it influential in 2001, 2004 and 2010. Other electorates with large Cook's distance are Barton (NSW) and Leichhardt (QLD) in 2016, and Canberra (ACT) in 2007.

6 Conclusion

This paper explores the effects of electoral socio-demographic characteristics on the two-party preferred vote in the 2001–2016 elections, using information from the corresponding Australian federal elections and Censuses. As a Census does not always occur in the same year as an election, Census data for the 2004–2013 elections are generated by employing a method of spatio-temporal imputation. This imputes electoral socio-demographics for the electoral boundaries in place at the time of the election — an approach that is distinctly different from previous work on modelling election outcomes, where Census and election data are typically

joined without addressing their temporal differences. Before estimating a model, these socio-demographic variables are standardized (to adjust for changing variable scales) and many variables (representing similar information) are combined into factors, resulting in a reduced predictor set. A spatial error model is then estimated for each election, accounting for the inherent spatial structure of the data.

Across the past six elections, most of the socio-demographics that drive the electoral two-party preferred vote are found to remain steady, whilst a few (typically weaker) effects vary over time. Industry and type of work are particularly influential, with energy-related and manufacturing/construction jobs, as well as administrative roles being strongly linked with the Liberal party in all elections. Incomes have a similarly consistent effect, with higher income areas supporting Liberal. Higher levels of unemployment shift from weak association with Labor to a significant Liberal effect over the years, and higher education levels are associated with Labor from 2007 (although significant only in 2016). It is also found that electorates with higher household mobility support Liberal, birthplace diversity favours Labor and more de facto relationships align with Labor preference — although marriages, family and household sizes have no material influence. Furthermore, the neighbourhood (spatial) effects are found to be positive in all elections, although significant only in 2001 and 2016, meaning that in the 2004–2013 elections, electorates effectively voted independently.

The findings in this paper complement the existing literature by modelling temporal trends, which as far as the authors are aware, has not been done previously for Australian elections using a regression framework. It is also the first study to model any Australian election since 2010 using Census information.

Additionally, a key contribution of this research is the wrangling of the raw data and imputed data sets for the 2004, 2007, 2010 and 2013 elections, which have been contributed to the *eechidna* R package — providing a rich, accessible data resource for future Australian electoral analysis.

7 Acknowledgements

This paper was produced using RMarkdown (Allaire et al. 2019) and knitr (Xie 2015). All corresponding code for this paper can be found in the github repository github.com/jforbes14/eechidna-paper, and the data used is available in the *eechidna* package (Forbes et al. 2019). All raw data was obtained from the Australian Electoral Commission, the Australian Bureau of Statistics and the Australian Government.

8 Software

All election and Census datasets, along with electoral maps and more, are available in the *eechidna* (Exploring Election and Census Highly Informative Data Nationally for Australia) R package, which can be downloaded from CRAN. The *eechidna* package makes it easy to look at the data from the Australian Federal elections and Censuses that occurred between 2001 and 2016. This study contributed a large revision to the *eechidna* package, which included the

addition of election and Census data for 2001–2010, voting outcomes for polling booths and imputed Census data for election years. For more details on using *eechidna*, please see the articles (vignettes) on the github page ropenscilabs.github.io/eechidna/.

The authors would like to sincerely thank Anthony Ebert, Heike Hofmann, Thomas Lumley, Ben Marwick, Carson Sievert, Mingzhu Sun, Dilini Talagala, Nicholas Tierney, Nathaniel Tomasetti, Earo Wang and Fang Zhou, all of whom have contributed to the *eechidna* package.

References

- Allaire, J, Y Xie, J McPherson, J Luraschi, K Ushey, A Atkins, H Wickham, J Cheng, W Chang & R Iannone (2019). *rmarkdown: Dynamic Documents for R*. R package version 1.12. <https://rmarkdown.rstudio.com>.
- Forbes, J, D Cook, A Ebert, H Hofmann, RJ Hyndman, T Lumley, B Marwick, C Sievert, M Sun, D Talagala, N Tierney, N Tomasetti, E Wang & F Zhou (2019). *eechidna: Exploring Election and Census Highly Informative Data Nationally for Australia*. R package version 1.3.0. <https://CRAN.R-project.org/package=eechidna>.
- Xie, Y (2015). *Dynamic Documents with R and knitr*. 2nd ed. Boca Raton, Florida: Chapman and Hall/CRC. <http://yihui.name/knitr/>.