# 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.

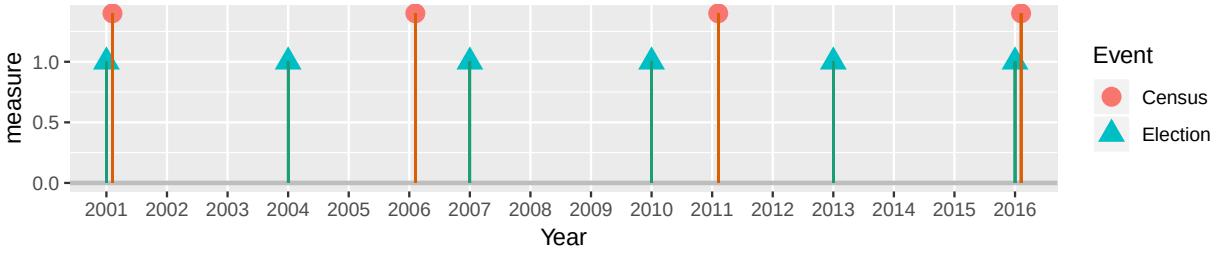


**Figure 1:** *Timeline of Australian elections and Censuses. Censuses happen regularly, elections not quite regularly, which means data needs to be interpolated by time.*

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 Coalition candidates, ignoring all other candidates. We express this as a percentage preference in favour of the Coalition — for example, a 2PP value of 45% indicates that 45% of voters ranked the Coalition 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.

### Country-wide trend

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

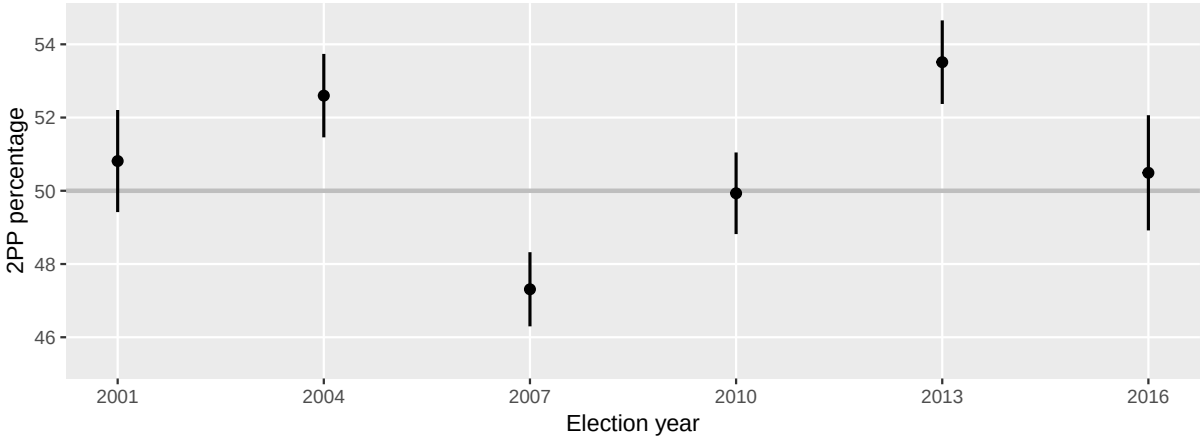


**Figure 2:** *Estimated two-party preferred vote for an electorate with average characteristics at each election.*

This shows that the baseline of party preference has varied over the elections, with the biggest swing occurring in the 2007 election where the average 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 Coalition party.

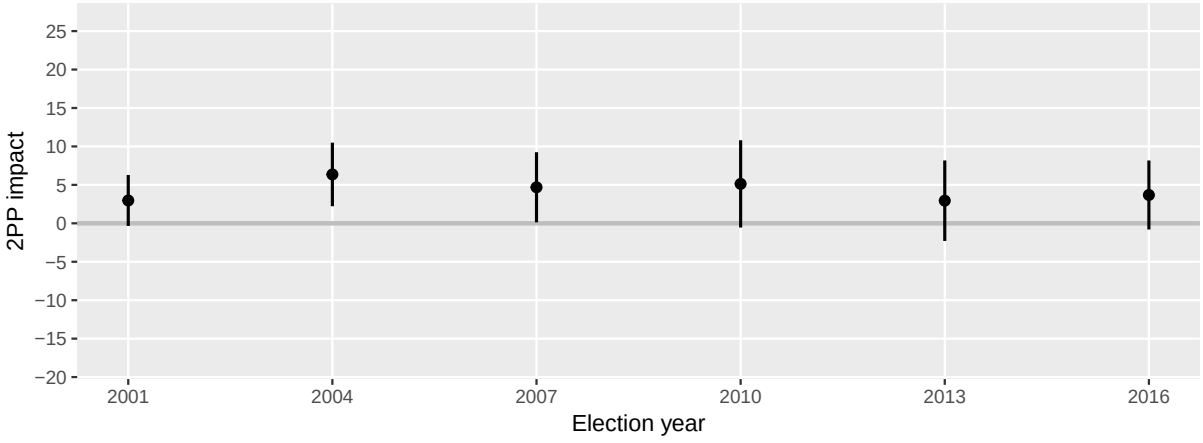


**Figure 3:** *Estimated impact of electorate median age on the two-party preferred vote.*

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.

When the uncertainty range covers the horizontal 0 line, the resulting impact can be considered statistically undetectable. Here age is a statistically significant effect only in 2004 and 2007.

### 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 Coalition, as the Incomes factor has a positive effect on Coalition preference. The effect also seems to have been increasing over time.

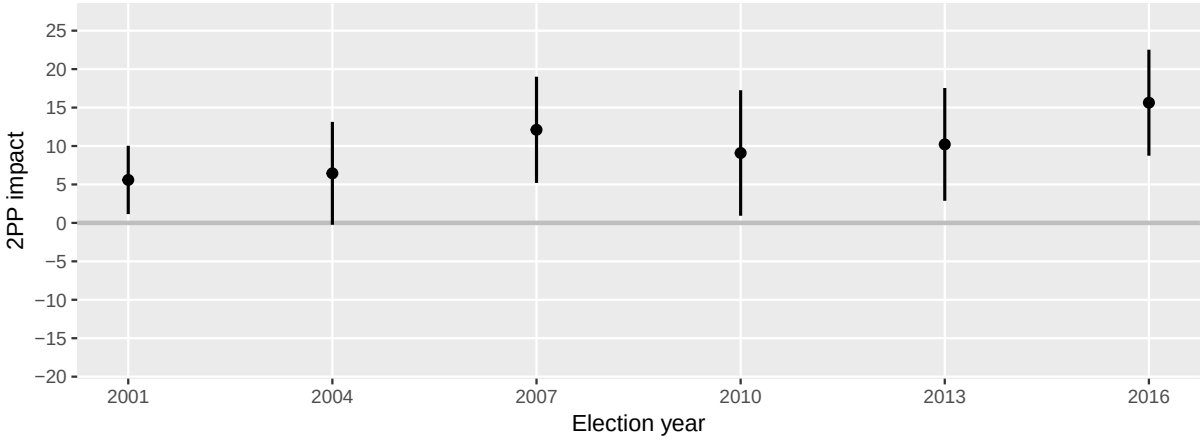


**Figure 4:** *Estimated impact of electorate income on the two-party preferred vote.*

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 Coalition, culminating in a significantly positive (but small) effect in 2016.

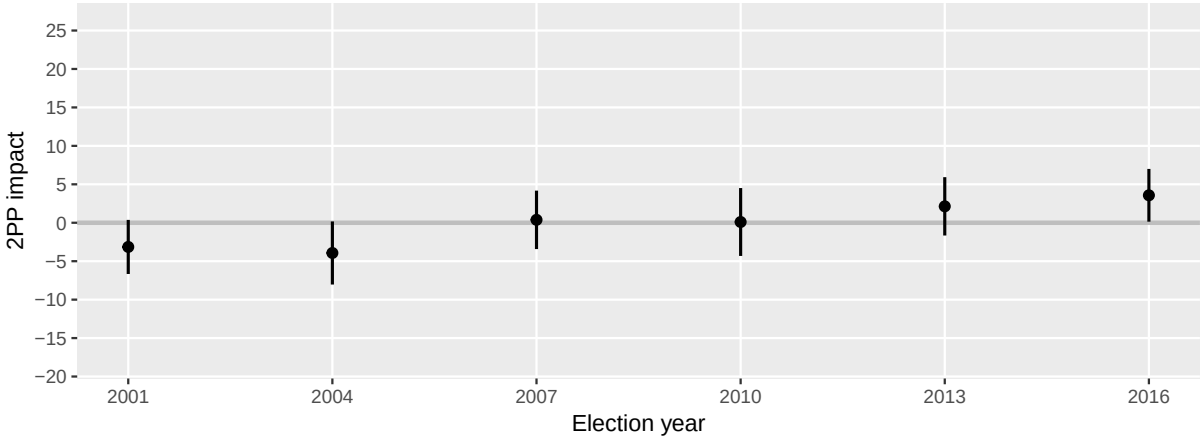


**Figure 5:** *Estimated impact of electorate unemployment on the two-party preferred vote.*

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 Coalition, with the magnitude of this effect slightly increasing over the years.

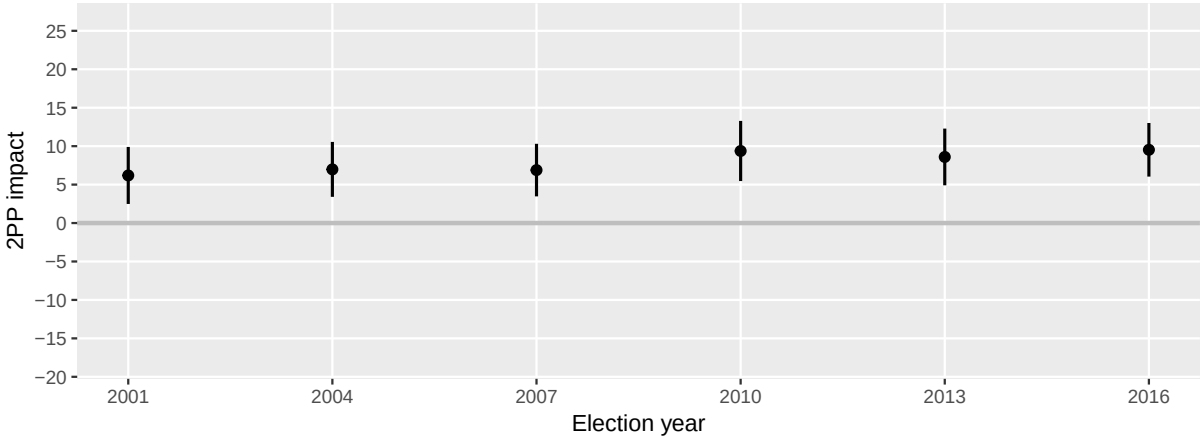


**Figure 6:** *Estimated impact of employment in the extractive industries on the two-party preferred vote.*

This is unsurprising, as the Coalition 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 Coalition.

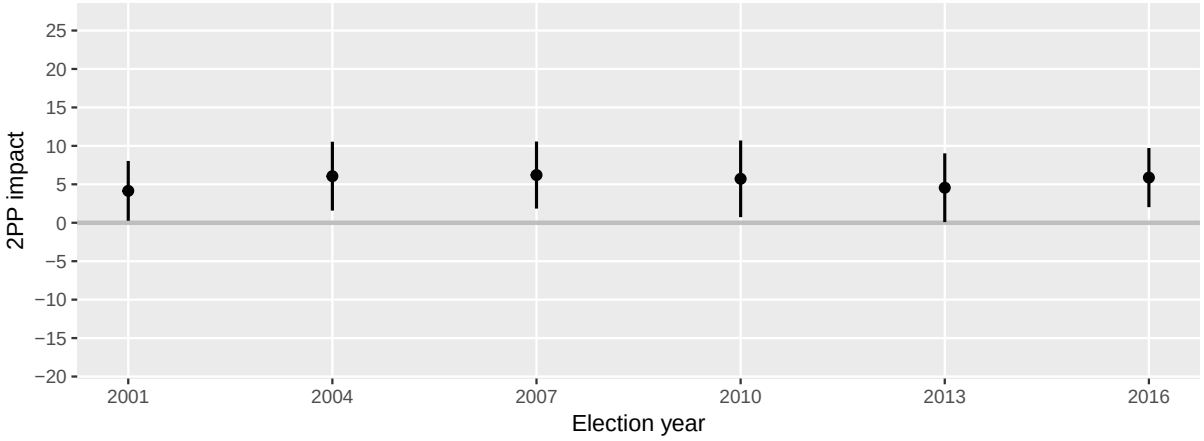


**Figure 7:** *Estimated impact of employment in the transformative industries on the two-party preferred vote.*

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 Coalition support.

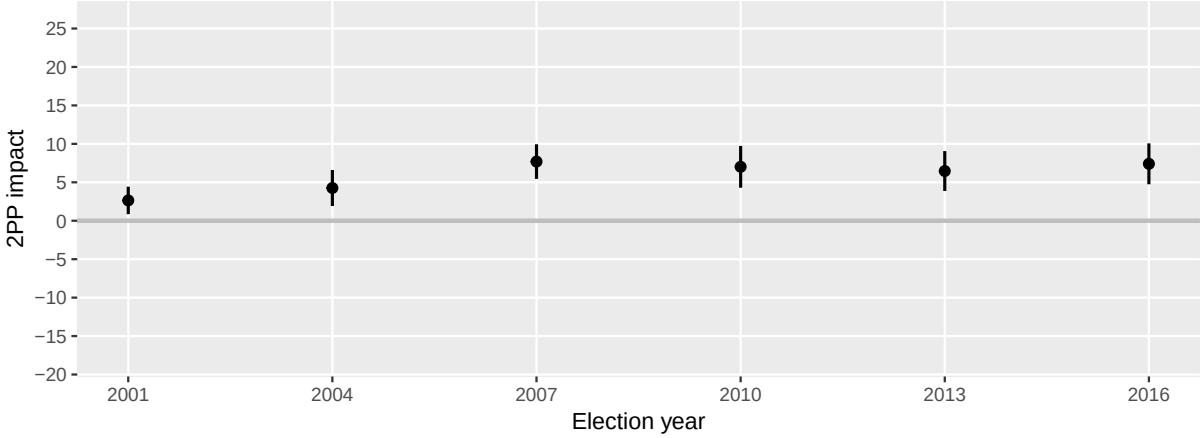


**Figure 8:** *Estimated impact of employment in management or administration on the two-party preferred vote.*

### 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 Coalition.

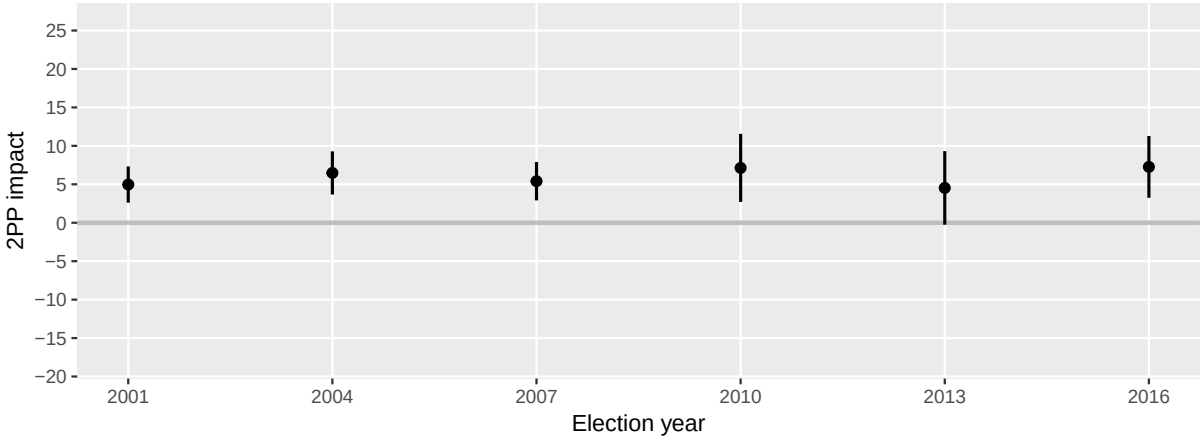


**Figure 9:** *Estimated impact of household mobility on the two-party preferred vote.*

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).

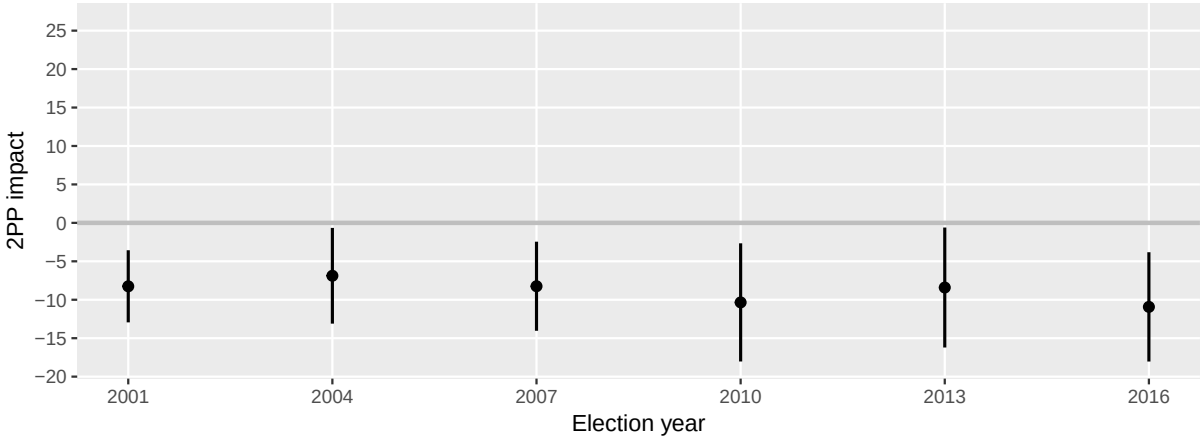


**Figure 10:** *Estimated impact of de facto relationships on the two-party preferred vote.*

## 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.

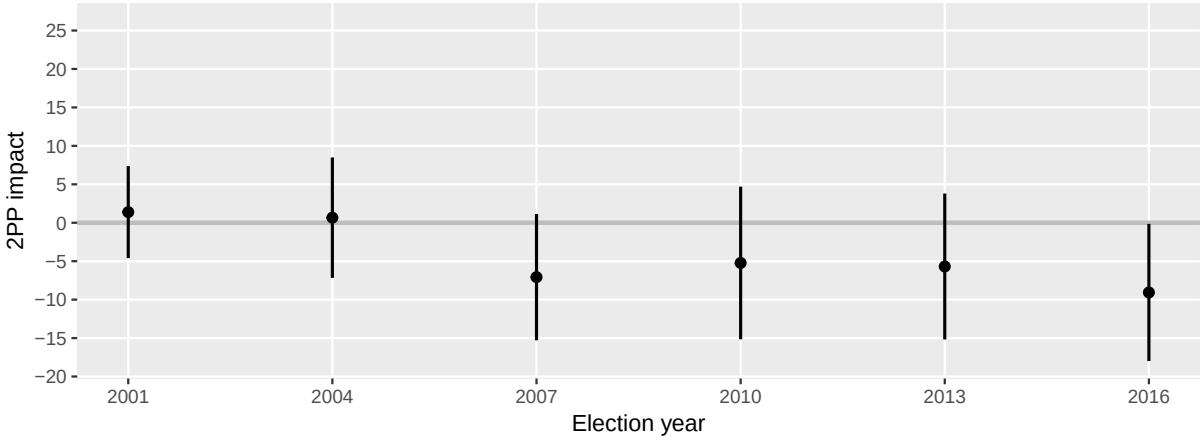


**Figure 11:** *Estimated impact of education on the two-party preferred vote.*

### 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 Coalition from 2004 onwards, with this effect being significant in 2007, 2010 and 2016.

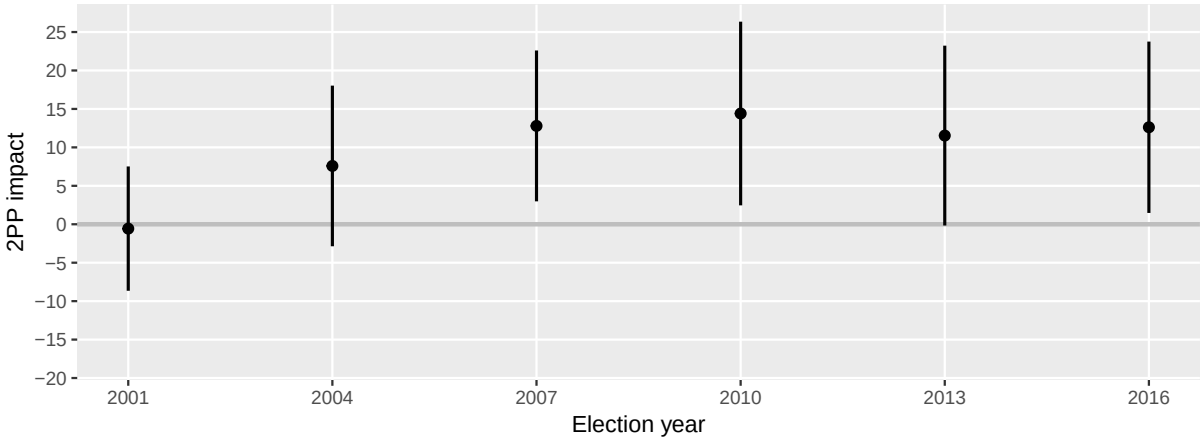


**Figure 12:** *Estimated impact of ethnic diversity on the two-party preferred vote.*

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

## 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 Coalition 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.

By this measure, the top five most surprising results in the last six elections were the following:

| Electorate | Election | Actual 2PP | Predicted 2PP |
|---|---|---|---|
| Capricornia | 2004 | 44.9 | 60.6 |
| Capricornia | 2001 | 43.1 | 58.8 |
| Herbert | 2013 | 56.2 | 41.6 |
| Wentworth | 2013 | 67.7 | 53.1 |
| Herbert | 2016 | 50.0 | 35.8 |

Capricornia is on the Queensland coast, taking in Rockhampton, Yeppoon and parts of Mackay. It was won in 2001 and 2004 by Labour MP Kirsten Livermore, defying the socio-demographic prediction which suggested it would be a safe Coalition seat. She also won it in 2007 and 2010, but with a smaller number of votes, so the result was not so different from predicted. It is currently held by Michelle Landry for the Coalition.

Herbert is another Queensland seat taking in Townsville. It was won in 2013 for the Coalition by Ewen Jones, and in 2016 by Cathy O'Toole for Labor by just 37 votes. The model predicts a much lower vote for the Coalition than tends to happen in practice.

Wentworth probably needs no explanation — the seat formerly held by Malcolm Turnbull which he won with a big majority in 2013. 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.

## 6 True Blue Seats

Another way the model can be used is to identify the seats with the strongest propensity to vote one way or the other, based on their socio-demographics.

The top five most blue-ribbon Coalition seats in the last six elections are listed below. Mitchell takes in the north-western outer suburbs of Sydney, including Baulkham Hills, Kellyville, and parts of Castle Hill. Maranoa occupies a large part of south west Queensland including towns such as Charleville, Kingaroy and Winton. Mallee is, of course, the north west corner of Victoria including Mildura, Swan Hill and Stawell.

| Electorate | Election | Actual 2PP | Predicted 2PP |
|---|---|---|---|
| Mitchell | 2001 | 71.3 | 76.4 |
| Mitchell | 2004 | 70.7 | 75.9 |
| Maranoa | 2013 | 72.3 | 75.6 |
| Maranoa | 2016 | 67.5 | 73.9 |
| Mallee | 2013 | 73.7 | 73.8 |

The top five seats most likely to vote Labour, based on their socio-demographics, are listed below. Batman has since been abolished, but occupied the inner north of Melbourne. Gellibrand takes in the inner western suburbs of Melbourne including Altona and Williamstown, while Fowler is in the western suburbs of Sydney, including Cabramatta and Liverpool.

| Electorate | Election | Actual 2PP | Predicted 2PP |
|---|---|---|---|
| Batman | 2007 | 24.1 | 22.3 |
| Batman | 2010 | 25.1 | 22.5 |
| Fowler | 2004 | 28.6 | 23.6 |
| Fowler | 2001 | 28.5 | 24.9 |
| Gellibrand | 2010 | 26.1 | 25.2 |

# 7 Acknowledgements

This analysis is based on a research paper coauthored by Rob Hyndman and Di Cook from the Department of Econometrics and Business Statistics at the Monash Business School, and Jeremy Forbes, a former Monash University Honours student in econometrics. The full analysis is available at https://robjhyndman.com/publications/elections/.

The code used for the analysis can be found in the github repository github.com/jforbes14/eechidna-paper.

All election and Census data since 2001, 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 uses data obtained from the Australian Electoral Commission, the Australian Bureau of Statistics and the Australian Government.