Why Waste Your Vote? Informal Voting in Compulsory Elections in Australia

EAMON MCGINN [#] AND SHIKO MARUYAMA [*]

Version dated: 5 April 2021

In Australia, where voting is compulsory, around 5% of votes are informal, not counting toward the outcome. Between 2004–2016, 32% of electorates reported more informal votes than votes in the margin between the winner and runner-up. Using exogenous changes in electorate boundaries, we test two hypotheses from the literature. We find the pivotal voter theory unsupported, except that better-educated voters respond to the margin more strategically. However, we do find that more candidates cause more informal votes. This choice-overload effect is observed regardless of voters' education, indicating the role of time and effort cost rather than just cognitive ability.

JEL: C21, D72, D73, D91, H11, P16

Keywords: informal voting, voter participation, pivotal voter model, redistricting, choice overload, compulsory voting, difference in differences, Australia

Eamon McGinn (Corresponding Author): UTS, UTS Business School PO Box 123, Broadway, NSW 2007, Australia (e-mail: Eamon.G.McGinn@student.uts.edu.au). Eamon holds a paid role at Deloitte Australia and is supported by an Australian Government Research Training Program Scholarship.

* Shiko Maruyama: UTS, UTS Business School PO Box 123, Broadway, NSW 2007, Australia (e-mail: Shiko.Maruyama@uts.edu.au).

I. Introduction

Australia is one of the few developed countries where voting is compulsory. However, in federal elections held between 2004 and 2016, around 5% of votes for the House of Representatives (the lower house) were classified as 'informal' (being either empty or improperly filled in), and did not count toward the election outcome. Such 'wasted' votes are a major issue, because high rates can raise concerns about the legitimacy of a government (Lijphart 1998), and can also disproportionally affect certain groups of citizens, typically the less advantaged (Kawai *et al.* 2021). Between 2004–2016, 32% of electorates received more informal votes than those in the margin between the winner and runner-up.

Why would voters who have travelled to the polls and queued up then choose to waste their vote? And what are the major determinants of informal voting? The literature on voter turnout offers one potential explanation, which concerns the probability that the vote will be pivotal (Downs (1957), Riker and Ordeshook (1968), and Palfrey (1985)). Levine and Palfrey (2007) find that voters are more likely to participate in voting when the election is expected to be close, and when the number of voters in an electorate is relatively small. The same logic may apply to informal voting.

An alternative explanation for voter participation is that voters are averse to the mental processing costs involved in correctly completing a ballot. Horiuchi and Woodruff (2017), Cunow (2014), and Augenblick and Nicholson (2015) show that, when presented with more candidates, voters are more likely to abstain from voting and to make errors on the ballot. This is similar to the findings in the broader literature relating to choice overload (Iyengar and Lepper 2000).

This paper advances our understanding of voting behavior by quantifying the causal effects of three important factors on the rate of informal voting: the competitiveness of the electorate, the number of voters in the electorate, and the

number of options on the ballot. The first two factors relate to the likelihood of a voter casting a pivotal vote, while the third relates to the mental processing cost hypothesis. To test these hypotheses, we use polling place-level data for Australian elections from 2004 to 2016, and exploit a natural experiment created by a feature of the electoral system – periodic changes to electorate boundaries that result in some voters experiencing an exogenous, discontinuous change in competitiveness, number of voters, and number of candidates. We also examine how these effects vary across groups with different education levels.

To the best of our knowledge, we are the first in the literature to quantify these different hypotheses in a unified statistical model of voter participation. The winning margin, the number of voters in an electorate, and the number of candidates are often interdependent and endogenously related, so studying all three simultaneously allows for better causal interpretation. More importantly, gauging the relative importance of these hypotheses has direct policy implications. Low participation due to a low probability of a voter casting a pivotal vote, and low participation due to the challenges of correctly completing a ballot, have different implications for policies that seek to increase formal voting rates.

There is disagreement in the literature about the causes of informal voting. This lack of consensus is likely attributable to the paucity of rigorous causal studies and the abundance of observational studies. Hill and Young (2006) argue that informal voting in Australia is positively related to the complexity of the voting process. This aligns with a time-series analysis by Nagler (2015), which shows that the more candidates on the ballot, the more informal voting occurs. While these two studies relate to the mental processing cost hypothesis, Galatas (2008) tests the pivotal voter model in Canada, finding a positive correlation between the anticipated closeness of the election and informal voting. By contrast, De Paola and Scoppa (2014), one of the few causal studies on informal voting, exploits differences in electoral rules for small and large municipalities in Italy and finds that closeness has no effect. Power and Garand

(2007) find that income inequality, urbanization, and compulsory voting tend to increase the level of informal voting in South America. Driscoll and Nelson (2014) find that a high level of informal voting in South America is associated with opposition to the party and president in power.

This paper contributes more generally to the literature on voter turnout and voter participation. Although this literature is well developed and contains many causal studies, findings are mixed (see literature reviews by Geys (2006), Blais (2006), Cancela and Geys (2016), and Stockemer (2017)). A number of causal studies, such as Fauvelle-Aymar and Francois (2006), Indridason (2008), and Garmann (2014), find that competitiveness increases turnout. Lyytikainen and Tukiainen (2019) find that a larger number of voters in an electorate decreases turnout. By contrast, Matsusaka (1993), exploiting different ballot propositions on the same ballot, finds no relationship between expected closeness and turnout. Similarly, using lab and field experiments, Gerber *et al.* (2020) find little correlation between closeness and turnout. And, in a field experiment, Bidwell *et al.* (2020) find no relationship between exposure to candidate debates and turnout.

A common approach in the causal literature on voter participation is to exploit changes to electorate boundaries, as we do in this paper (see Ansolabehere *et al.* (2000), Desposato and Petrocik, (2003), Fraga (2016), and Jones and Walsh (2018) for US examples). A major concern associated with this approach is the potential endogeneity of boundary changes due to gerrymandering. Since an independent, nonpartisan commission is responsible for making such changes in Australia, the endogeneity concern is minimal.

Another methodological concern in the voter participation literature is the impact of travel costs on voting behavior. Many causal papers, such as Funk (2010), Godefroy and Henry (2016), and Schelker and Schneiter (2017), find that turnout decreases with travel costs. However, the effects of travel costs and other relevant aspects of elections, such as closeness and the number of candidates, often change simultaneously, making it challenging to distinguish

one effect from another. Our setting allows us to abstract from travel costs and focus on quantifying alternative hypotheses regarding voter behavior.

By considering how the effect of election characteristics varies across groups with different education levels, this paper also contributes to the extensive literature on the relationship between voting behavior and socio-economic status (see Geys (2006) and Cancela and Geys (2016) for thorough reviews of this literature). Our subgroup analysis, moreover, sheds light on Feddersen and Pesendorfer's (1996) hypothesis that uninformed voters choose to abstain in order to allow informed voters to determine the election outcome. This model has received empirical support in lab experiments (Battaglini *et al.* 2008) and in natural experiments (Lassen 2005).

We find that a higher winning margin and a larger number of voters do *not* increase the rate of informal voting, as hypothesized. By contrast, we find that having more candidates on the ballot leads to higher levels of informal voting. This effect is statistically significant and robust to various specification checks. Further, the magnitude of the effect is of practical relevance: a back-of-the-envelope calculation indicates that, if the number of options on a ballot were halved, informal voting would fall by 31%, reducing its share of the overall vote from 5.3% to 3.7%.

The analysis of educational heterogeneity yields interesting patterns. Though we do not find the expected positive effect of the margin on informal voting for the overall sample, this effect *is* found in better-educated areas, indicating that better-educated voters vote more strategically. Together with the observation that the level of informal voting decreases with more education, this finding supports the hypothesis of Feddersen and Pessendorfer (1996) that uninformed voters abstain in deference to informed voters. Yet voters in better-educated areas respond to the number of options on the ballot in the same way as those in less well-educated areas, indicating that this behavioral response is better explained by time and effort cost rather than just cognitive ability. Further analysis of different types of informal votes indicates that, regardless of their

level of education, voters faced with numerous candidates engage in both intentional and unintentional informal voting, supporting the mental processing cost hypothesis.

Overall, these findings suggest that informal voting is related to the challenges involved in understanding and ranking multiple candidates, rather than to the probability of a voter casting a pivotal vote. This suggests that better designed ballots and simplified voting rules could reduce the rate of informal voting and achieve better representation of voters' preferences.

II. Background on Elections in Australia

As the research design exploits institutional factors specific to Australia, we first provide an overview of the Australian electoral system. Further discussions of the system and voter turnout are provided in Appendices A and B, respectively.

A. The Australian Electoral System

Australia is a federation of six states and two territories. National elections occur roughly every three years, with consistent voting rules across all states and territories. The Parliament in Australia consists of two houses: The House of Representatives (henceforth House) and the Senate. The House has 150 members, with each member elected for a term of three years to represent a single geographic area, called an 'electorate'. Each electorate has a population of around 100,000 voters. The Senate, meanwhile, has 76 senators; each state elects 12 senators to represent it, while each territory elects two senators. Thus, at elections, each electorate offers a selection of different House candidates, whereas within a state or territory the Senate candidates are the same for all electorates.

Voting for the House and the Senate coincides, although Senators serve six years, with half of the Senate elected roughly every three years. The candidates for each electorate for both Houses of Parliament are officially

announced around 3–4 weeks before election day, and the order of candidates on the ballot is randomized at the electorate level. Voters arrive at a polling place in their local electorate, have their name marked off the electoral roll, and are issued with two voting papers (ballots): one for the House and one for the Senate. On both ballots, voters must rank their preferred choices. While all candidates for the House must be ranked, voting for the Senate does not require this exhaustive ranking. Until 2016, Senate voters could either mark one preferred party or rank all candidates. As of 2016, voters need to rank at least six parties or 12 candidates.

In the House, the winner in each electorate is selected through a process known as full-preferential preference voting, or instant-runoff voting. In this process, ballots are initially counted for each voter's first preference, then the candidate with the lowest share of the vote is eliminated and their votes are redistributed to the candidate with second preference on the ballot. This process is repeated until one candidate has a majority of the vote. In the Senate, each state and territory's representatives are selected using a single transferable vote system of proportional representation. In this system, votes are counted in a similar way to the House, but since each state and territory has multiple Senators, a Senator is elected once they receive a certain proportion of the overall vote (normally 14.3% for states and 33% for territories), rather than a majority.

Every Australian citizen aged 18 and over is required to vote. Failure to vote can attract fines ranging from $20–210 Australian Dollars (plus legal costs) and can result in a criminal conviction. In addition to voting in person at a local polling place on election day (ordinary voting), postal and pre-poll voting options are available. As a result, turnout is relatively high in Australia – around 91% in 2016 (AEC 2017).

Despite voting being compulsory, it is not illegal to submit an empty or defaced ballot rather than a legitimate vote. This type of vote is recorded in the voting data as 'informal'. Figure 1 shows examples of informal votes. Informal

voting can also occur due to misunderstanding of the voting rules or to a mistake when filling in the ballot (for example, marking two candidates as a first preference). The most common type of informal voting is where a voter, mistakenly or otherwise, ranks only one candidate on the ballot, accounting for an average of 28% of informal votes. The next most common type, averaging 23.6% of informal votes, is a blank ballot (see Appendix A).

In elections held between 2004–2006, the rate of informal voting was 5–6% in the House and 3–4% in the Senate.

National elections are administered by the Australian Electoral Commission (AEC). The AEC is a federal government agency that oversees the organizing, conducting, and supervising of federal elections and referendums. The AEC's existence and role are an important piece of the institutional voting framework and critical to this paper's research design. The AEC's structure means that many of the administrative aspects of voting, including electorate boundary changes, are conducted independent of political considerations.

B. Changing Electorate Boundaries

The AEC regularly changes electorate boundaries – a practice referred to as redistribution. For example, during 2000–2017, there were 20 redistributions,[1] with at least one redistribution conducted between every election. Redistributions are undertaken to ensure that each electorate gains representation in the House in proportion to its population. This process is conducted with minimal and balanced political input.[2]

Our analysis uses voting data at the polling place level. A polling place is the location where voters report on election day. Most polling places are located

[1] ACT x2, NSW x3, Northern Territory x3, Queensland x3, South Australia x2, Tasmania x2, Victoria x2, and Western Australia x3.

[2] Senior public servants are responsible for conducting the redistribution. Two political representatives are included in the panel that responds to any objections to the redistribution. Any objections, and the response, are thoroughly documented in the Report of the augmented Electoral Commission that is delivered before finalizing the redistribution.

in schools, church halls, or community centers, which means they do not physically move location; however, redistributions often result in polling places switching from one electorate to another. When a polling place switches electorate, the local voters who have moved with it will continue to vote at the same polling place, but will now be voting in a different electorate. This means that these voters experience a discontinuous change in the margin, number of voters, and number of candidates on the ballot.

Figure 2 shows examples of polling places switching electorate. The main figure highlights the number nationally that switched electorate between the 2010 and 2013 elections. The insets show a particular instance in western Melbourne where two neighboring electorates exchanged boundaries and polling places.

III. Research Design and Econometric Specification

We test three hypotheses drawn from the literature: (H1) when the expected margin in an election is bigger, informal voting will be higher; (H2) when there are more voters in an electorate, informal voting will be higher; and (H3) when there are more candidates on the ballot, informal voting will be higher. The margin is a standard indicator of the closeness of an election, and is defined as the winning candidate's vote share minus 50%. In the Australian context, where instant-runoff voting is used, the margin refers to the final round of preference distribution that results in the selection of a winning candidate.[3]

Raw correlations in the data support the three hypothesized relationships. Figure 3 shows binned scatter plots of informal voting percentages in the House at each polling place against the margin, electorate size, and number of options on the ballot. All three panels show positive, statistically significant

[3] Hence, our margin indicator can considerably differ from the margin defined based on first preferences. To examine whether our conclusions are driven by the particular definition we use, we have also considered the latter definition, but it does not materially alter our conclusions. The latter definition also lacks theoretical justification because it does not directly determine the probability of being a pivotal voter in an election.

relationships, although the effect size for the margin does not appear to be practically meaningful, with the estimated effect on informality being only 0.4 percentage points when moving from a 0% to a 25% margin. A similarly weak relationship is reported by Kawai *et al.* (2021).

The statistically significant relationships in Figure 3 do not necessarily imply causal relationships. Consider the following linear model:

$$(1)\quad Informal_{it} = \theta_1 Margin_{it} + \theta_2 \ln(Voters)_{it} + \theta_3 \ln(N\ Options)_{it} + \boldsymbol{X}_{it}\beta + \epsilon_{it}\ ,$$

where $Informal_{it}$ denotes the percentage of informal votes in the House (ranging from 0 to 100) recorded at polling place i in year t. $Margin_{it}$, $\ln(Voters)_{it}$, and $\ln(N\ Options)_{it}$ are the margin (ranging from 0 to 1), the natural log of the number of voters, and the natural log of the number of candidates on the ballot at polling place i in year t, respectively. These particular scales and transformations have been selected in order to give the coefficient parameters a comparable and meaningful interpretation. $\boldsymbol{X_{it}}$ is a vector of covariates such as demographic characteristics, and ϵ_{it} denotes the error term. Assuming potential over-time correlation of errors within a polling place, we use standard errors clustered at the polling place level.[4]

The OLS estimates for θ_1, θ_2, and θ_3 do not produce causal estimates due to endogeneity problems. Likely omitted variables include knowledge of and interest in politics, which systematically vary across polling places. Simultaneity bias may also arise if voters' behavior affects which candidates are available in the electorate, which in turn affects the margin and the number of options. These endogeneity issues have been widely acknowledged in existing studies (see Cox and Katz (2002), Carson and Crespin (2004), and Jones and Walsh (2018)).

[4] To calculate the clustered standard errors, we use the LFE package in R. Clustered standard errors produced by LFE are known to be slightly more conservative than those produced by Stata.

To address the endogeneity concern, changes in electorate boundaries have been used; see Ansolabehere *et al.* (2000), Desposato and Petrocik (2003), Fraga (2016), and Jones and Walsh (2018) for US examples. In this research design, treated polling places are those that switched electorate, while the control group consists of polling places that did not switch electorate. A simple difference-in-differences (DID) model that exploits the exogenous boundary change is:

$$Informal_{it} = \mu \cdot Changed.Electorate_{it} + \gamma_t + \alpha_i + \boldsymbol{X}_{it}\beta + \epsilon_{it}\,, \quad (2)$$

where $Changed.Electorate_{it}$ is a dummy variable indicating whether polling place i in year t switched electorate between the previous election and year t. γ_t denotes time fixed effects, while α_i denotes polling place fixed effects. The DID estimator, μ, identifies the causal effect of an electorate change on $Informal_{it}$ under the common trends assumption (*i.e.*, the occurrence of an electorate change is orthogonal to the over-time change of ϵ_{it} conditional on $\boldsymbol{X}_{it}$).

A potential concern relating to the use of boundary changes is raised by Henderson *et al.* (2016), who argue that electorate boundary changes are politically influenced in many US jurisdictions. We argue that the exogeneity of boundary changes is a reasonable assumption in the Australian context. Australia has a clear and established administrative process for conducting a redistribution. To corroborate our exogeneity assumption, we conduct an efficiency gap analysis, quantitatively comparing redistributions in the US and Australia (Appendix C). The results show that the assumption is plausible in Australia.

Another potential concern is that while votes are counted in an electorate regardless of where people vote, voters are allowed to vote at any polling place. Because we use polling place-level data, our identification strategy relies on the assumption that voters do not systematically change their choice of polling place in response to a redistribution. We argue that this is a reasonable assumption because polling places are located in nearby schools or other public venues, and

rarely move physically. Another concern is that some control polling places may experience some form of treatment if the boundaries of their electorate change, notwithstanding that these particular polling places remain in the same electorate. This concern is partially addressed by the fact that the degree of change undergone by these control polling places should be substantially smaller than that experienced by the treated polling places. Electorates in our data each contain about 55 polling places on average, but only a handful of them change electorate. We conduct additional analysis to test whether the results are sensitive to the inclusion or exclusion of these control polling places.

Our main interest is not the effect of electorate change *per se* but the effects of the margin, number of voters, and number of candidates on the ballot. To achieve causal inference, we instrument these three treatment variables by the exogenous changes in these variables due to electorate change. Building on Equations (1) and (2) leads to our baseline DID-IV regression:

$$(3)\quad Informal_{it} = \delta_1 Margin_{it} + \delta_2 \ln(Voters)_{it} + \delta_3 \ln(N\ Options)_{it} + \delta_4 Changed.Electorate_{it} + \gamma_t + \alpha_i + \boldsymbol{X}_{it}\beta + \epsilon_{it}$$

where $Margin_{it}$, $\ln(Voters)_{it}$, and $\ln(N\ Options)_{it}$ are instrumented by *Change.in.Margin*$_{it}$, *Change.in.*ln(*Voters*)$_{it}$ and *Change.in.*ln(*Options*)$_{it}$, which are changes in the margin, the log of the number of voters in the electorate, and the log of the number of options on the ballot resulting from polling place *i* switching electorate between year *t* and the previous election. These change variables are defined as the difference between the actual values voters in *i* faced and the counterfactual values they would have faced without boundary changes (see Appendix D for further information on relevant variables). Consequently, these change variables take the value of zero for polling places in the control group. δ_1, δ_2, and δ_3 yield estimates of the causal effects of changes in the margin, number of voters, and number of candidates on the ballot. δ_4 captures the causal effect of the electorate change *per se*, not including the effects of the

margin, number of voters, and number of candidates. Our decision to keep the effect of the electorate change *per se* is motivated by findings from Hayes and McKee (2009) that, after redistricting, voters were less likely to complete their ballot fully, compared to voters unaffected by redistricting.

For our instruments to be valid, we need an additional assumption: that *Change.in.Margin*$_{it}$, *Change.in.*ln(*Voters*)$_{it}$, and *Change.in.*ln(*Options*)$_{it}$ are orthogonal to the over-time change of ϵ_{it} conditional on $\boldsymbol{X}_{it}$, analogous to the common trends assumption for (2). Identifying δ_1, δ_2, δ_3, and δ_4 in (3) is therefore more demanding than identifying μ in (2). We argue that the validity of our instruments is credible because the AEC administers electorate redistributions mechanically, without political influence or consideration (see Appendix A). The change in the number of voters may be correlated with recent population changes and hence with electorate changes (as ensuring equal representation across electorates is the purpose of redistributions); however, this does not invalidate our instruments because the population size and recent population changes are included in $\boldsymbol{X}_{it}$.

The timing assumed in (3) requires care when interpreting the results. *Change.in.Margin*$_{it}$, *Change.in.*ln(*Voters*)$_{it}$, and *Change.in.*ln(*Options*)$_{it}$ could affect voting behavior, either immediately or with a lag. It has become common in the literature to study the relationship between margin and voting outcome during the same time period. This is known as the 'ex-post' approach, and is based on the proposition that the ex-post outcome of an election is a good proxy for voters' ex-ante beliefs (see Geys (2006) for further information).

Although electorate changes are administered in an apolitical way, there is still the possibility that the treatment and control polling places differ in meaningful ways, making the exogeneity assumption dubious. To address this concern, we estimate two variants of (3) as robustness checks, using propensity score matching and distance-based restrictions. The propensity score is estimated by regressing *Changed.Electorate*$_{it}$ on a range of observable characteristics of polling place *i* that should not be affected by treatment:

median age, mean income, unemployment rate, population density, population growth rate, average value of a newly constructed house, percentage of households where English is a second language, and percentage of people with post-high school education. These variables are measured in the year of the previous election. We use a logit model to estimate the propensity score, and matching is done based on the nearest neighbor technique without replacement (Ho *et al.* 2007). Further details on matching are provided in Appendix E.

The distance-based approach restricts the sample to polling places within 2.5 kilometers of treated polling places, in the spirit of spatial discontinuity analysis. Distance-based restrictions may provide a better control for time-varying unobservable characteristics associated with each geographic location heterogeneously which cannot be addressed with propensity score matching. A good example is regional political sentiment. When we report the results, we contrast the three approaches, labeling them 'Baseline', 'Propensity Score Matched', and 'Distance Limited'.

Another bias concern arises when redistributions change important margins of the election not captured in our model and those changes are related to the three treatment variables of interest. To address this concern, we extend our DID-IV models to include two additional treatment variables: the tenure of the incumbent candidate (in years) and the share of votes for progressive parties. These two variables are also instrumented by the two change variables induced by boundary change. As shown in Appendix F, this extension does not alter the main results.

The structure of the Australian electoral system provides another potential set of control groups for treated polling places that switch electorate. Voters are required to submit ballots for both the House and Senate at the same time. Because every electorate has different House candidates, the margin, electorate size, and number of options vary across electorates. By contrast, with the Senate, every electorate within a state has the same expected margin, electorate size, and number of candidates. Since redistributions do not affect Senate

voting, we can use it as another control group. Using the Senate data as a control group, we form a House-Senate-DID-IV (HS-DID-IV) as follows:

$$
\begin{aligned}
& Informal_{ijt} \\
(4) \quad & = \delta_1 Margin_{ijt} + \delta_2 \ln(Voters)_{ijt} + \delta_3 \ln(N\ Options)_{ijt} + \varphi_{it} \\
& + \sigma Senate_{ijt} + \sigma^{2016} Senate_{ijt} \cdot Y2016_t + \epsilon_{ijt} .
\end{aligned}
$$

For HS-DID-IV, the sample is restricted to polling places that switched electorate. There are three main differences from our baseline DID (3). First, HS-DID-IV includes both Houses of Parliament, indicated by subscript *j*. The difference between the two Houses is captured by a dummy variable $Senate_{ijt}$, which takes a value of 1 if the observation is for the Senate and 0 otherwise. We also include the interaction of the Senate and Year 2016 dummies to capture the policy change to the Senate ballot rules in the 2016 election (see Appendix A). Second, this specification excludes the *Changed.Electorate* variable, as all observations in this specification have switched electorate. Third, instead of time and polling place fixed effect terms, we include fixed effects, φ_{it}, for each polling place-time pair. This is because we have two observations for each polling place-time pair. We do not include $\boldsymbol{X}_{it}$ in this DID regression since it does not vary between the House and Senate observations. For observations from the Senate, we define *Margin*, *Voters*, and *N Options* slightly differently to House observations. Margin is defined as the percentage point difference between the first- and second-ranked parties on the first round of vote counting. This is because the Senate uses a proportional representation system and so there is no margin between a winning and losing candidate. *Voters* is defined as the number of voters in the state, as the state is the relevant electorate for the Senate. *N Options* is defined as the number of 'above the line' options (meaning the number of parties), as most voters choose to vote for parties rather than candidates. For more information on each of these differences, see Appendix A. Identifying δ_1, δ_2, and δ_3 relies on the same instruments as (3). However, HS-DID-IV (4) may yield different estimates of these parameters, because the

baseline causal effect of the electorate change *per se* (with no change in *Margin*, *Voters*, and *N Options*) is identified by the control group for each specification.

The availability of the two alternative control groups allows for a triple differences approach (DDD-IV):

$$(5)\quad Informal_{ijt} = \delta_1 Margin_{ijt} + \delta_2 \ln(Voters)_{ijt} + \delta_3 \ln(N\ Options)_{ijt} + \delta_4 Changed.Electorate_{ijt} + \varphi_{it} + \sigma Senate_{ijt} + \sigma^{2016} Senate_{ijt} \cdot Y2016_t + \epsilon_{ijt}.$$

This DDD-IV model is estimated using both Houses of Parliament and polling places that switched and did not switch electorate. We report the results of standard DID-IV (3), HS-DID-IV (4), and DDD-IV (5), the latter two acting as robustness checks.[5]

IV. Data

Data on voting and elections is sourced from administrative datasets compiled by Australian government agencies. The primary data source is the AEC's voting data, 'First Preferences by Candidate by Polling Place', for the 2004, 2007, 2010, 2013, and 2016 federal elections (AEC 2018).[6] The unit of observation in our analysis is defined by the polling place, election year, and House of Parliament (House or Senate) associated with a count of informal votes and counts of first preference votes for each candidate. The AEC data also provides voter numbers and the number of options on the ballot in each electorate. To enhance the AEC data, we use data from the Australian

[5] Comparing the strengths and weaknesses of the two control groups, we consider voting behavior in the House in similar polling places is preferable because we conjecture significant differences in voting behavior between the House and Senate. In the House, for a vote to be counted as formal, voters must rank all candidates on the ballot, while in the Senate a voter does not need to rank all candidates, making voting in the Senate less demanding in terms of time, research, and decision-making effort. In fact, informal voting is systematically more prevalent in the House than the Senate (see Appendix D).

[6] Polling place-level data are only available for the 2004 election on.

Broadcasting Corporation (ABC) on the incumbent party in each electorate, the winning party, and the winning party's margin in each election.[7]

There are currently around 8300 polling places in Australia. We select polling places to be included in the analysis as follows. We first select polling places that were used in every election from 2004 to 2016. This reduces the number of polling places to 6360 across 150 electorates. The data remaining is referred to as 'All Data' in Table 1. We then impose the following restrictions to select the sample for our analysis. First, we exclude polling places that at any point between 2004 and 2016 belonged to an electorate that was created, nullified, or renamed over that period.[8] Next, we exclude observations with missing values for the variables used in the analysis. Lastly, we exclude polling place observations that, after the above restrictions have been applied, appear only once during the period 2007–2016. Based on this data, we identify polling places that switched electorate. Data for 2004 is used only to determine whether a polling place switched electorate between the 2004 and 2007 elections. This results in a final dataset that contains 23,096 polling place-year observations for the four elections across 143 House electorates. Inclusion of the Senate yields 46,192 observations at the {polling place, House of Parliament, year} level. This final dataset is referred to as 'DID Data' in the following summary tables.

Table 1 summarizes the polling place sample for the House. In any given election, 2–8% of polling places switched electorate. In 2013, fewer polling places switched electorate because only relatively minor redistributions had taken place since 2010.

Table 2 reports the summary statistics of main variables in the House sample. The mean informal vote share in the House is 5.3%, while in the treated and control samples it is 6.0% and 5.2%, respectively. The average margin is

[7] In the literature, betting market data is often used to anticipate the likely margin in an election. However, it was not possible to source data on historical betting odds for elections at the electorate level in Australia.

[8] Electorates are not frequently created, nullified, or renamed. In this data: the electorate of Charlton, New South Wales, was eliminated prior to the 2013 election, while the electorate of Burt, Western Australia, was created for the 2013 election. Three electorates were renamed over the period 2001–2016.

9.1%. We transform the number of options and number of voters into log form in the DID analysis. Voters at each polling place vote for the same set of House candidates, alongside, in their electorate, about 97,000 other voters on average ($\approx e^{11.48}$). Although the AEC strives to maintain a roughly equal population size in each electorate (around 100,000), variations remain: the smallest electorate contains about 57,000 registered voters while the largest contains about 143,000. The mean number of candidates on the ballot in each polling place is 7.6 ($\approx e^{1.96}$). The minimum number is 3, and the maximum 19. See Appendix D for the distribution of these treatment variables. The "*Change in*" variables are always zero for polling places in the control group, as only treated polling places are affected by the change to electorate boundaries.

The rest of Table 2 lists the control variables used in the causal analysis. We use National Regional Profile data from the Australian Bureau of Statistics (ABS) (ABS 2016a), from which we draw regional data on median age, mean income, unemployment rate, population density, population growth, house values, English as a second language, and percentage of people with a tertiary degree. This data is not available for polling place level but is available for Statistical Area Level 2 (SA2), as defined by the ABS.[9] As a result, a GIS program was used to match latitude and longitude coordinates for polling places to their respective SA2.[10] When used in regressions, population growth is split into two variables: one for areas that experienced population growth and one for those that experienced population decline (the absolute value is used in the regression).

[9] Each SA2 aims to represent a community that interacts together socially and economically (ABS 2016b). There are around 2200 SA2s in Australia, with populations ranging between 3,000–25,000 and an average population of around 10,000.

[10] Most SA2s contain multiple polling places. For example, in 2016 each SA2 contained an average of 3.3 polling places, with the maximum being 46 polling places within a single SA2, the Sydney-Haymarket-The Rocks SA2.

V. Results

A. Main Results

Table 3 presents the main results estimated by using the House sample. Column (1) shows the results of a simple linear regression mirroring the relationships shown in Figure 3: informal voting is positively and statistically significantly associated with the margin, electorate size, and number of options. Columns (2) and (3) add a range of socioeconomic covariates and polling place fixed effects, respectively, showing that the sign and magnitude of the associations between informal voting and the three treatment variables are not robust. In particular, the sign and statistical significance of *Margin* both change across specifications, indicating systematic geographical differences between high- and low-margin elections. Columns (4)–(6) show the results of the DID-IV regressions, which exploit exogenous changes to the margin, number of voters, and number of options resulting from electorate redistributions. Column (4) reports the results of our preferred baseline DID-IV. As robustness checks, Columns (5) and (6) show the estimates of the same DID-IV model with the propensity score matched sample and the distance limited sample, respectively.

The three DID-IV models consistently show that informal voting increases statistically significantly when ln(*N Options*) increases. The effects of *Margin* and ln(*Voters*) are consistently negative, but they are barely statistically significant at the 90% level or insignificant. Hence, our estimates yield no evidence for the hypotheses we posited regarding the competitiveness of the election and the number of voters in the electorate (H1 and H2), while providing strong support for the hypothesis regarding the number of options (H3). The finding for H1 and H2 is similar to that in lab and field experiments in Gerber *et al.* (2020). The finding for H3 is in line with the literature that focuses on the mental processing costs of voting, broadly known as choice-overload (Iyengar and Lepper 2000). The parameter estimate of –1.945 for *Margin* in Column (4) indicates that a one percentage point increase in the margin leads to a 0.01945

percentage point decrease in informal voting. Similarly, doubling the number of voters reduces informal voting by 2.95 percentage points, approximately. The mechanism underlying these negative effects is not clear, but for *Margin* could be related to a 'bandwagon effect', where voters support a dominant candidate who is expected to win (Grillo 2017).

The parameter estimate for *Change in* ln(*N Options*) is statistically and practically significant, indicating that if the number of options on the ballot doubles, the level of informal voting increases by 2.73 percentage points, approximately. With an existing average level of informal voting of around 5.3%, this would be equivalent to a 52% increase. As the number of options tends to be between 4 and 8, changes of 25% to 100% are not uncommon in real life. A back-of-the-envelope calculation suggests that, if the number of options in each electorate was halved, the number of observed informal votes would fall by 31%, reducing the share of informal votes from 5.3% to 3.7% of the total.

The estimated coefficient on *Changed Electorate* is negative and statistically significant across all specifications. This indicates that, even without changes in the margin, electorate size, and number of options, voters whose electorate is redistributed are less likely to submit an informal vote than their counterparts in unchanged electorates. This may be due to voters taking more interest in the 'new' candidates or in issues in their new electorate, and so being more engaged in the political process. This result is the opposite of that reported in Hayes and McKee (2009), who look at voter turnout in Texas and find that redistricting reduces turnout. However, their estimate should be interpreted as the overall effect of redistributions that include changes in the margin, electorate size, and number of options. In addition, Hayes and McKee (2009) do not use fixed effects to control for unobserved factors, and there may be endogeneity issues with their research design, given the politicized nature of redistricting in Texas.

To explore effect heterogeneity, we repeat the baseline DID-IV regression using subsamples defined according to the percentage of people with tertiary

education in the area of each polling place. As reported in Table 4, the results reveal substantial educational heterogeneity, in particular in the effect of *Margin*. Similar to the overall effect in Table 3, the effect of *Margin* is negative and statistically insignificant across the three lowest education groups. By contrast, voters in the best-educated areas (Quartile 4) are less likely to vote informally as a reaction to the competitiveness of their electorate, as predicted by theory (H1). Better-educated voters may be better than other groups at considering the strategic implications of their voting decisions, and may be less prone to bandwagon effects. These results support the hypothesis of Feddersen and Pessendorfer (1996), which is that uninformed voters abstain so that the election can be determined by informed voters: not only does overall informal voting decrease with education (as shown at the bottom of Table 4), but also, when the election is competitive, better-educated voters are more engaged while less well-educated voters are more likely to abstain.

By contrast, the effect of the number of options on the ballot is robust – consistently positive and statistically significant across all quartiles, and, interestingly, it shows no monotone relationship with education. This result supports the literature on mental processing costs but does not support a simple story based on voter error. That story would see better-educated voters more able to avoid mistakes, and so we should observe a smaller effect for them. The fact that the effect is strongest for Quartile 4 indicates that the time cost of research is likely also at play.

The effect of electorate change *per se* shows educational heterogeneity. The effect is negative and significant as before for Quartiles 1 and 2, but it becomes weaker for Quartile 3 and disappears for Quartile 4. This supports the view that the better-educated behave more in line with a rational decision-making model of voting, since an electorate change *per se* should not influence the net benefit of voting.

B. Robustness Checks

Table 5 confirms the robustness of our results by expanding the main results in two ways. First, it includes the results of the HS-DID-IV and DDD-IV specifications in Panels B and C, respectively, comparing these with the DID-IV results reported in Panel A. Second, in addition to the baseline results, the propensity score matched results, and the distance limited results in Columns (1)–(3), we include two alternative sets of control groups. The control group used in Column (4) is made up only of polling places in electorates that did not have any boundary changes that year. This addresses the concern that the control polling places that belong to the same electorate as the treated polling places may experience some form of treatment. In Column (5), by contrast, we restrict the control group to only polling places in electorates that had boundary changes. The advantage of this approach is that comparison is made between treated and control groups that are proximate and hence very similar. Panel B has only one column because its control groups comprise Senate observations from the treated polling places.

Across all models and data restrictions, the parameter estimates for ln(*N Options*) is positive and statistically significant. The range of estimates is fairly small, from a minimum of 2.62 to a maximum of 3.57, with the preferred baseline result being at the lower end of this range. Both *Margin* and ln(*Voters*) are not consistently statistically significant across specifications and vary in sign. In DDD-IV, the negative effect of *Changed Electorate* disappears.

In Appendix F, we further confirm the robustness of our results, by conducting a falsification test and other specification tests to address concerns regarding alternative explanations. In Appendix G, we show that an extended set of observable characteristics are generally not strongly associated with treatment. Appendix H presents the full result tables, including covariates.

C. Additional Analysis of Types of Informal Voting

We have found a robust, positive effect of the number of options on a ballot on informal voting across all education groups. This finding indicates that the effect is not only due to the cognitive demands of completing a large ballot but also to the time and mental processing cost of it, including prior research conducted into candidates. To corroborate this hypothesis, in this section we further analyze different types of informal voting.

There are a number of scenarios resulting in an informal vote. For example, a ballot left blank and one on which the voter writes their name are both recorded as informal. The AEC classifies informal votes into one of two categories: *intentionally informal* votes and *unintentionally informal* votes. The latter are defined as ballots with incomplete numbering, non-sequential numbering, and/or ticks and crosses, and those on which the voter is identified; all other informal votes are classified as intentionally informal (AEC 2016).[11] This classification is somewhat arbitrary because drawing a clear line between intentional and unintentional informal voting is difficult; paying close attention to avoid an unintentional mistake can be a conscious decision to some degree. While acknowledging this limitation, we follow the AEC's classification to further our understanding.

Data on the type of informal vote is not systematically gathered or available at a polling place level, which prevents us from conducting causal analysis.[12] The AEC has, however, undertaken reviews of informal voting at the electorate level. Using data on the elections of 2007, 2010, and 2013, we regress the percentage (0–100) of different types of informal votes on electorate-level characteristics and time period fixed effects. This analysis does not allow causal interpretation because it cannot exploit boundary changes as a source of causal

[11] The AEC classifies informal votes into one of the following categories: Totally blank; Incomplete numbering (number 1 only); Incomplete numbering (other); Ticks and crosses; Other symbols; Non-sequential numbering; Scribbles, slogans, and other protest vote marks; Illegible numbering; Voter identified; and Other.

[12] Communication with the AEC indicated that this data is also not available unpublished.

identification, but it nevertheless provides additional insight into how informal voting correlates with socioeconomic characteristics of electorates.

The results are shown in Table 6. As some AEC categories are uncommon, we aggregate the categories as follows: the Non-numeric category includes ballots with ticks, crosses, and symbols, while the Other category includes ballots where the voter is identified, where the vote is illegible, and the AEC's Other category.

The first column reports the results of an electorate-level regression for any type of informal voting. The level of informal voting is higher when the number of options on the ballot is larger and in electorates with higher mean income, higher unemployment, higher population density, lower population growth rate, lower education, and more residents whose primary language is not English.

The subsequent columns concern different types of informal voting. The margin and number of voters do not show a robust pattern with statistical significance. The number of options is positively and statistically significantly associated with four types of informal voting: higher levels of blank, non-sequential, incomplete informal votes, and the Other group, whereas negatively associated with one-only and non-numeric.

Turning to socioeconomic factors, English as a second language, tertiary education, and income are statistically significant determinants of many types of informal voting, and these three characteristics show similar patterns. Higher levels of English as a second language tend to lead to higher levels of informality across the board, but there is a notably strong correlation with the number of blank, one-only, non-sequential, and non-numeric ballots – likely reflecting errors or misunderstandings in completing the ballot. This is similar to the findings in Power and Roberts (1995) that recently enfranchised voters in Latin America are particularly likely to make errors on complex ballots. Galatas (2008) also finds that the percentage of immigrants is robustly positively correlated with the proportion of informal votes in Canada. Higher levels of

tertiary education tend to reduce almost all types of informal voting. While the correlation is strongest on the number of blank, one-only, and non-numeric ballots, education levels are weakly related to non-sequential numbering on ballots and incomplete numbering. As non-sequential and incomplete numbering tend to constitute unintentional informal voting, these results support the idea that better-educated voters may consider the strategic implications of their voting decisions more carefully than other groups. This interpretation is similar to results seen in Driscoll and Nelson (2014) and Cohen (2018), which both find that, in Latin America, voters with high levels of political literacy protest poor government performance by submitting blank ballots in compulsory elections. As with education level, a higher income is strongly correlated with the number of blank, one-only, and non-numeric ballots, but the signs are all opposite, indicating that, when controlling for the level of education, higher income increases the three types of informal voting. We do not have a clear explanation for this finding.

These results provide some important observations. The number of options on a ballot is associated with both intentional and unintentional informal voting. It is, moreover, positively correlated with non-sequential ballots, which are common in areas with more non-English speakers. However, the number of options is also positively correlated with blank ballots, a leading type of intentional informal voting. It is, additionally, positively associated with incomplete ballots but not with informal ballots containing scribbles (clearly stemming from an anti-democratic intention). These observations reveal the challenges and costs of voters having to complete larger ballots.

VI. Conclusion

We find support for the hypothesis that when more options are available on the ballot, informal voting will be higher (H3); however, we do not find support for the hypotheses that when the expected margin in an election is

bigger, informal voting will be higher (H1); nor that when there are more voters in an electorate, informal voting will be higher (H2).

Once a voter has incurred travel and time costs to reach a polling place, whether voters vote informally is largely driven by the time and mental processing costs of reflecting on the candidates (including research time in advance) and of ranking their options, rather than by their likelihood of being pivotal in the election. This is supported by non-causal analysis that finds that more options on the ballot is associated with higher levels of blank, non-sequential, and incomplete informal votes – all indicative of time and mental processing costs.

The response of voters with higher levels of education is, however, better aligned with existing theory, suggesting that better-educated voters consider the strategic implications of their voting decisions more carefully than other groups. We also find that the response of these voters to the number of candidates is similar to those with lower levels of education, indicating that behavior relating to the number of options is explained not only by differences in cognitive ability but also time and mental processing cost. This is supported by non-causal analysis which shows that voters in better-educated areas are less likely to submit unintentional informal votes.

From 2004–2016, around 32% of contests received more informal votes than those in the winning margin. Policies that reduce informal voting may, therefore, affect the final composition of Parliament and, ultimately, economic, social, and other policies. Our findings suggest that the rate of formal voting could be increased through strategies that make it easier for voters to research, understand, and rank candidates, such as allowing a non-exhaustive ranking of candidates for the House, as happens in some Australian state elections.

Do the insights in this paper apply to voluntary voting regimes such as those in the US? The existence of compulsory voting in Australia enables us to focus on the voting decision *sans* travel costs. While this facilitates clear interpretation, it means our results are identified from decisions by voters at the

margin of formal voting, not voters at the margin of turnout in a voluntary voting regime. The external validity of our results warrants future research.

Our results do, however, contain some implications beyond the Australian context, particularly for the US. A number of US jurisdictions are considering transitioning from first-past-the-post voting to instant-runoff voting, which requires voters to rank multiple candidates on a ballot.[13] The results in this paper demonstrate a potential downside to this shift – a likely increase in informal voting when a ballot is large.

References

ABC. 2020. *ABC Elections.* Accessed August 2018. http://www.abc.net.au/news/elections/.

ABS. 2016a. "1379.0.55.001 - National Regional Profile, 2010-14."

—. 2016b. "1270.0.55.001 - Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016."

—. 2018. *Federal electoral boundary GIS data for free download.* 10 July. Accessed August 2018. https://www.aec.gov.au/Electorates/gis/gis_datadownload.htm.

AEC. 2016. "Informal voting: 2016 House of Representatives elections." Australian Electoral Commission.

—. 2017. *Voter turnout: 2016 House of Representatives and Senate elections.* AEC. https://www.aec.gov.au/About_AEC/research/files/voter-turnout-2016.pdf.

—. 2018. *Official election results.* Accessed August 2018. https://results.aec.gov.au/.

[13] Currently, instant-runoff voting is only used in a small number of municipal elections in the US, but this number has grown significantly over the last ten years. Instant-runoff voting has shown recent success, endorsed by *The New York Times* (2018), and adopted for some elections in New York City.

Ansolabehere, Stephen, James M. Snyder Jr, and Charles Stewart III. 2000. "Old voters, new voters, and the personal vote: Using redistricting to measure the incumbency advantage." *American Journal of Political Science* 44 (1): 17-34.

Augenblick, Ned, and Scott Nicholson. 2015. "Ballot position, choice fatigue, and voter behaviour." *The Review of Economic Studies* 83 (2): 460-480.

Battaglini, M., R. B. Morton, and T. R. Palfrey. 2008. "Information aggregation and strategic abstention in large laboratory elections." *American Economic Review* 98 (2): 194-200.

Bidwell, Kelly, Katherine Casey, and Rachel Glennerster. 2020. "Debates: Voting and Expenditure Responses to Political Communication." *Journal of Political Economy* 128 (8).

Blais, Andre. 2006. "What affects voter turnout?" *Annu. Rev. Polit. Sci.* 9: 111-125.

Cancela, Joao, and Benny Geys. 2016. "Explaining voter turnout: A meta-analysis of national and subnational elections." *Electoral Studies* 42: 264-275.

Carson, Jamie L., and Michael H. Crespin. 2004. "The effect of state redistricting methods on electoral competition in United States House of Representatives races." *State Politics & Policy Quarterly* 4: 455-469.

Cohen, Mollie J. 2018. "Protesting via the Null Ballot: An Assessment of the Decision to Cast an Invalid Vote in Latin America." *Political Behavior* 40 (2): 395-414.

Cox, Gary W., and Jonathan N. Katz. 2002. *Elbridge Gerry's salamander: The electoral consequences of the reapportionment revolution.* Cambridge University Press.

Cunow, Saul. 2014. "More is Less (Representation): Choice Set Size, Information Acquisition, and Correct Voting in Multimember

Districts." *Montreal Voting Experiment Workshop, University of Montreal.*

De Paola, Maria, and Vincenzo Scoppa. 2014. "The causal impact of closeness on Electoral participation exploiting the Italian dual ballot system." *Public Choice* 160 (3-4): 467-479.

Desposato, Scott W., and John R. Petrocik. 2003. "The variable incumbency advantage: New voters, redistricting, and the personal vote." *American Journal of Political Science* 47: 18-32.

Downs, Anthony. 1957. "An economic theory of political action in a democracy." *Journal of Political Economy* 65:135-150.

Driscoll, Amanda, and Michael J. Nelson. 2014. "Ignorance or opposition? Blank and spoiled votes in low-information, highly politicized environments." *Political Research Quarterly* 67 (3): 547-561.

Fauvelle-Aymar, Christine, and Abel François. 2006. "The impact of closeness on turnout: An empirical relation based on a study of a two-round ballot." *Public Choice* 127 (3-4): 461-483.

Feddersen, T. J., and W. Pesendorfer. 1996. "The swing voter's curse." *The American Economic Review* 408-424.

Fraga, Bernard L. 2016. "Redistricting and the causal impact of race on voter turnout." *The Journal of Politics* 78: 19-34.

Funk, Patricia. 2010. "Social incentives and voter turnout: evidence from the Swiss mail ballot system." *Journal of the European Economic Association* 8 (5): 1077-1103.

Galatas, Steven. 2008. "None of the Above? Casting Blank Ballots in Ontario Provincial Elections." *Politics & Policy* 36 (3): 448-473.

Garmann, Sebastian. 2014. "A note on electoral competition and turnout in run-off electoral systems: Taking into account both endogeneity and attenuation bias." *Electoral Studies* 34: 261-265.

Gerber, A., M. Hoffman, J. Morgan, and C. Raymond. 2020. "One in a Million: Field Experiments on Perceived Closeness of the Election and

Voter Turnout." *American Economic Journal: Applied Economics* 12 (3): 287–325.

Geys, Benny. 2006. "Explaining voter turnout: A review of aggregate-level research." *Electoral studies* 25: 637-663.

Godefroy, R., and E. Henry. 2016. "Voter Turnout and Fiscal Policy." *European Economic Review 89*: 389-406.

Grillo, A. 2017. "Risk aversion and bandwagon effect in the pivotal voter model." *Public Choice* 172: 465–482.

Hayes, D., and S. C. McKee. 2009. "The participatory effects of redistricting." *American Journal of Political Science* 53 (4): 1006-1023.

Henderson, John A., Jasjeet S. Sekhon, and Rocio Titiunik. 2016. "Cause or effect? Turnout in Hispanic majority-minority districts." *Political Analysis* 24: 404-412.

Hill, L., and S. Young. 2006. "Cause and effect? Informal and compulsory voting in Australia." *Australian Journal of Political Science* 41 (3).

Ho, D., K. Imai, G. King, and E. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 199–236.

Horiuchi, Yusaku, and Alexandra Woodruff. 2017. "Randomized Ballot Order Can Increase Invalid Votes: Evidence from Australia."

Indridason, Indridi H. 2008. "Competition & turnout: the majority run-off as a natural experiment." *Electoral Studies* 27: 699-710.

Iyengar, S, and M Lepper. 2000. "When choice is demotivating: Can one desire too much of a good thing?" *Journal of Personality and Social Psychology* 79: 995-1006.

Jones, Daniel B., and Randall Walsh. 2018. "How Do Voters Matter? Evidence from US Congressional Redistricting." *Journal of Public Economics* 158: 25-47.

Kawai, Kei, Yuta Toyama, and Yasutora Watanabe. 2021. "Voter Turnout and Preference Aggregation." *American Economic Journal: Microeconomics* , forthcoming.

Lassen, D. D. 2005. "The effect of information on voter turnout: Evidence from a natural experiment." *American Journal of political science* 49 (1): 103-118.

Levine, David K., and Thomas R. Palfrey. 2007. "The paradox of voter participation? A laboratory study." *American Political Science Review* 101: 143-158.

Lijphart, Arend. 1998. "The problem of low and unequal voter turnout-and what we can do about it." *Political Science Series* 54: 1-13.

Lyytikainen, Teemu, and Janne Tukiainen. 2019. "Are Voters Rational?" *European Journal of Political Economy* 59: 230-242.

Matsusaka, J. G. 1993. "Election closeness and voter turnout: Evidence from California ballot propositions." *Public Choice* 76 (4): 313-334.

Nagler, Matthew G. 2015. "Trading off the benefits and costs of choice: Evidence from Australian elections." *Journal of Economic Behavior & Organization* 114: 1-12.

Palfrey, Thomas R., and Howard Rosenthal. 1985. "Voter participation and strategic uncertainty." *American Political Science Review* 79: 62-78.

Power, T., and J. C. Garand. 2007. "Determinants of invalid voting in Latin America." *Electoral Studies* 26 (2): 432-444.

Power, T., and Timmons Roberts. 1995. "Compulsory voting, invalid ballots, and abstention in Brazil." *Political Research Quarterly* 48 (4): 795-826.

Riker, William H., and Peter C. Ordeshook. 1968. "A Theory of the Calculus of Voting." *American political science review* 62: 25-42.

Schelker, M., and M. Schneiter. 2017. "The elasticity of voter turnout: Investing 85 cents per voter to increase voter turnout by 4 percent." *Electoral Studies* 49: 65-74.

Stockemer, Daniel. 2017. “What affects voter turnout? A review article/meta-analysis of aggregate research.” *Government and Opposition* 52: 698-722.

The New York Times. 2018. *A Congress for Every American*. https://www.nytimes.com/interactive/2018/11/10/opinion/house-representatives-size-multi-member.html.

Figures and Tables

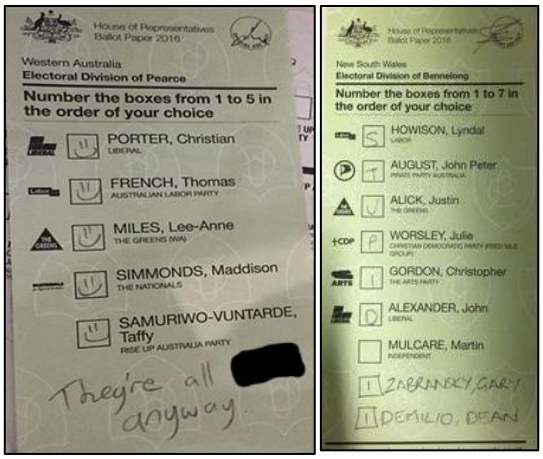

FIGURE 1. EXAMPLES OF INFORMAL VOTES

Source: Google Images

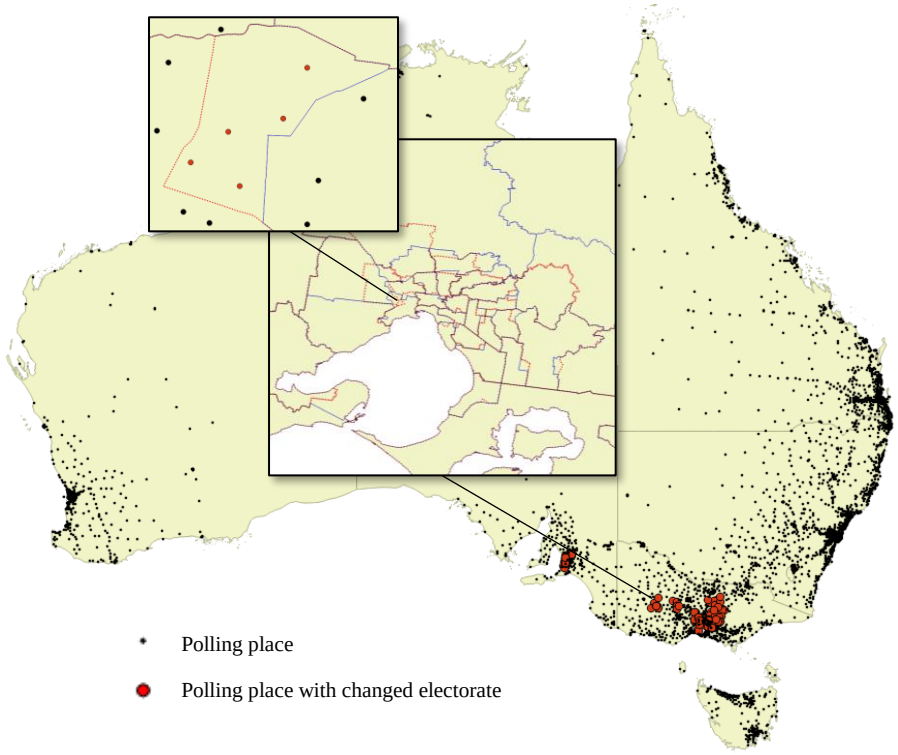

FIGURE 2. POLLING PLACES IN AUSTRALIA WITH INSETS OF BOUNDARY CHANGES

Notes: The main map shows polling places which did or did not switch electorate between the 2010 and 2013 elections. The insets show the electorate boundaries in western Melbourne for 2010 in solid, blue lines and the boundaries for 2013 in dashed, red lines.

Source: Authors' calculations based on AEC (2018).

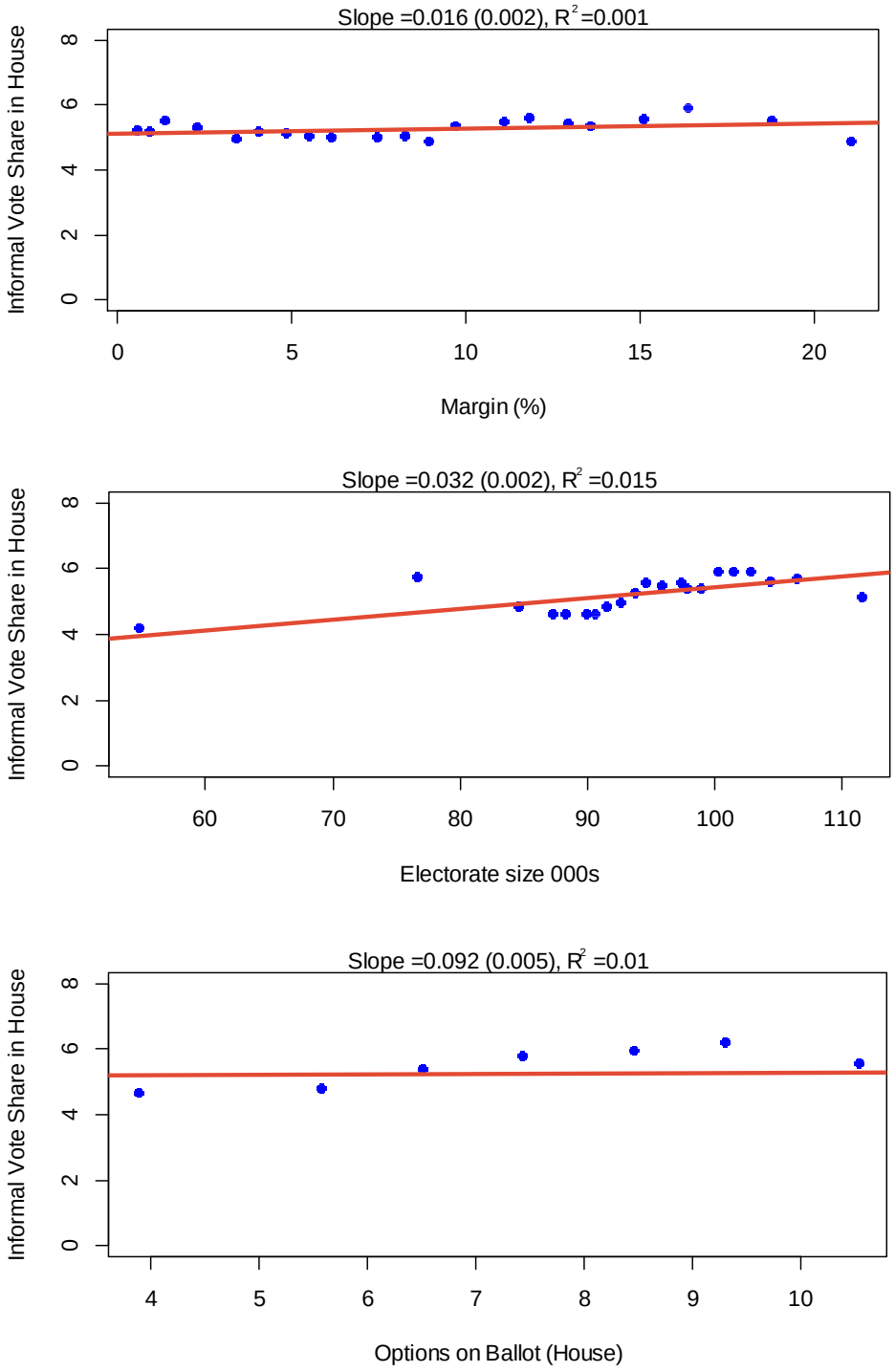

FIGURE 3. INFORMAL VOTING POSITIVELY CORRELATED WITH MARGIN, ELECTORATE SIZE, AND NUMBER OF OPTIONS ON BALLOT

Notes: Dots represent binned averages where each dot represents the same number of observations. The fitted lines are derived from simple OLS regressions, with the slope, standard error, and R^2 listed at the top of each panel.

Source: Authors' calculations based on AEC (2018).

TABLE 1. NUMBER OF POLLING PLACES AND ELECTORATES IN THE HOUSE SAMPLE

	2004	2007	2010	2013	2016
Panel A - All Data					
Polling Places	6,360	6,360	6,360	6,360	6,360
Changed electorate	NA	623	661	156	591
Did not change	NA	5,737	5,699	6,204	5,769
Electorates	150	150	150	150	150
Panel B - DID Data					
Polling Places		5,731	5,861	5,841	5,663
Changed electorate		418	426	156	415
Did not change		5,313	5,435	5,685	5,248
Electorates		143	143	143	143

Notes: 'All Data' includes all polling places used in every election. 'DID data,' a subset of 'All Data,' includes polling place observations used for DID analysis after sample selection.

Source: Authors' calculations based on AEC data (2018).

TABLE 2. SUMMARY STATISTICS FOR KEY VARIABLES

	DID Data		Treated		Control	
	Mean	SD	Mean	SD	Mean	SD
Informal (%)	5.28	2.57	5.97	2.98	5.23	2.54
Margin	0.09	0.06	0.09	0.06	0.09	0.06
ln(*Voters*)	11.48	0.10	11.49	0.09	11.48	0.10
ln(*N Options*)	1.96	0.35	1.88	0.30	1.97	0.35
Changed Electorate	0.06	0.24	1	0	0	0
Change in Margin	0.0002	0.02	0.004	0.07	0	0
Change in ln(*Voters*)	0.0001	0.01	0.001	0.05	0	0
Change in ln(*N Options*)	-0.001	0.08	-0.01	0.31	0	0
Covariates						
Median Age	39.64	5.16	39.17	5.24	39.67	5.15
Mean Income (000)	49.57	15.56	47.30	13.30	49.72	15.69
Unemployment (%)	5.60	2.10	5.88	2.21	5.58	2.10
Population Density	1,154.92	1,593.66	1,091.80	1,728.75	1,159.04	1,584.40
Population Growth (%)	1.32	1.98	1.43	2.31	1.31	1.96
Population Decline (%)	0.41	2.48	0.58	3.24	0.39	2.42
House Value (000)	445.24	563.72	385.64	387.45	449.13	573.14
English 2nd Language (%)	13.96	16.11	16.30	20.84	13.81	15.74
Tertiary Degree (%)	17.00	10.71	14.70	9.29	17.15	10.77
Observations	23,096		1,415		21,681	

Notes: Tertiary Degree refers to any post-high school degree. In Australia, this is often a 3-year university degree.

Source: Authors' calculations based on data from AEC (2018) and ABC (2020).

TABLE 3. MAIN RESULTS

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. Covariates	Fixed Effects	DID-IV Baseline	DID-IV Propensity Score Matched	DID-IV Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	2.177***	−0.217	−3.195***	−1.945	−3.334*	−2.569*
	(0.345)	(0.283)	(0.356)	(1.437)	(1.901)	(1.409)
ln(*Voters*)	3.534***	1.387***	0.837**	−2.950	−4.263	−1.476
	(0.192)	(0.156)	(0.351)	(2.046)	(2.899)	(2.128)
ln(*N Options*)	0.468***	0.858***	2.079***	2.732***	3.070***	2.717***
	(0.067)	(0.053)	(0.057)	(0.431)	(0.598)	(0.425)
Changed Electorate				−0.280***	−0.416***	−0.170**
				(0.067)	(0.104)	(0.068)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Clusters (Polling Place)	5,930	5,930	5,930	5,930	5,481	1,955
Observations (Polling Place × Year)	23,096	23,096	23,096	23,096	12,859	7,483
Treated Observations	NA	NA	NA	1,415	1,168	1,415
Control Observations	NA	NA	NA	21,681	11,691	6,068
R^2	0.030	0.432	0.354	0.344	0.350	0.400

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Column (1) shows the results of a linear regression without controls; Column (2) introduces a range of controls; and Column (3) introduces polling place fixed effects. Column (4) presents the baseline DID-IV, which uses voting in the House with treated polling places (those that switched electorate) and control polling places (those that did not). Column (5) shows the same model but with the propensity score matched sample. Column (6) shows results where the sample is limited to polling places within 2.5 kilometers of treated polling places.
* p<0.1; ** p<0.05; *** p< 0.01

TABLE 4. SUBGROUP ANALYSIS BASED ON QUARTILES OF TERTIARY EDUCATION LEVELS

	Dependent variable: Informal %			
	Quartile 1 (lowest)	Quartile 2	Quartile 3	Quartile4 (highest)
Margin	−2.766	−2.992	−3.405	6.650*
	(2.674)	(3.107)	(3.673)	(3.562)
ln(*Voters*)	−9.365	−2.492	0.406	−6.354
	(7.538)	(2.700)	(4.066)	(7.244)
ln(*N Options*)	3.431***	1.603*	2.326**	5.393***
	(1.070)	(0.883)	(0.909)	(1.037)
Changed Electorate	−0.252*	−0.428***	−0.149	−0.003
	(0.146)	(0.141)	(0.107)	(0.164)
Covariates	✓	✓	✓	✓
Time FE	✓	✓	✓	✓
Polling Place FE	✓	✓	✓	✓
Mean of Informal %	5.66	5.55	5.46	4.44
Observations (Polling Place × Year)	5,894	5,660	5,711	5,767
R^2	0.267	0.365	0.423	0.066

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Quartile 1 represents areas with the lowest percentage of people with tertiary degrees while Quartile 4 represents areas with the highest percentage. * p<0.1; ** p<0.05; *** p< 0.01

TABLE 5. ALTERNATIVE SPECIFICATIONS

	Dependent variable: Informal %				
	Baseline	Propensity Score Matched	Distance Limited	Controls: Non-Treated Electorates	Controls: Treated Electorates
	(1)	(2)	(3)	(4)	(5)
Panel A - DID-IV (House, treated and control polling places)					
Margin	−1.945	−3.334*	−2.569*	−1.806	−3.629*
	(1.437)	(1.901)	(1.409)	(1.648)	(2.070)
ln(*Voters*)	−2.950	−4.263	−1.476	−3.736*	−6.891
	(2.046)	(2.899)	(2.128)	(2.184)	(4.460)
ln(*N Options*)	2.732***	3.070***	2.717***	2.620***	2.885***
	(0.431)	(0.598)	(0.425)	(0.493)	(0.729)
Changed Electorate	−0.280***	−0.416***	−0.170**	−0.469***	0.034
	(0.067)	(0.104)	(0.068)	(0.078)	(0.143)
Observations (Polling Place × Year)	23,096	12,859	7,483	16,141	8,370
R^2	0.344	0.350	0.400	0.377	0.350
Panel B - HS-DID-IV (House and Senate, treated polling places only)					
Margin			0.275		
			(2.117)		
ln(*Voters*)			2.976		
			(4.926)		
ln(*N Options*)			3.344***		
			(0.727)		
Observations (Polling Place × Year × House)			2,830		
R^2			0.124		
Panel C - DDD-IV (House and Senate, treated and control polling places)					
Margin	−0.290	0.958	−0.445	−0.631	0.496
	(1.857)	(2.026)	(1.896)	(1.792)	(2.165)
ln(*Voters*)	2.603	0.536	2.500	2.378	3.122
	(4.469)	(2.374)	(4.522)	(4.262)	(5.042)
ln(*N Options*)	3.272***	3.571***	3.253***	3.230***	3.371***
	(0.659)	(0.552)	(0.663)	(0.631)	(0.740)
Changed Electorate	1.397	0.718	0.091	1.669	0.687
	(1.385)	(0.752)	(0.545)	(1.670)	(0.677)
Observations (Polling Place × Year × House)	46,192	25,696	14,966	32,282	16,740
R^2	−0.519	0.139	−0.021	−0.520	−0.180

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Time period fixed effects and polling place fixed effects are included for models in Panels A and C, while the model in Panel B include polling place-time fixed effects. All models include covariates. Column (1) presents the baseline results, Column (2) the propensity score matched results, Column (3) the distance limited results, Column (4) the results with control groups in non-treated electorates, and Column (5) the results with control groups in treated electorates.
* p<0.1; ** p<0.05; *** p< 0.01

TABLE 6. DETERMINANTS OF DIFFERENT TYPES OF INFORMAL VOTING (ELECTORATE LEVEL ANALYSIS)

	Dependent variable: Informal type share of total votes							
		Intentional		Unintentional				
	Total	Blank	Scribble	One only	Non-sequential	Non-numeric	Incomplete	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Margin	1.583	1.268*	0.467	−0.578	0.598	0.425	−0.559	−0.037
	(3.259)	(0.679)	(0.489)	(1.879)	(0.898)	(0.670)	(0.436)	(0.176)
ln(*Voters*)	−0.117	−0.021	0.111***	−0.133	−0.007	−0.079***	0.013	−0.0004
	(0.171)	(0.060)	(0.036)	(0.082)	(0.044)	(0.026)	(0.017)	(0.010)
ln(*N Options*)	2.479***	0.829***	−0.047	−0.565**	1.771***	−0.208*	0.582***	0.117***
	(0.579)	(0.137)	(0.076)	(0.271)	(0.241)	(0.121)	(0.090)	(0.029)
Median Age	−0.060	0.011	0.012	−0.037	−0.017	0.0001	−0.028***	−0.001
	(0.085)	(0.020)	(0.013)	(0.045)	(0.024)	(0.015)	(0.010)	(0.004)
ln(Mean Income (000))	5.585***	1.851***	0.051	2.812**	−0.794	2.067***	−0.177	−0.224*
	(1.994)	(0.419)	(0.246)	(1.257)	(0.634)	(0.468)	(0.241)	(0.127)
Unemployment (%)	0.428**	0.017	−0.004	0.322***	−0.039	0.106**	0.031	−0.006
	(0.198)	(0.041)	(0.024)	(0.106)	(0.052)	(0.041)	(0.023)	(0.009)
Population Density	0.001***	0.0002**	0.00003	0.0005***	−0.00004	0.0002***	0.00001	0.00002*
	(0.0003)	(0.0001)	(0.00004)	(0.0002)	(0.0001)	(0.0001)	(0.00004)	(0.00001)
Population Growth (%)	−0.305*	−0.060	0.130***	−0.328***	0.037	−0.141***	0.013	0.043***
	(0.185)	(0.040)	(0.032)	(0.108)	(0.059)	(0.041)	(0.023)	(0.011)
Population Decline (%)	−0.340	0.377	0.768**	−2.007**	0.838	−0.487	−0.059	0.230***
	(1.729)	(0.509)	(0.342)	(0.935)	(0.597)	(0.320)	(0.225)	(0.088)
ln(House Value (000))	0.874	−0.322*	−0.130	0.971**	0.313	−0.129	0.199*	−0.028
	(0.864)	(0.175)	(0.098)	(0.420)	(0.230)	(0.172)	(0.102)	(0.038)
English 2nd Language (%)	0.176***	0.039***	0.017***	0.056***	0.025***	0.027***	0.009***	0.003**
	(0.025)	(0.005)	(0.002)	(0.014)	(0.005)	(0.005)	(0.003)	(0.001)
Tertiary Degree (%)	−0.386***	−0.093***	−0.021***	−0.172***	−0.019	−0.074***	−0.010*	0.002
	(0.050)	(0.010)	(0.006)	(0.028)	(0.012)	(0.011)	(0.005)	(0.003)
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Share of informal (%)	100.0	23.6	16.1	28.1	13.9	11.3	4.5	2.6
Observations	429	429	429	429	429	429	429	429
R^2	0.687	0.748	0.567	0.517	0.468	0.589	0.499	0.316

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the electorate level. Categories reported by the AEC are further aggregated: the Non-numeric category includes ballots with ticks, crosses, and symbols while the Other category includes ballots where the voter is identified, where the vote is illegible, and the AEC's Other category. The dependent variable is defined as the number of a specific type of informal vote divided by the total number of votes (including informal) in each electorate. * p<0.1; ** p<0.05; *** p< 0.01

Appendix A – Details on the Australian Electoral System

Australian Elections

Australia is a federation of six states and two territories, with separate elections held for representatives at state and national levels. National elections, the subject of our study, are held roughly every three years. While each state has different rules for its own elections, rules for national elections are consistent across all states and territories.

The Parliament in Australia consists of two houses: The House of Representatives (abbreviated as House) and the Senate. The House has 150 members, with each member elected for a term of three years to represent a single geographic area (formally called an 'electoral division' but also referred to as an 'electorate' or 'seat'). Each electorate has a population of around 100,000 voters. The Senate has 76 senators; each state is represented by 12 Senators while each territory has two Senators. Senators serve for a term of six years, and at each election half of the Senate is contested. While each electorate has different House candidates, all electorates within a state or territory have the same Senate candidates.

Voting for the House and Senate normally takes place simultaneously. Voters arrive at a polling place in their local electorate, have their name marked off the electoral roll, and are issued with two voting papers (ballots): one for the House and one for the Senate. The voter fills out the two ballots, then submits both.

On both ballots, voters must rank their preferred choices. While they must rank all available House candidates, Senate voting does not require exhaustive ranking. Until 2016, voters could either indicate one preferred party or rank all candidates on the ballot. In 2016, this was changed to a requirement to rank at least six parties or at least 12 candidates. Examples of correctly filled out ballots for the 2016 House and Senate elections are provided in Figure A1. The candidates for each electorate are officially announced around 3–4 weeks before

election day. For example, in 2016 this announcement was made 22 days before the election.[14] The order of candidates on the ballot is randomized at the electorate level.

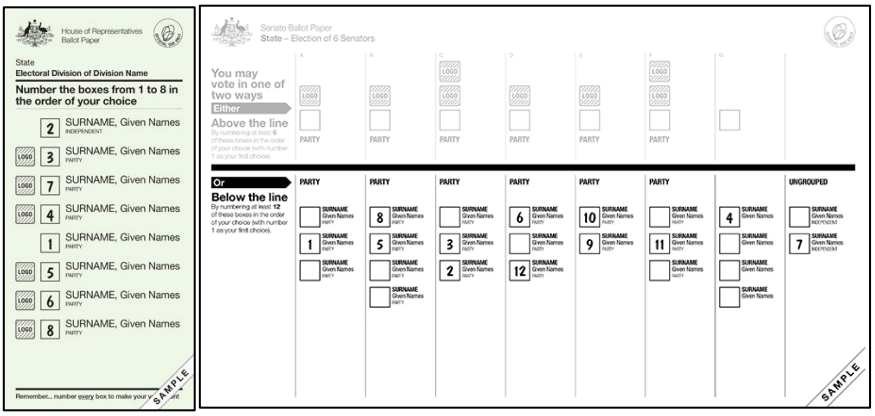

FIGURE A1. CORRECTLY COMPLETED HOUSE OF REPRESENTATIVES (LEFT PANEL) AND SENATE (RIGHT PANEL) BALLOTS

Notes: The Senate ballot is completed 'below the line' and according to the rules introduced in 2016, where below the line rankings do not need to be exhaustive.

Source: AEC (2016a).

For the House, the winner in each electorate is selected through a process known as full-preferential preference voting, or instant-runoff voting. In this process, the voter ranks all candidates in order of preference. Ballots are initially counted for each voter's first preference, after which the candidate with the lowest share is eliminated. Their votes are then redistributed to the candidate with second preference on the ballot. This process is repeated until one candidate has a majority of the vote.

For the Senate, each state and territory's representatives are selected using a single transferable vote system of proportional representation. In this system, votes are counted in a similar way to the House, but as each state and territory has multiple Senators, a Senator is elected once they receive a certain proportion of the overall vote (normally 14.3% for states and 33% for territories), rather than a majority.

[14] Dates for both the election and close of nominations are set out in the writ for the election, which is proposed by the Government and issued by the Governor General.

While data on the full preference listing for ballots is not available, some data on the flow of preferences in the counting process is made available by the AEC. This allows for a reconstruction of how often preferences are used in the counting of votes. Analysis of the data shows that, with roughly 13.5 million ballots counted in the 2016 election, approximately 10.5 million first preferences were upheld following the preference flow process. Further, while some 12.5 million voters (92% of the total) were obliged to rank five or more candidates, fifth or lower preferences came into play only in around 10,500 ballots (less than 0.1%). This suggests that the requirement to exhaustively rank all candidates on the ballot is rarely needed when votes are counted – yet it may contribute significantly (in both statistical and practical terms) to the level of informal voting.

Compulsory Voting

Voting in Australia is compulsory. Every Australian citizen aged 18 and over is required to be listed on the electoral roll and to vote in local government, state/territory, and national elections. Enrolment rates are high by international standards, with over 96% of eligible Australians enrolled (AEC 2019a).[15]

Failure to vote can attract fines ranging from $20–210 Australian Dollars (plus legal costs) and can result in a criminal conviction.[16] If a person who fails to vote does not pay the initial $20 fine or provide a valid and sufficient reason

[15] The accuracy of enrolment data is enhanced by data sharing between government agencies, and the AEC can directly enrol voters and update their details where sufficient information is available from other agencies. This is known as the Federal Direct Enrolment and Update (FDEU) program, which uses information from state and territory driver's licences, the Department of Human Services, and the Australian Tax Office (AEC 2019b). A voter cannot de-register themselves from the electoral roll, except in certain medical circumstances or in the case of permanent international relocation. Prisoners are required to be registered and to vote.

[16] Failure to vote at a federal election without a valid and sufficient reason is an offence under section 245 of the Commonwealth Electoral Act 1918 and initially carries a $20 penalty. The penalty is first issued by the AEC in a letter to the voter. The letter requires the voter to either pay the penalty or provide a valid and sufficient reason for not voting. Acceptable reasons include situations such as sickness, physical obstruction, natural disasters, personal accident, and urgent public duty, but do not include dislike of candidates or not preferring one candidate over another.

for not voting, the AEC may prosecute them and seek a penalty up to the current $210 maximum plus legal costs.[17]

In addition to voting in person on election day at a local polling place (or any polling place within the voter's state or territory), referred to as ordinary voting, other options are available. Voters can vote at selected polling places within their electorate prior to election day, which is known as pre-poll voting. They can also vote in a different state or territory, at an interstate voting center, or send their ballot in by post.[18]

As a result of compulsory voting, with its associated enforcement mechanisms, and options such as pre-polling and postal voting, turnout is relatively high in Australia – around 91% in 2016, although that represents a decline from an average of approximately 95% in the 1980s and 1990s (AEC 2017).

Informal Voting

Despite voting being compulsory, in order to satisfy the law, voters need only visit a polling place and have their name marked off the electoral roll. This means they can submit an empty or defaced ballot rather than a legitimate vote. This type of vote is recorded in the voting data as 'informal'. Postal voters must return their ballot, not merely request a ballot, in order to be marked as having voted. However, the ballot they return may still be informal. Informal voting can also occur as a result of misunderstanding of the voting rules, or of entering data incorrectly on the ballot (for example, marking two candidates as the first preference).

[17] If the non-voter decides not to pay the court fine, the court may impose penalties such as community service orders, seizure of goods, or jail.

[18] Voters can also apply to submit a postal vote, where the ballots are mailed to the voter's home, and the voter completes them in the presence of a witness and returns them via mail. Overseas voters can vote at the Australian embassy/consulate or via post. Both pre-poll and postal voting have been trending up over time, with postal voting increasing from around 5.0% in 2014 to around 8.5% in 2010. Pre-poll voting was introduced only in 2010 but has increased from around 7.0% of votes in 2010 to just under 20% in 2016 (Muller 2016). For federal elections, there is not any general availability of phone or internet-based voting (Lundie 2016).

Despite higher expenditure on political advertising and closer electoral outcomes, the rate of informal voting has been rising. In elections held between 2004–2016, it has accounted for around 5–6% of votes for the House and around 3–4% of votes for the Senate.

Informal votes are classed as either intentionally informal or unintentionally informal. While data on intentionality is not systematically gathered or available at the polling place level, a 2016 review provides data at the national level (AEC 2016b). As shown in Table A1, around 60% of informal votes for the House between 2001–2013 were classed as unintentional, while around 40% were classed as intentional, with intentional informal votes trending upwards over time. Table A2 shows the shares by type of informal vote. The most common type is where a voter ranks only one candidate on the ballot, accounting for an average of 28% of informal votes. The next most common type, averaging 24%, is a blank ballot.

TABLE A1. UNINTENTIONAL AND INTENTIONAL INFORMAL VOTES IN THE HOUSE OF REPRESENTATIVES 2001–2013

Year	Unintentional	Intentional	Total Informal
2001	3.2%	1.6%	4.8%
2004	3.2%	1.9%	5.1%
2007	2.5%	1.5%	3.9%
2010	2.8%	2.6%	5.5%
2013	3.6%	2.2%	5.8%

Notes: Unintentional informal votes are defined as ballots with incomplete numbering, non-sequential numbering, and/or ticks and crosses, and ballots in which the voter is identified; all other informal votes are classified as intentionally informal. Due to a significant methodological change from 2016, the AEC does not recommend comparison of 2016 figures with those of previous years (AEC 2018).

Source: Authors' calculations based on AEC data (2016b).

TABLE A2. TYPES OF INFORMAL VOTING AT THE ELECTORATE LEVEL (2007–2013)

	Mean number of votes	Mean share of informal votes (%)
One only	1,321.5	28.1
Blank	1,071.3	23.6
Scribble	696.3	16.1
Non-sequential	611.9	13.9
Non-numeric	528.8	11.3
Incomplete	216.6	4.5
Other	112.0	2.6
Total	4,558.4	100

Notes: The results above aggregate categories reported by the AEC. The Non-numeric category aggregates ballots with ticks, crosses, and symbols, while the Other category aggregates ballots where the voter is identified, where the vote is illegible, and the AEC's Other category.

Source: Authors' calculations based on AEC data (2009, 2011, and 2016c).

Australia's rate of informal voting is high compared to that of other OECD nations. For example, informal votes account for around 0.2% and 0.4% of all votes in the UK and US, respectively. Both these countries have first-past-the-post election systems, with voluntary voting. Other comparator countries such as Canada (0.7%), New Zealand (1.5%), and Japan (1.7%) are similarly low relative to Australia. Countries with compulsory voting tend to have higher rates of informality, more comparable to Australia's: e.g., Singapore (2.1%), Argentina (4.1%), Belgium (5.8%), and Brazil (16%) (International IDEA 2019).

Administration of Elections

National elections are administered by the Australian Electoral Commission (AEC). The AEC is a federal government agency that oversees the organizing, conducting, and supervising of federal elections and referendums. The AEC is also responsible for drawing electorate boundaries and conducting redistributions, maintaining the electoral roll, publishing official records of election results, fining and prosecuting voters who do not vote, monitoring the activities of registered political parties, and distributing public funding to political parties. The AEC's existence and role are an important piece of the institutional voting framework and are critical to the research design of this

paper. The AEC's structure means that many of the administrative aspects of voting are conducted independent of political considerations.

Changing Electorate Boundaries

There is a clear and established process for conducting a redistribution. The first step is the establishment of a Redistribution Committee for the relevant state or territory. The committee calculates the enrolment quota for each electorate, based on an estimate of the population divided by the number of House seats to which that state or territory is entitled. The committee then divides the state or territory into electorates, ensuring that the population of each electorate is as close to the enrolment quota as possible. This division process is legally required to consider factors such as economic, social, and regional communities of interest, means of communication and travel, physical features and area, and existing boundaries. The public can comment on the proposed boundaries, but once new boundaries are set, Parliament has no power to reject or amend them. This process is conducted with minimal and balanced political input, meaning that redistribution is exogenous to political considerations and past political outcomes.[19]

Following completion of a redistribution, the AEC must notify all voters whose electorate has changed as a result of the creation of a new electorate, the renaming of an existing electorate, or a change in boundaries of an existing electorate. Affected voters are notified by letter, and the changes are also announced in national and state/territory newspapers.

[19] Senior public servants are responsible for conducting the redistribution. Two political representatives are included in the panel that responds to any objections to the redistribution. Any objections, and the response, are thoroughly documented in the Report of the augmented Electoral Commission that is delivered before finalizing the redistribution.

Appendix B – Informal Voting and Turnout

There is a strong relationship between informal voting and turnout in Australia, shown in Figure B1, which uses electorate-level data for 2004–2016. Non-Vote, on the y-axis, is the percentage of enrolled voters who do not turn out to vote. The relationship between informal voting and non-voting is positive and statistically significant at conventional levels. This relationship remains in unreported regressions that include fixed effects for year and electorate, suggesting there may be consistent factors which affect both decisions.

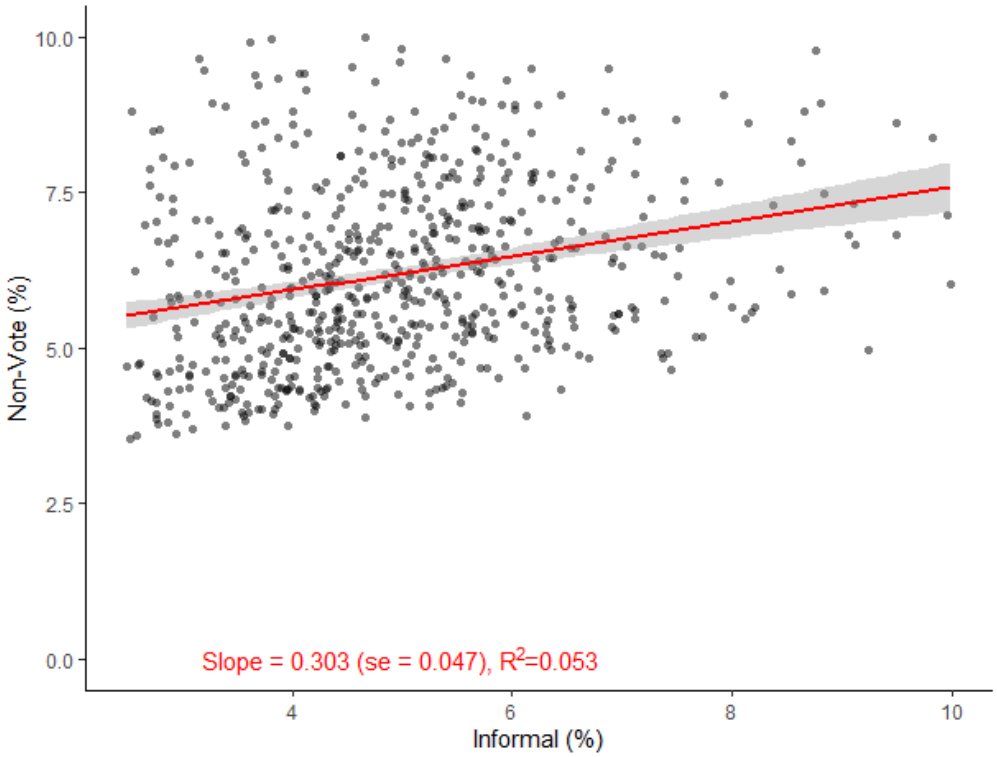

FIGURE B1. RELATIONSHIP BETWEEN INFORMAL VOTING AND TURNOUT

Notes: Data is at the electorate level, covering 2004–2016. Non-Vote is defined as 100 minus the turnout percentage recorded by the AEC. The coefficient of the slope remains statistically significant at the 5% level of significance in a regression with year and electorate fixed effects.

Source: Authors' calculations based on multiple AEC data sources.

Appendix C – Efficiency Gap Analysis

In order to measure gerrymandering, Stephanopoulos and McGhee (2015) propose a relevant statistic named the efficiency gap. The efficiency gap is defined as the absolute value of the difference between the parties' respective 'wasted votes', divided by the total number of votes cast in an election.[20] In gerrymandered electorates, the efficiency gap will be higher, since the goal of gerrymandering is, essentially, to maximize wasted votes for the opposition party and minimize wasted votes for one's own party.

Where information is available on vote margin and the final number of seats won by a party, Stephanopoulos and McGhee (2015) show that the efficiency gap calculation can be simplified into the formula $e = s - (2 \times v)$ where *e* is the efficiency gap, *s* is the seat margin (defined as the share of seats won by a party minus 50 percent), and *v* is the vote margin (defined as the share of votes received by a party minus 50 percent). Applying the simplified efficiency gap formula to the results of Australian House elections between 2001–2016 yields the estimated efficiency gaps shown in Table C1.

TABLE C1. OBSERVED EFFICIENCY GAP IN AUSTRALIAN ELECTIONS

Year	Efficiency Gap
2001	3.8%
2004	3.8%
2007	0.7%
2010	0.2%
2013	5.1%
2016	1.6%

Notes: Efficiency Gap is defined as in Stephanopoulos and McGhee (2015), essentially a measure of the difference in the proportion of 'wasted votes'.

Source: Authors' calculations based on Wikipedia data (2018).

[20] Stephanopoulos and McGhee (2015) define a wasted vote as the number of votes for the winning party more than required to win plus all votes for the losing party. Their paper provides detailed worked examples of the calculation.

The results indicate that, over this period, elections for the House of Representatives have seen an average efficiency gap of around 2.3% in favor of the more conservative political grouping, the Liberal-National Party Coalition. To put this into context, Jackman (2015) estimates that Wisconsin, a state where gerrymandering is suspected, has recorded efficiency gaps of between 10% and 13% in favor of the Republican Party.

Appendix D – Figures for Variables of Interest

Kernel density estimates of the rate of informal voting in the House and Senate are shown in Figure D1, below. Informal voting in the House has a higher valued mode and longer right tail than the Senate, likely due to the requirement to rank all candidates on the House ballot.

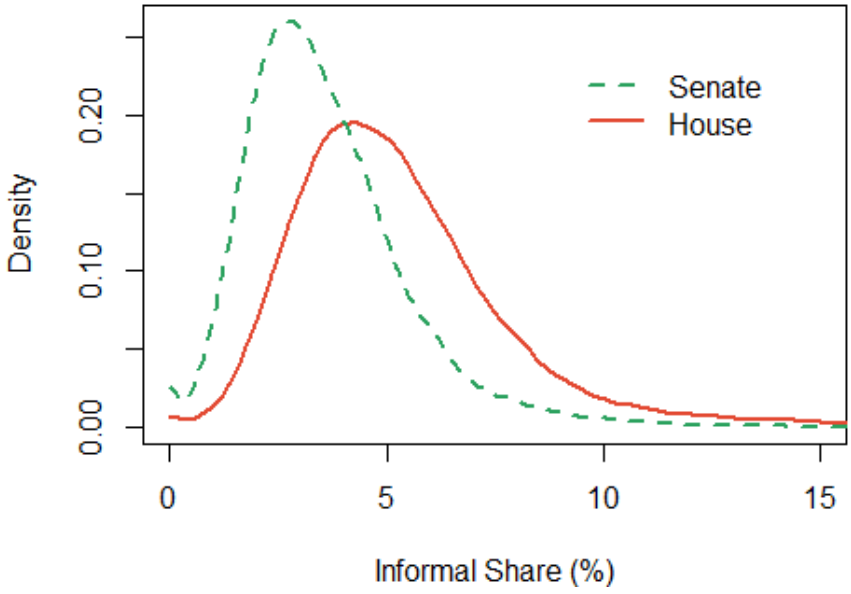

FIGURE D1. RATES OF INFORMAL VOTING IN THE HOUSE AND SENATE 2004–2016

Source: Authors' calculations based on AEC data (2018).

Figure D2 shows the same kernel density plots by year. The lower level of informal voting in 2007 is noteworthy. A potential explanation is that in the 2007 election a long-serving government was replaced by a new government, and so voters may have been more interested and engaged; however, the 2013 election also saw a change in government without an associated reduction in informality.

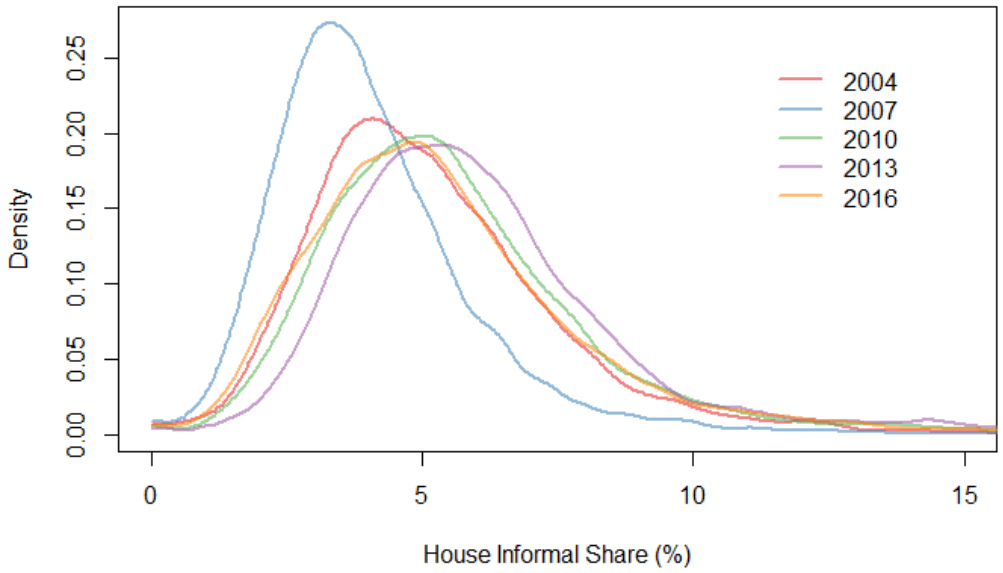

FIGURE D2. KERNEL DENSITY ESTIMATES OF INFORMAL VOTING SHARE IN THE HOUSE BY YEAR

Kernel density estimates for the margin over time are shown in Figure D3, demonstrating a consistency in the distribution of margin over time.

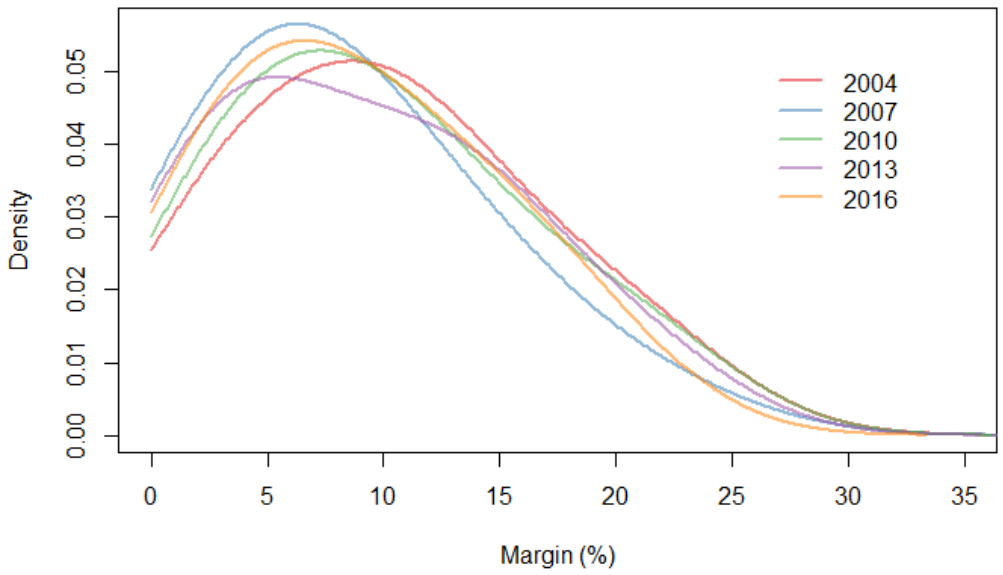

FIGURE D3. KERNEL DENSITY ESTIMATES OF MARGIN BY YEAR

The distribution of electorate sizes over time is shown in Figure D4. The size of electorates has been increasing over time, in tandem with population growth, but sizes are still generally grouped around the target of 100,000 voters.

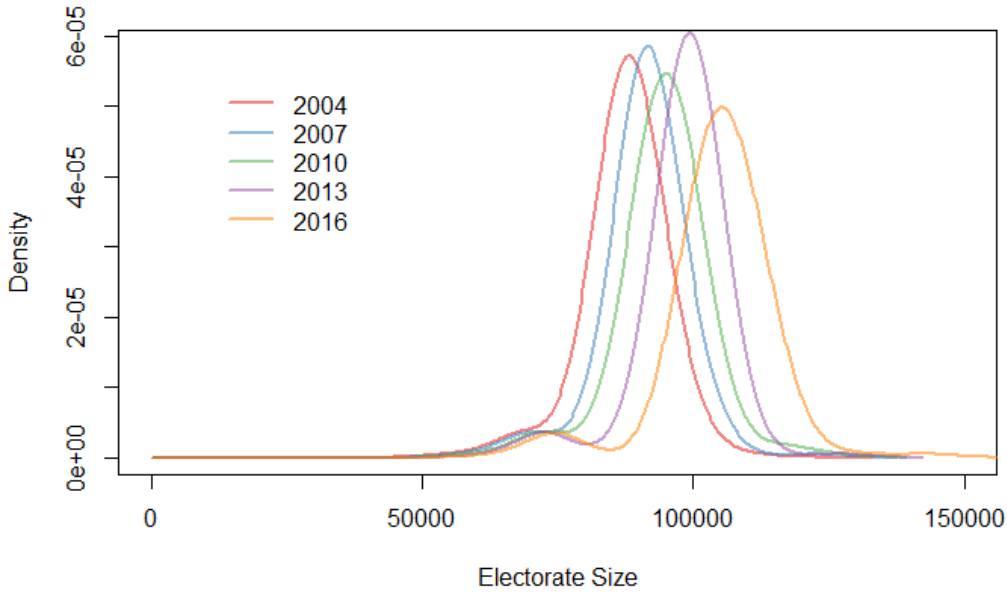

FIGURE D4. KERNEL DENSITY ESTIMATES OF ELECTORATE SIZE BY YEAR

The distribution of the number of options in the House is shown in Figure D5. The number is fairly consistent over time and generally sits between 5–10.

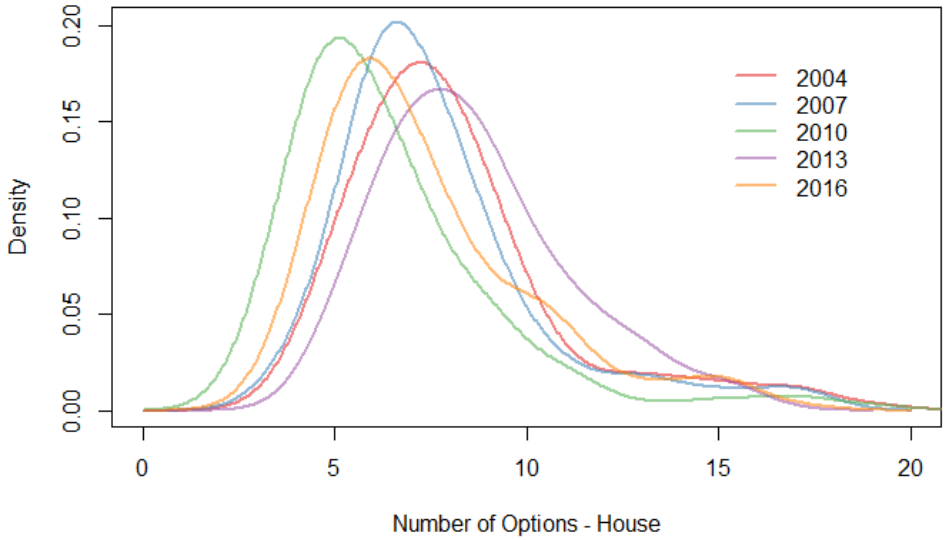

FIGURE D5. KERNEL DENSITY ESTIMATES OF THE NUMBER OF OPTIONS IN THE HOUSE BY YEAR

Figure D6 shows that, in total, around 32% of contests received more informal votes than votes in the margin between 2004–2016.

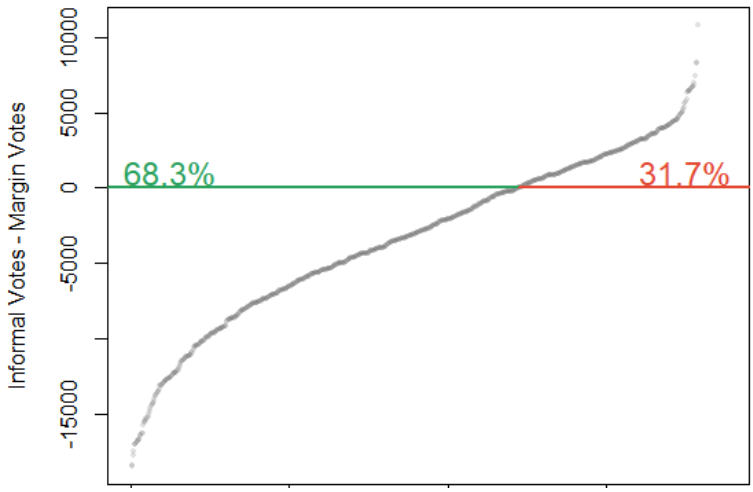

FIGURE D6. DIFFERENCE BETWEEN NUMBER OF INFORMAL VOTES AND MARGIN IN THE HOUSE

Notes: Data is at the electorate level, covering the years 2004–2016.

Source: Authors' calculations based on multiple AEC and ABC data sources (ABC 2020).

Appendix E – Details of Propensity Score Matching

The post-matching means for both the treated and control groups are presented in Table E1. While the treated and control groups show some small differences in control variables (see Table 2 in the main text), after propensity score matching the absolute value of the standardized mean difference is less than 0.1 for all variables. This satisfies the rule of thumb provided in Flury and Riedwyl (1986) for matching to successfully produce treated and control groups that are similar in observable characteristics. For the propensity score matching process, we use a single population growth variable that includes areas with both population growth and decline. This is different from the main regression where we split the population growth rate into two variables. We do this to ensure that the matching process accurately matches areas with similar growth rates.

TABLE E1. PRE AND POST MATCHING MEANS FOR TREATED AND CONTROL GROUPS

	Post Matching		
	Treated	Control	Std. Mean Diff
Median Age	38.44 (5.43)	38.42 (4.64)	0.00
Mean Income (000)	44.32 (13.71)	44.47 (14.23)	-0.01
Unemployment (%)	5.72 (2.2)	5.64 (2.29)	0.04
Population Density	1,016.56 (1,680.84)	998.16 (1,274.48)	0.01
Population Growth Rate (%)	1.04 (4.3)	1.08 (1.96)	-0.02
House Value (000)	333.05 (415.97)	332.91 (323.24)	0.00
English 2nd Language (%)	16.11 (20.98)	14.87 (17.26)	0.06
Tertiary Degree (%)	14.41 (9.48)	14.59 (8.99)	-0.02

Notes: For matching we use the nearest neighbor technique without replacement (Ho *et al*. 2007). Standard errors are in parenthesis. In all cases, the absolute value of the standardized mean difference is < 0.1, which satisfies the rule of thumb provided in Flury and Riedwyl (1986).

Appendix F –Further Robustness Checks

Visualizing the Common Trends Assumption

Undertaking matching also allows for a visual test of whether the common trends assumption is met in the data. For demonstration purposes, Figure F1 shows the rate of informal voting in the House over time across four particular groups: polling places that never moved during our study period; polling places that moved only in 2016 and experienced an *increase* in the number of options on the ballot; polling places that moved only in 2016 and experienced a *decrease* in the number of options on the ballot; and polling places that moved only in 2016 and experienced no change in the number of options. The Figure also includes 95% confidence intervals around each point. The observations for the 'Never Moved' group have been re-weighted based on a matching procedure similar to that described in Appendix E. For this Figure, polling places are matched based on observable characteristics in 2007, 2010, and 2013.

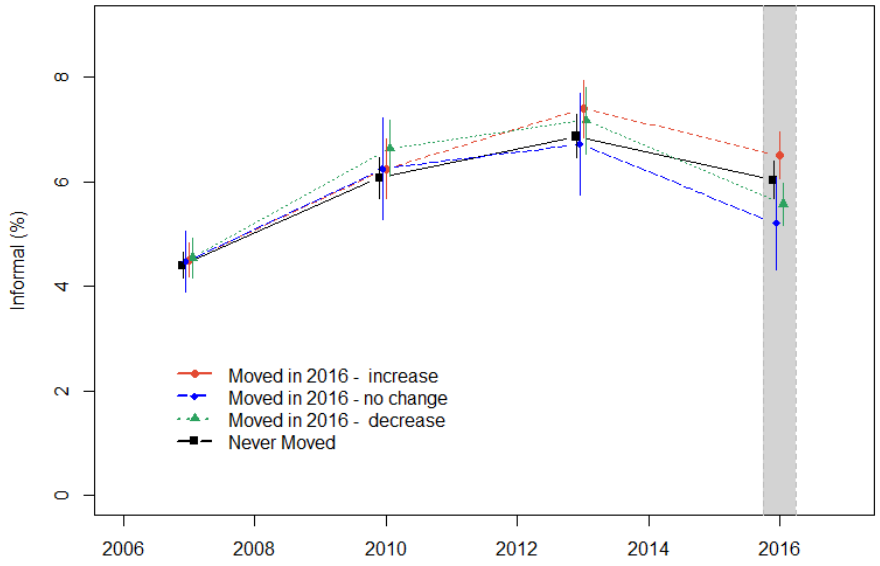

FIGURE F1. TIME TRENDS FOR GROUPS EXPERIENCING DIFFERENT CHANGES IN THE NUMBER OF OPTIONS ON THE BALLOT

Notes: Polling places in the 'increase' group saw an increase in the number of options on the ballot in 2016. 'No change' and 'decrease' groups are defined similarly. The 'Never Moved' group did not switch electorate at any time between 2007 and 2016. The Figure also includes 95% confidence intervals around each point. The 'Never Moved' group has been re-weighted using propensity score matching at the polling place level, using observable characteristics in 2007, 2010, and 2013.

Source: Authors' calculations based on AEC data (2018).

Visually, it appears that each group broadly follows the same time trend, with observations in 2007, 2010, and 2013 recording substantial overlap in confidence intervals between all four groups. In 2016, the group which saw an increase in the number of options appeared to experience higher levels of informal voting, while those in the 'No change' and 'decrease' groups saw lower levels. This visual analysis supports the common trends assumption and also demonstrates some of this paper's main findings.

Placebo Tests

To address the concern that our causal estimates are biased due to pre-existing differential trends in the error term, we conduct a placebo test in which the dependent variable is changed to variables for which no genuine treatment effect is expected. The variables selected are related to the election process but should not, theoretically, be affected by the margin, the number of voters, and the number of options available on the ballot: (1) the total number of votes recorded in the House, (2) the total number of votes recorded in the Senate, (3) the percentage of informal votes in the Senate in the previous election, (4) the percentage of Senate 'Donkey votes', defined as a voter voting for the first party on the ballot as their first preference, in the previous election, and (5) the share of votes for non-major parties (i.e., not the Liberals, Nationals, Labor, or the Australian Greens). None of these outcomes should be affected by an electorate change or by changes in the margin, number of voters, or number of options in the House.

Table F1 presents the results based on the baseline DID in Panel A and the propensity score matched DID in Panel B. The treatment is not found to be statistically significant at conventional levels in 34 of the 40 parameters. A statistically significant result is found for the relationship between *Changed Electorate* and Senate voting outcomes. There appears to be no ready theoretical explanation for this, and the estimated parameter values are small in magnitude.

Overall, this set of placebo tests provides support for the treatment effect estimated in the main results being a genuine causal effect.

TABLE F1. PLACEBO TEST

	Dependent variable:				
	Total House Votes	Total Senate Votes	Senate Informal $\%_{t-1}$	Senate Donkey $\%_{t-1}$	Senate Other $\%_{t-1}$
	(1)	(2)	(3)	(4)	(5)
Panel A - Baseline					
Margin	−61.992	−66.894	−0.008	0.006	0.033
	(200.876)	(200.325)	(0.009)	(0.008)	(0.029)
ln(*Voters*)	−143.103	−143.928	−0.003	−0.017	−0.019
	(313.100)	(311.992)	(0.013)	(0.012)	(0.041)
ln(*N Options*)	35.921	35.076	−0.003	−0.002	0.008
	(50.084)	(50.030)	(0.003)	(0.002)	(0.007)
Changed Electorate	6.001	6.122	−0.003***	0.003***	0.010***
	(10.358)	(10.368)	(0.0005)	(0.0004)	(0.001)
Covariates	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	23,096	23,096	23,051	23,051	23,051
R^2	0.106	0.106	0.261	0.061	0.622
Panel B - Propensity Score Matched					
Margin	83.402	84.354	−0.010	−0.009	−0.030
	(275.902)	(274.923)	(0.012)	(0.009)	(0.035)
ln(*Voters*)	251.543	259.219	0.008	−0.015	−0.044
	(432.408)	(430.482)	(0.019)	(0.014)	(0.058)
ln(*N Options*)	29.934	29.514	−0.002	0.001	0.010
	(66.126)	(66.156)	(0.004)	(0.002)	(0.009)
Changed Electorate	7.341	7.352	−0.003***	0.002***	0.004**
	(15.178)	(15.182)	(0.001)	(0.0004)	(0.002)
Covariates	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	12,859	12,859	12,859	12,859	12,859
R^2	0.106	0.106	0.285	0.021	0.688

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Each column reports the results with different dependent variables: the total number of votes recorded in the House, the total number of votes recorded in the Senate, the percentage of informal votes in the Senate, the percentage of Donkey votes in the Senate, and the share of votes for non-major parties (i.e., not the Liberals, Nationals, Labor, or the Australian Greens). Donkey voting is defined as a voter voting for the first party on the ballot as their first preference. * p<0.1; ** p<0.05; *** p< 0.01

Checking Omitted Causal Variables

Other variables not included in the main analysis might be contributing to informal voting and could be driving the estimated effect for the margin and number of options. One possibility is that the presence of a strong incumbent candidate might be correlated with the margin and number of options, while

affecting the likelihood of voters submitting an informal vote. The role of incumbency advantage in elections is frequently discussed in the literature. Ansolabehere *et al.* (2000), Friedman and Holden (2009), and Desposato and Petrocik (2003) all analyze incumbency advantage with reference to electorate boundaries, while Lee (2008) and Carson *et al.* (2007) consider the source of incumbency advantage more broadly.

Another potential omitted causal variable underlying our causal estimates is the share of the progressive votes. In Australia, high (low) shares of the progressive vote often lead to minor progressive (conservative) parties contesting the electorate, thereby increasing the number of parties and number of options on the ballot. For example, in the 2019 Federal election, out of the 10 electorates that were contested by a socialist part member, 9 were won by members of a progressive party – often by a significant margin.[21]

To account for the potential role of incumbency and progressive votes in informal voting, we estimated the same models including variables ln(*Tenure*), the number of years that the incumbent has represented the electorate at the time of the election, and *Progressive Vote Share*, progressive parties' share of the vote, defined as first preference votes for the Australian Labor and Australian Greens parties. Similarly to the DID-IV regressions, ln(*Tenure*) and *Progressive Vote Share* are instrumented by their exogenous changes due to electorate boundary changes.

As reported in Table F2, the inclusion of ln(*Tenure*) and *Progressive Vote Share* does not affect the overall conclusions relating to *Margin*, ln(*Voters*), or ln(*N Options*). In particular, the sign and magnitude of ln(*N Options*) remain similar to the main results. The scenarios described above are therefore unlikely to be driving our main results.

[21] Another path by which progressive vote share may cause concerns around omitted variable bias starts with the finding from Hill and Jones (2017) which shows that progressive parties spend more on minority groups relative to conservative parties, Pettersson-Lidbom (2008) find that progressive parties spend and tax 2–3% more than right-wing parties. These findings are important since Bechtel and Hainmueller (2011) show that voters respond to expenditure rises by increasing their vote for incumbents for at least two future elections.

TABLE F2. MAIN RESULTS – INCLUDING TENURE AND PROGRESSIVE SHARE

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. Covariates	Fixed Effects	DID-IV Baseline	DID-IV Propensity Score Matched	DID-IV Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	3.364***	−0.030	−3.547***	−2.057	−4.113*	−2.757*
	(0.355)	(0.285)	(0.361)	(1.557)	(2.299)	(1.498)
ln(*Voters*)	3.708***	1.400***	0.863**	−2.943	−4.412	−1.324
	(0.200)	(0.157)	(0.352)	(2.131)	(3.327)	(2.234)
ln(*N Options*)	0.730***	0.865***	2.155***	2.780***	2.864***	2.802***
	(0.067)	(0.054)	(0.057)	(0.444)	(0.655)	(0.437)
ln(*Tenure*)	−0.250***	−0.075***	−0.060***	0.005	0.039	−0.054
	(0.026)	(0.021)	(0.021)	(0.153)	(0.212)	(0.150)
Progressive Vote Share (%)	2.419***	0.162	1.245***	−0.467	−3.586	−0.927
	(0.136)	(0.115)	(0.123)	(2.419)	(2.810)	(2.059)
Changed Electorate				−0.282***	−0.419***	−0.174**
				(0.069)	(0.108)	(0.070)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Observations (Polling Place × Year)	23,096	23,096	23,096	23,083	12,859	7,470
R^2	0.057	0.433	0.360	0.337	0.293	0.390

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Column (1) shows the results of a simple linear regression; Column (2) introduces a range of socioeconomic covariates; and Column (3) introduces fixed effects for polling place. Columns (4)–(6) show the results of the DID-IV regressions. Column (4) presents the main output, which uses voting for the House, with treated polling places being those that switched electorate and control polling places being those that did not. Column (5) shows the same model but after propensity score matching has been applied to make the observable characteristics of treated and control groups similar, and Column (6) shows results where the sample is limited to polling places within 2.5 kilometers of treated polling places. We also attempt a variation of the above specification in which the change in the progressive vote share is split into an increase and a decrease variable. This variation does not materially affect the results presented above.

* p<0.1; ** p<0.05; *** p< 0.01

Alternative Definition of Margin

A potential concern with interpreting our results is related to the complexity of defining and understanding the margin in the instant-runoff voting system used for the House. In this system, the margin is defined based on the share of votes in the final round of vote counting, not on the share of first preference votes (*i.e.*, the number one preference marked on the ballot). We use this definition of the margin, assuming it to be a good proxy for the competitiveness of the election, but voters cannot easily calculate and respond to this margin when submitting their vote (since it requires calculation of the flow of preferences throughout the runoff process).

To address this concern, Table F3 presents a version of the main results where the margin is defined based on share of first preferences. The results in Table F3 are similar to the main results, except that *Margin* is not statistically significant in any specification. This result supports the view that the first preference margin is less relevant to the chances of a would-be pivotal voter. The effect of ln(*N Options*) remains positive and statistically significant, showing robustness to this change.

TABLE F3. MAIN RESULTS – MARGIN DEFINED ON FIRST PREFERENCES

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. Covariates	Fixed Effects	DID-IV Baseline	DID-IV Propensity Score Matched	DID-IV Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	0.898***	0.597***	0.238*	−0.150	−0.824	−0.498
	(0.150)	(0.125)	(0.140)	(0.649)	(0.767)	(0.635)
ln(*Voters*)	3.512***	1.203***	0.580*	−2.855	−4.756**	−1.388
	(0.193)	(0.158)	(0.301)	(1.796)	(2.256)	(1.856)
ln(*N Options*)	0.423***	0.888***	2.175***	2.924***	3.110***	2.927***
	(0.067)	(0.053)	(0.049)	(0.382)	(0.453)	(0.376)
Changed Electorate				−0.258***	−0.399***	−0.139**
				(0.057)	(0.078)	(0.057)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Clusters (Polling Place)	5,930	5,930	5,930	5,930	5,477	1,955
Observations (Polling Place × Year)	23,101	23,101	23,101	23,101	12,853	7,488
Treated Observations	NA	NA	NA	1,420	1,168	1,420
Control Observations	NA	NA	NA	21,681	11,685	6,068
R^2	0.030	0.433	0.349	0.336	0.346	0.389

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Column (1) shows the results of a simple linear regression; Column (2) introduces a range of socioeconomic covariates; and Column (3) introduces fixed effects for polling place. Columns (4)–(6) show the results of the DID-IV regressions. Column (4) presents the main output, which uses voting for the House, with treated polling places being those that switched electorate and control polling places being those that did not. Column (5) shows the same model but after propensity score matching has been applied to make similar the observable characteristics of treated and control groups; and Column (6) shows results where the sample is limited to polling places within 2.5 kilometers of treated polling places.
* p<0.1; ** p<0.05; *** p< 0.01

Sensitivity Analysis for Turnout

Our results may be driven not by voters on the margin between a decision to vote formally or informally, but by those on the margin between a decision to turn out or vote informally. If this proves to be the case, simplifying ballots in order to reduce informal voting might result only in lower voter turnout. To test this possibility, Table F4 shows a sensitivity analysis where informal votes are removed in order to increase the rate of non-voting by 1 percentage point. This mimics the removal of informal voters who are on the margin of not turning out.

The results are similar to those in the main results, and are also similar to additional, unreported sensitivity analyses based on a –1 and +2 percentage point change in non-voting. This sensitivity analysis suggests that our results are not driven by voters on the margin of not turning out to vote.

TABLE F4. SENSITIVITY ANALYSIS – INCREASING NON-VOTING BY 1% FROM INFORMAL VOTERS

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. Covariates	Fixed Effects	DID-IV Standard	DID-IV Propensity Score Matched	DID-IV Distance Limited
Margin	2.206***	−0.209	−3.241***	−1.900	−3.350**	−2.537**
	(0.348)	(0.285)	(0.309)	(1.249)	(1.440)	(1.221)
ln(*Voters*)	3.559***	1.397***	0.887***	−2.943*	−4.359*	−1.457
	(0.194)	(0.158)	(0.304)	(1.785)	(2.225)	(1.850)
ln(*N Options*)	0.469***	0.862***	2.090***	2.748***	3.102***	2.733***
	(0.067)	(0.054)	(0.049)	(0.375)	(0.456)	(0.368)
Changed Electorate				−0.279***	−0.427***	−0.168***
				(0.059)	(0.079)	(0.059)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Clusters (Polling Place)	5,930	5,930	5,930	5,930	5,477	1,955
Observations (Polling Place × Year)	23,096	23,096	23,096	23,096	12,853	7,483
Treated Observations	NA	NA	NA	1,415	1,168	1,415
Control Observations	NA	NA	NA	21,681	11,685	6,068
R^2	0.030	0.433	0.355	0.344	0.355	0.400

Notes: The data has been manually adjusted to artificially increase the rate of non-voting by 1 percentage point by reducing the number of informal votes. Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level Column (1) shows the results of a simple linear regression; Column (2) introduces a range of socioeconomic covariates; and Column (3) introduces fixed effects for polling place. Columns (4)–(6) show the results of the DID-IV regressions. Column (4) presents the main output, which uses voting for the House, with treated polling places being those that switched electorate and control polling places being those that did not. Column (5) shows the same model but after propensity score matching has been applied to make similar the observable characteristics of treated and control groups; and Column (6) shows results where the sample is limited to polling places within 2.5 kilometers of treated polling places. * p<0.1; ** p<0.05; *** p< 0.01

Appendix G – Selection into Treatment

Table G1 sets out the results of a regression where, in Column (1), the dependent variable is whether a polling place has changed electorate, and in Column (2), whether the polling place is in an electorate where any polling places have shifted. If this treatment indicator could be easily predicted by observable characteristics, the research design might be problematic. This is a concern despite the identification relying on the randomness of electorate changes *after controlling for those controls*, because the strong correlation of electorate changes with certain variables makes other relevant omitted variables highly plausible. The results indicate that observable characteristics are generally not strongly associated with treatment, and that, overall, the regression has low explanatory power, as measured by unadjusted and adjusted R^2. This supports use of the research design and of treating changes to electorate boundaries as an exogenous change.

TABLE G1. FIRST STAGE REGRESSIONS

	Dependent variable:	
	Polling Place Changed Electorate	Electorate had Polling Place change
	(1)	(2)
Median Age	−0.001	−0.005***
	(0.0004)	(0.001)
Mean Income (000)	0.0001	0.0002
	(0.0002)	(0.0003)
Unemployment (%)	−0.001	0.010***
	(0.001)	(0.002)
Population Density	0.00000	0.00000
	(0.00000)	(0.00000)
Population Growth (%)	0.001	0.006***
	(0.001)	(0.002)
Population Decline (%)	0.001	0.004***
	(0.001)	(0.001)
House Value (000)	0.00000	0.00002**
	(0.00000)	(0.00001)
English 2nd Language (%)	0.001***	0.0003
	(0.0001)	(0.0003)
Tertiary Degree (%)	−0.002***	−0.003***
	(0.0003)	(0.001)
Constant	0.103***	0.527***
	(0.018)	(0.036)
Observations	23,096	23,096
R^2	0.008	0.011
Adjusted R^2	0.007	0.010

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Column (1) shows the results of a linear probability model where the dependent variable is equal to one if the polling place changed electorate in a given year. Column (2) shows the results of a linear probability model where the dependent variable is equal to one if the polling place was in an electorate where any polling places changed electorate in a given year.

* p<0.1; ** p<0.05; *** p< 0.01

Appendix H – Detailed Model Outputs

This appendix includes full results for the main tables in the text. Table H1 provides additional summary statistics for key variables, particularly for the Senate.

TABLE H1. SUMMARY STATISTICS FOR KEY VARIABLES

	DID Data		Treated		Control	
	Mean	SD	Mean	SD	Mean	SD
House						
Informal (%)	5.28	2.57	5.97	2.98	5.23	2.54
Margin	0.09	0.06	0.09	0.06	0.09	0.06
ln(*Voters*)	11.48	0.10	11.49	0.09	11.48	0.10
ln(*N Options*)	1.96	0.35	1.88	0.30	1.97	0.35
ln(*Tenure*)	2.12	0.68	2.01	0.68	2.12	0.68
Progressive Vote Share (%)	0.52	0.16	0.53	0.18	0.52	0.16
Changed Electorate	0.06	0.24	1	0	0	0
Change in Margin	0.0002	0.02	0.004	0.07	0	0
Change in ln(*Voters*)	0.0001	0.01	0.001	0.05	0	0
Change in ln(*N Options*)	-0.001	0.08	-0.01	0.31	0	0
Change in ln(*Tenure*)	0.0002	0.24	0.003	0.97	0	0
Change in Progressive Vote Share (%)	0.01	0.03	0.10	0.09	0	0
Senate						
Informal (%)	3.52	1.83	3.97	2.26	3.50	1.79
Margin	0.12	0.10	0.12	0.09	0.12	0.10
ln(*Voters*)	14.40	0.64	14.69	0.43	14.38	0.65
ln(*N Options*)	3.38	0.37	3.39	0.28	3.38	0.38
Changed Electorate	0	0	0	0	0	0
Change in Margin	0	0	0	0	0	0
Change in ln(*Voters*)	0	0	0	0	0	0
Change in ln(*N Options*)	0	0	0	0	0	0
Covariates						
Median Age	39.64	5.16	39.17	5.24	39.67	5.15
Mean Income (000)	49.57	15.56	47.30	13.30	49.72	15.69
Unemployment (%)	5.60	2.10	5.88	2.21	5.58	2.10
Population Density	1,154.92	1,593.66	1,091.80	1,728.75	1,159.04	1,584.40
Population Growth (%)	1.32	1.98	1.43	2.31	1.31	1.96
Population Decline (%)	0.41	2.48	0.58	3.24	0.39	2.42
House Value (000)	445.24	563.72	385.64	387.45	449.13	573.14
English 2nd Language (%)	13.96	16.11	16.30	20.84	13.81	15.74
Tertiary Degree (%)	17.00	10.71	14.70	9.29	17.15	10.77
Observations	23,096		1,415		21,681	

Notes: Tertiary Degree refers to any post-high school degree. In Australia this is often a three-year university degree.

Source: Authors' calculations based on data from AEC (2018) and ABC (2020).

Table H2 presents the main results of the analysis including parameter estimates for all covariates except polling place fixed effects.

TABLE H2. MAIN RESULTS – FULL VERSION

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. Covariates	Fixed Effects	DID-IV Baseline	DID-IV Propensity Score Matched	DID-IV Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	2.177***	−0.217	−3.195***	−1.945	−3.334*	−2.569*
	(0.345)	(0.283)	(0.356)	(1.437)	(1.901)	(1.409)
ln(*Voters*)	3.534***	1.387***	0.837**	−2.950	−4.263	−1.476
	(0.192)	(0.156)	(0.351)	(2.046)	(2.899)	(2.128)
ln(*N Options*)	0.468***	0.858***	2.079***	2.732***	3.070***	2.717***
	(0.067)	(0.053)	(0.057)	(0.431)	(0.598)	(0.425)
Changed Electorate				−0.280***	−0.416***	−0.170**
				(0.067)	(0.104)	(0.068)
2010		1.602***	2.093***	2.336***	2.642***	2.591***
		(0.029)	(0.039)	(0.125)	(0.176)	(0.133)
2013		1.602***	2.046***	2.222***	2.432***	2.947***
		(0.041)	(0.073)	(0.173)	(0.287)	(0.206)
2016		0.777***	1.763***	2.323***	2.592***	2.533***
		(0.051)	(0.098)	(0.308)	(0.466)	(0.348)
Median Age		−0.032***	−0.010	−0.019	−0.017	0.092***
		(0.004)	(0.015)	(0.016)	(0.029)	(0.029)
Mean Income (000)		0.021***	−0.016***	−0.015***	−0.013	−0.043***
		(0.002)	(0.004)	(0.004)	(0.009)	(0.007)
Unemployment (%)		0.127***	−0.127	−0.071	−0.324	0.081
		(0.012)	(0.188)	(0.193)	(0.312)	(0.426)
Population Density		0.0002***	0.00003	0.00005	0.0002	0.0002
		(0.00003)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Population Growth (%)		−0.023**	0.001	0.002	0.017	0.027*
		(0.009)	(0.011)	(0.011)	(0.017)	(0.017)
Population Decline (%)		0.014**	−0.004	−0.003	0.024*	0.004
		(0.007)	(0.008)	(0.008)	(0.014)	(0.013)
House Value (000)		0.0001***	−0.00003	−0.00004*	−0.0001	−0.00003
		(0.00003)	(0.00002)	(0.00002)	(0.0001)	(0.00002)
English 2nd Language (%)		0.071***	−0.063***	−0.073***	−0.067**	−0.059***
		(0.003)	(0.013)	(0.016)	(0.026)	(0.020)
Tertiary Degree (%)		−0.129***	0.038	0.074	−0.041	0.150
		(0.003)	(0.075)	(0.077)	(0.117)	(0.105)
Constant	−36.408***	−12.874***				
	(2.151)	(1.764)				
Polling Place FE			✓	✓	✓	✓
Clusters (Polling Place)	5,930	5,930	5,930	5,930	5,481	1,955
Observations (Polling Place × Year)	23,096	23,096	23,096	23,096	12,859	7,483
Treated Observations	NA	NA	NA	1,415	1,168	1,415
Control Observations	NA	NA	NA	21,681	11,691	6,068
R^2	0.030	0.432	0.354	0.344	0.350	0.400

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Time period fixed effects use 2007 as the reference year.

* p<0.1; ** p<0.05; *** p< 0.01

Table H3 presents results when the data is subset according to quartiles of tertiary education.

TABLE H3. SUBGROUP ANALYSIS BASED ON QUARTILES OF TERTIARY EDUCATION LEVELS: WITH DETAILED COEFFICIENT ESTIMATES

	Dependent variable: Informal %			
	Quartile 1 (lowest)	Quartile 2	Quartile 3	Quartile 4 (highest)
Margin	−2.766	−2.992	−3.405	6.650*
	(2.674)	(3.107)	(3.673)	(3.562)
ln(*Voters*)	−9.365	−2.492	0.406	−6.354
	(7.538)	(2.700)	(4.066)	(7.244)
ln(*N Options*)	3.431***	1.603*	2.326**	5.393***
	(1.070)	(0.883)	(0.909)	(1.037)
Changed Electorate	−0.252*	−0.428***	−0.149	−0.003
	(0.146)	(0.141)	(0.107)	(0.164)
2010	2.898***	2.071***	2.280***	2.758***
	(0.434)	(0.171)	(0.244)	(0.375)
2013	2.634***	2.153***	2.132***	2.157***
	(0.542)	(0.382)	(0.315)	(0.541)
2016	3.472***	2.325***	1.733***	2.396***
	(1.183)	(0.496)	(0.600)	(0.892)
Median Age	−0.113**	0.007	−0.006	−0.082**
	(0.045)	(0.031)	(0.036)	(0.041)
Mean Income (000)	−0.008	−0.007	−0.015	−0.015**
	(0.012)	(0.012)	(0.012)	(0.007)
Unemployment (%)	−0.612	0.081	−0.590	0.748
	(0.426)	(0.288)	(0.544)	(0.532)
Population Density	−0.0004	0.002***	0.001**	0.00003
	(0.001)	(0.001)	(0.0003)	(0.0001)
Population Growth (%)	0.003	0.033	−0.013	−0.050
	(0.021)	(0.031)	(0.018)	(0.031)
Population Decline (%)	−0.009	0.047	0.046***	0.001
	(0.039)	(0.031)	(0.017)	(0.015)
House Value (000)	0.0001	0.0001	−0.0002	0.00003
	(0.0005)	(0.0004)	(0.0002)	(0.00004)
English 2nd Language (%)	−0.036	−0.063*	−0.079**	−0.074***
	(0.047)	(0.038)	(0.032)	(0.026)
Tertiary Degree (%)	0.120	−0.209	−0.142	0.352**
	(0.513)	(0.256)	(0.166)	(0.137)
Polling Place FE	✓	✓	✓	
Mean of Informal %	5.66	5.55	5.46	4.44
Observations (Polling Place × Year)	5,894	5,660	5,711	5,767
R^2	0.267	0.365	0.423	0.066

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Quartile 1 represents areas with the lowest percentage of people with tertiary degrees, while Quartile 4 represents areas with the highest percentage of people with tertiary degrees. Time period fixed effects use 2007 as the reference year. * p<0.1; ** p<0.05; *** p< 0.01

Table H4 and H5 show alternative versions of the main results that include both alternative control groups (Columns (4) and (5)) and different model specifications.

TABLE H4. ALTERNATIVE MODEL SPECIFICATIONS – PANEL A – DID-IV (HOUSE; TREATED AND CONTROL POLLING PLACES) : WITH DETAILED COEFFICIENT ESTIMATES

	Dependent variable: Informal %				
	Model specifications				
	Baseline	Propensity Score Matched	Distance Limited	Controls: Non-Treated Electorates	Controls: Treated Electorates
	(1)	(2)	(3)	(4)	(5)
Margin	−1.945	−3.334*	−2.569*	−1.806	−3.629*
	(1.437)	(1.901)	(1.409)	(1.648)	(2.070)
ln(*Voters*)	−2.950	−4.263	−1.476	−3.736*	−6.891
	(2.046)	(2.899)	(2.128)	(2.184)	(4.460)
ln(*N Options*)	2.732***	3.070***	2.717***	2.620***	2.885***
	(0.431)	(0.598)	(0.425)	(0.493)	(0.729)
Changed Electorate	−0.280***	−0.416***	−0.170**	−0.469***	0.034
	(0.067)	(0.104)	(0.068)	(0.078)	(0.143)
2010	2.336***	2.642***	2.591***	2.186***	2.812***
	(0.125)	(0.176)	(0.133)	(0.131)	(0.286)
2013	2.222***	2.432***	2.947***	2.130***	
	(0.173)	(0.287)	(0.206)	(0.185)	(0.000)
2016	2.323***	2.592***	2.533***	2.316***	3.280***
	(0.308)	(0.466)	(0.348)	(0.293)	(0.909)
Median Age	−0.019	−0.017	0.092***	−0.010	0.036
	(0.016)	(0.029)	(0.029)	(0.022)	(0.068)
Mean Income (000)	−0.015***	−0.013	−0.043***	−0.013**	−0.015
	(0.004)	(0.009)	(0.007)	(0.006)	(0.014)
Unemployment (%)	−0.071	−0.324	0.081	−0.084	−0.029
	(0.193)	(0.312)	(0.426)	(0.246)	(0.568)
Population Density	0.00005	0.0002	0.0002	0.00003	−0.0001
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0003)
Population Growth (%)	0.002	0.017	0.027*	0.006	0.012
	(0.011)	(0.017)	(0.017)	(0.017)	(0.031)
Population Decline (%)	−0.003	0.024*	0.004	−0.0005	−0.006
	(0.008)	(0.014)	(0.013)	(0.011)	(0.017)
House Value (000)	−0.00004*	−0.0001	−0.00003	−0.00002	−0.0002**
	(0.00002)	(0.0001)	(0.00002)	(0.00004)	(0.0001)
English 2nd Language (%)	−0.073***	−0.067**	−0.059***	−0.085***	−0.027
	(0.016)	(0.026)	(0.020)	(0.021)	(0.039)
Tertiary Degree (%)	0.074	−0.041	0.150	0.163	−0.260
	(0.077)	(0.117)	(0.105)	(0.099)	(0.222)
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	23,096	12,859	7,483	16,141	8,370
R^2	0.344	0.350	0.400	0.377	0.350

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Time period fixed effects use 2007 as the reference year. Propensity score matching is applied to Column (2) .

* p<0.1; ** p<0.05; *** p< 0.01

TABLE H5. ALTERNATIVE MODEL SPECIFICATIONS – PANEL C – DDD-IV (HOUSE AND SENATE; TREATED AND CONTROL POLLING PLACES) : WITH DETAILED COEFFICIENT ESTIMATES

	Dependent variable: Informal %				
	Model specifications				
	Baseline	Propensity Score Matched	Distance Limited	Controls: Non-Treated Electorates	Controls: Treated Electorates
	(1)	(2)	(3)	(4)	(5)
Margin	−0.290	0.958	−0.445	−0.631	0.496
	(1.857)	(2.026)	(1.896)	(1.792)	(2.165)
ln(*Voters*)	2.603	0.536	2.500	2.378	3.122
	(4.469)	(2.374)	(4.522)	(4.262)	(5.042)
ln(*N Options*)	3.272***	3.571***	3.253***	3.230***	3.371***
	(0.659)	(0.552)	(0.663)	(0.631)	(0.740)
Changed Electorate	1.397	0.718	0.091	1.669	0.687
	(1.385)	(0.752)	(0.545)	(1.670)	(0.677)
Senate	−13.972	−8.553	−14.958	−12.996	−16.307
	(13.579)	(7.030)	(14.585)	(12.622)	(16.154)
Senate × 2016	0.293	0.036	0.436	0.605	−0.427
	(0.656)	(0.455)	(0.561)	(0.868)	(0.283)
Polling Place × Year FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year × House)	46,192	25,696	14,966	32,282	16,740
R^2	−0.519	0.139	−0.021	−0.520	−0.180

Notes: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level. Time period fixed effects use 2007 as the reference year. Propensity score matching is applied to Column (2).

* p<0.1; ** p<0.05; *** p< 0.01

Appendix References

ABC. 2020. *ABC Elections.* http://www.abc.net.au/news/elections/.

AEC. 2009. *Analysis of Informal Voting House of Representatives 2007 Election* .

—. 2011. *Analysis of Informal Voting House of Representatives, 2010 Federal Election.*

—. 2016a. *How to vote.* 29 November. https://www.aec.gov.au/Voting/How_to_Vote/.

—. 2016b. *Informal voting: 2016 House of Representatives elections.* Australian Electoral Commission.

—. 2016c. *Analysis of informal voting 2013 House of Representatives elections.*

—. 2017. *Voter turnout: 2016 House of Representatives and Senate elections.* AEC. https://www.aec.gov.au/About_AEC/research/files/voter-turnout-2016.pdf.

—. 2018. *Analysis of informal voting: 2016 House of Representatives elections.* https://www.aec.gov.au/About_AEC/research/files/analysis-of-informal-voting-2016.pdf.

—. 2019a. *Size of the electoral roll and enrolment rate 2019.* https://www.aec.gov.au/Enrolling_to_vote/Enrolment_stats/national/2019.htm.

—. 2019b. *Direct Enrolment and Update.* https://www.aec.gov.au/Enrolling_to_vote/About_Electoral_Roll/direct.htm.

Ansolabehere, Stephen, James M. Snyder Jr, and Charles Stewart III. 2000. "Old voters, new voters, and the personal vote: Using redistricting to measure the incumbency advantage." *American Journal of Political Science* 17-34.

Bechtel, Michael M., and Jens Hainmueller. 2011. "How lasting is voter gratitude? An analysis of the short-and long-term electoral returns to beneficial policy." *American Journal of Political Science* 55 (4): 852-868.

Carson, Jamie L, Erik J. Engstrom, and Jason M. Roberts. 2007. "Candidate quality, the personal vote, and the incumbency advantage in Congress." *American Political Science Review* 101 (2): 289-301.

Desposato, Scott W., and John R. Petrocik. 2003. "The variable incumbency advantage: New voters, redistricting, and the personal vote." *American Journal of Political Science* 47: 18-32.

Flury, Bernhard K, and Hans Riedwyl. 1986. "Standard distance in univariate and multivariate analysis." *The American Statistician* 40 (3): 249-251.

Friedman, J. N., and R. J. Holden. 2009. "The rising incumbent reelection rate: What's gerrymandering got to do with it?" *The Journal of Politics* 71 (2): 593-611.

Hill, Andrew J., and Daniel B. Jones. 2017. "Does partisan affiliation impact the distribution of spending? Evidence from state governments' expenditures on education." *Journal of Economic Behavior & Organization* 143: 58-77.

Ho, D., K. Imai, G. King, and E. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 199–236.

International IDEA. 2019. *Invalid Votes.* https://www.idea.int/data-tools/question-view/443.

Jackman, Simon. 2015. "Assessing the Current Wisconsin State Legislative Districting Plan." Expert Analysis by Political Scientist, Stanford, CA: Stanford University.

Lee, D S. 2008. "Randomized experiments from non-random selection in US House elections." *Journal of Econometrics* 142 (2): 675-697.

Lundie, Rob. 2016. *Electronic voting at federal elections.* https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook45p/ElectronicVoting.

Muller, Damon. 2016. *The 2016 federal election.* https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook45p/FederalElection2016.

Pettersson-Lidbom, Per. 2008. "Do parties matter for economic outcomes? A regression-discontinuity approach." *Journal of the European Economic Association* 6 (5): 1037-1056.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2015. "Partisan gerrymandering and the efficiency gap." *The University of Chicago Law Review* 831-900.

Wikipedia. 2018. *List of Australian federal elections.* https://en.wikipedia.org/wiki/List_of_Australian_federal_elections.