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

Preliminary and incomplete
2020-07-20

Current Version: `http://eamonmcginn.com/papers/Informal_Voting.pdf`.

By Shiko Maruyama*and Eamon McGinn†

Voter turnout tends to range from 50-70% globally. Low, for something as critical as electing policymakers, but much higher than basic economic theory would predict. In Australia, where voting is compulsory, there are still around 5-6% of voters who submit an informal vote, which does not count towards the total. We make use of a natural experiment, based on exogenous changes in electorate boundaries, to identify what factors influence the number of wasted votes. In an advance from the existing literature, we can test a range of potential contributing factors together, in a single model, and with causal interpretation. We find that factors that feature in the traditional theory on voter decisions, competitiveness and number of other voters, do not affect the rate of informal voting. Instead we find that more candidates on the ballot results in higher levels of informal voting. Halving the number of options would reduce informal voting by 15%. This effect is present regardless of the level of education, indicating it is likely a decision to abstain rather than an error. We conclude that making the way voters can express their preferences as simple as possible would be beneficial for reducing wasted votes.
JEL: C21, D72, D73, D91, H11, P16
Keywords: informal voting, voter participation, rational voting, choice overload, compulsory voting, difference in differences

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

† McGinn (Corresponding Author): Univeristy of Technology Sydney, UTS Business School PO Box 123, Broadway, NSW 2007, Australia, eamon.g.mcginn@student.uts.edu.au). Acknowledgments: This research is supported by an Australian Government Research Training Program Scholarship. Thanks to Peter Siminski, Kentaro Tomoeda, Shiko Maruyama and Mario Fiorini

In Australia, which is one of the few developed countries where voting is compulsory, around 5-6% of votes in recent elections are determined to be "informal", because they are empty or improperly filled in, and so don't count towards the outcome of the election.

Why would a voter who has already travelled to the polls and queued up to get their ballot waste their vote? Wasted votes are a major concern as high rates can cause concerns about the legitimacy of a government (Lijphart, 1998) and can disproportionately affect disadvantaged citizens leading to unequal influence (Kawai, Toyama and Watanabe, 2020). This is of concern as, from 2004-2016, around 32% of contests in Australia had more informal votes than the margin of victory.

This paper offers a better understanding of voting behavior by making use of a natural experiment that arises in the Australian electoral system. In Australia, an independent commission makes changes to the boundaries of electoral divisions. These boundary changes result in some voters experiencing a drastic exogenous change in the nature of the election – competitiveness, number of voters and number of candidates all change. This exogenous change allows us to quantify the causal effects of three important factors contributing to the rate of informal votes: the competitiveness of the electoral division, the number of voters in the electoral division, and the number of options shown on the ballot. This paper also contributes to the literature on voter turnout as the identification approach means we can abstract from travel costs to quantitatively compare alternative hypotheses on voter behavior, which isn't possible in other papers that focus on voter turnout in other contexts. Analysing the impacts of travel costs on turnout has been done causally in papers such as Funk (2010) Godefroy and Henry (2016), Schelker and Schneiter (2017) and non-casually in papers such as Gimpel and Schuknecht (2003) and Haspel and Knotts (2005). These papers generally find that there is an increase in turnout when costs are decreased.

We focus on these factors because the empirical literature on voter turnout, which generally builds on theoretical work from Downs (1957), Riker and Ordeshook (1968) and Palfrey and Rosenthal (1985), generally identifies three key determinants of voter turnout and wasted votes: a voter is more likely to vote in a smaller electoral division, when there is expected to be a close margin between the winner and the runner up, and when travel and time costs are low (Levine and Palfrey, 2007). The general finding in the empirical literature is support, although mixed, for the expected relationships from the theory [see literature reviews by Geys (2006), Blais (2006), Cancela and Geys (2016) and Stockemer (2017)]. These results are generally confirmed by papers with a focus on causal analysis such as Fauvelle-Aymar and François (2006), Indridason (2008), De Paola and Scoppa (2014) and Garmann (2014) which all find that competitiveness increases turnout; and Lyytikäinen and Tukiainen (2019) find that the number of voters in an electoral division decreases turnout. In contrast, Matsusaka (1993), who uses use of

different ballot propositions on the same ballot, finds that there is no relationship between closeness and turnout.

An alternative explanation for voter turnout and informal voting behavior is that voters are averse to the mental processing costs associated with correctly completing a ballot. Horiuchi and Lange (2019), Cunow (2014) and Augenblick and Nicholson (2016) show that, when presented with more candidates, voters can abstain from voting, reduce the dimensionality of the issue space, and make voting errors.

In a completely different approach, Feddersen and Pesendorfer (1996), build a theoretical model where uninformed voters may choose to abstain in order to allow informed voters to control the outcome of the election. This model has received empirical support in lab experiments (Battaglini, Morton and Palfrey, 2008) and in natural experiments (Lassen, 2005).

There is also considerable disagreement on what causes informal voting. In causal analysis, De Paola and Scoppa (2014) find that closeness has no effect on informal voting, this is different to non-causal studies such as Galatas (2008) which do find the expected relationship between closeness and informal voting. In non-causal analyses, Power and Garand (2007) find that social, personal and institutional factors contribute to the level of informal voting in South America. In contrast, Driscoll and Nelson (2014) attribute informal voting in South America primarily to political concerns. In Australia, Hill and Young (2006) conclude that informal voting is related to complexity of voting and high levels of voter turnout due to compulsory voting. This align with analysis by Nagler (2015), which uses time series data, and shows that more candidates on the ballot is associated with more informal votes and also with causal analysis of Swiss elections by Hoffman, León and Lombardi (2017) who find that compulsory voting increases the level of informal voting.

We make use of a natural experiment, based on exogenous changes in electorate boundaries, to identify what factors influence the decision to waste a vote. In an advance from the existing literature, we are able to test a range of potential contributing factors together, in a single model, and with causal interpretation. Directly testing and comparing these different pathways is important because they have different implications for policies to increase voting and enhance the functioning of the electoral system.

We find that factors that feature in the traditional theory on voter decisions, competitiveness and number of other voters, do not affect the rate of informal voting. Instead we find that more candidates on the ballot results in higher levels of informal voting. Using the results, a back of the envelope calculation indicates that, if the number of options on each ballot were reduced by half than the total number of observed informal votes in the data would reduce by 14.9% and see the share of informal votes falling from 5.4% in total to 4.6%. This effect is important

because, from 2004-2016, around 32% of contests had more informal votes than the margin. Policies which affect the level of informal voting may, therefore, affect the final composition of Parliament

These findings are robust across model specifications and subgroups within the data. The findings suggest that the decision not to vote is more related to the costs of voting (understanding and deciding on a ranking of the candidates) rather than the potential benefits of an individual's vote being pivotal. This suggests that increasing the rate of formal voting could be achieved by strategies that make it easier for voters to research, understand and rank the candidates on the ballot.

One subgroup, areas with high levels of tertiary education, is of particular interest. Informal votes are less prevalent in these better-educated areas and voters are more strategic, reacting to the competitiveness of their electoral division as predicted by theory. Surprisingly, voters in these better educated areas still respond to the number of options in the same way as other voters, indicating that this behavior is not explained by differences in levels of education. This suggests that wasted votes are likely to be associated with a decision by the voter to abstain from voting rather than an error. This conjecture is supported by additional, non-causal analysis of the frequency of different types of informal votes. More options on a ballot are associated with higher levels of blank, non-sequential and incomplete informal votes and voters in areas with higher levels of education are less likely to unintentionally vote informally.

This paper is structured so that Section I gives a summary of the Australian electoral system, the main point in this background section is that boundary redistributions are made by an independent body and so are exogenous to the political process. In Section II, the research design, a Difference-In-Differences (DID) approach, is described in detail and integrated with existing literature on similar research designs. The data sources are set out in Section III and the main results of the analysis are presented in Section IV along with a range of robustness checks in Section V. Finally, Section VI concludes and also provides context for the results by considering how they relate to electoral outcomes.

I. Background on elections in Australia

As the research design makes use of many institutional factors relating to the Australian electoral system, we first overview the Australian electoral system.

A. The Australian Electoral System

Australia is a federation of six states and two territories, and there are separate elections for representatives at both the state and national level. National elections, the subject of our study, are elections of the national Parliament and occur roughly every three years. While each state has different rules for its own elections, rules for national elections are consistent across all states in Australia. In recent

elections, despite greater expenditure on political advertising and closer electoral outcomes, there has been an increasing trend in voters casting ballots that don't count towards the final result (referred to as informal votes and explained further below), informal voting has also been linked with lower socio-economic status, there has also been a growing number of candidates being presented on the ballot (Australian Electoral Commission, 2016*b*).

The Parliament in Australia is made up of two houses: The House of Representatives (henceforth 'House') and the Senate. The House has 150 members with each member being 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 electoral division has a population of around 100,000 voters. The Senate has 76 senators, each state in Australia has 12 senators to represent it while each territory has 2 senators to represent it. Each senator serves for a term of 6 years so that, at each election, half of the senate is contested. Thus, every electoral division has different candidates for the House while candidates for the Senate are the same for all electoral divisions within a state or territory.

Voting for the House and the Senate normally occurs at the same time and location. That is, voters arrive at a Polling Place in their local electoral division, 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 and then submits both at the same time.

On both ballots, the voter must list their preferences over the available candidates. Examples of correctly filled out ballots for the House and Senate are provided in Figure I.A. The candidates in each electoral division are officially announced around 3-4 weeks before the election day, for example, in 2016, this announcement was 22 days before the election and the order of candidates on the ballot is randomized at the electoral division level.

In the House, the winner in each electoral division is selected through a process known as full-preferential preference voting or Instant Run Off voting. In this process, the voter ranks all the candidates in order of preference. Ballots are initially counted for each voter's first preference, the candidate with the lowest share of first preferences is eliminated with votes redistributed according to the second preference on the ballot. This process continues until one candidate is the remaining choice of a majority of the voters. In the Senate, each state'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 has multiple Senators, a Senator is elected once they receive a certain proportion of the overall vote (14.3% for states and 33% for territories) rather than a majority.

In the Senate, until 2016, a voter could vote 'above the line', where they would number a single box associated with a party (and have their preferences dis-

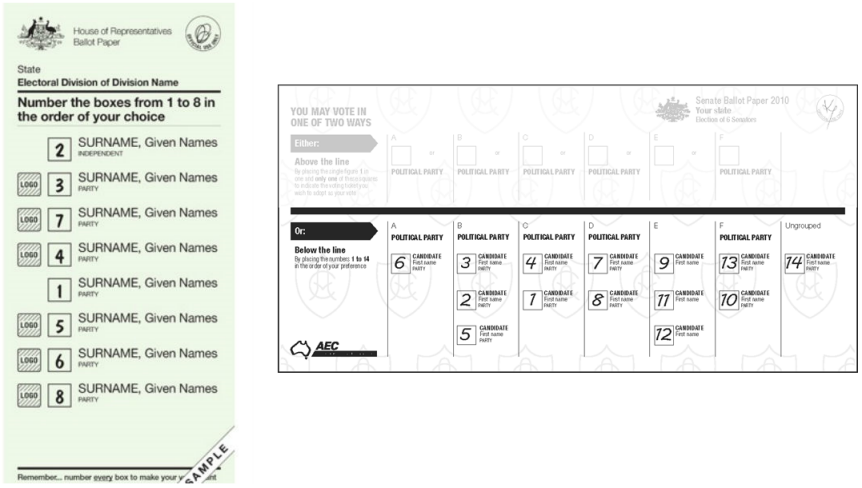

Figure 1. Correctly completed House (Left) and Senate (Right) ballots

Note: The Senate ballot is completed 'below the line' and according to the rules for voting in the Senate introduced in 2016, where below the line rankings do not need to be exhaustive.
Source: Australian Electoral Commission (2016*c*)

tributed according to the preferences of that party) or they could vote 'below the line', where they would need to rank all candidates on the ballot. In 2016, an adjustment was introduced to require numbering of at least 6 parties 'above the line' and allow ranking of at least 12 candidates below the line. Figure I.A shows a ballot for the Senate in 2016 that is correctly filled out 'below the line'.

Voting in Australia is compulsory. Every Australian citizen who has reached 18 years of age is required to be listed on the electoral roll and to vote in both state and national elections. Enrolment rates are high by international standards with over 96% of eligible Australians enrolled to vote (Australian Electoral Commission, n.d.). The accuracy of enrolment 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 government agencies. This is known as the Federal Direct Enrolment and Update (FDEU) program and uses information from state and territory driver's licences, the Department of Human Services and the Australian Tax Office Australian Electoral Commission (2019). Further, a voter cannot de-register themselves from the voting role with the exception of medical circumstances and permanent international relocation.[1]

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

[1]Prisoners are required to be registered and to vote.

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 situations such as dislike of candidates or lack of preference over candidates. If the non-voter does not pay the initial $20 fine or provide a valid and sufficient reason for not voting then the AEC may prosecute the non-voter in court and seek a penalty up to the current maximum of $210 plus legal costs. Then, 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.

In addition to voting in person at the local polling place on election day, referred to as an ordinary vote, there are also other options available. Voters can cast their vote at selected polling places within their electoral division prior to election day, called pre-poll voting. A voter can also cast an absent vote, where they vote outside their own electoral division but are provided with a ballot for their home electoral division. Voters can also apply to make a postal vote where the ballots are mailed to the voter's home, the voter completes the ballots in the presence of a witness, and returns them via mail. Overseas voters can vote at the Australian 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 only introduced in 2010 but has increased from around 7% of votes in 2010 to just under 20% in 2016 (Muller, n.d.). For federal elections, there is not any widespread availability of phone or internet based voting (Lundie, n.d.).

As a result of compulsory voting, the surrounding enforcement mechanisms and approaches such as postal votes and pre-polling, turnout is relatively high in Australia -– around 91% in 2016 – although there has been a decline from an average of around 95% in the 1980s and 1990s (Australian Electoral Commission, 2017). There is also a strong relationship between informal voting and turnout in Australia, shown in Figure I.A. Figure I.A uses data at the electoral division level for the years 2004-2016 where Non-Vote is the percentage of enrolled voters that do not turnout to vote. The relationship between informal voting and non-voting is positive and statistically significant at conventional levels, this remains true in unreported regressions that include fixed effects for year and electoral division, which suggests that there may be consistent factors which affect both decisions.

Despite voting being compulsory, a voter can submit an empty ballot or a defaced ballot rather than a legitimate vote. This type of vote is recorded in the voting data as an 'informal' vote. Examples of informal votes are provided in Figure 3. It should be noted that informal voting can also occur due to misunderstanding the voting rules or due to a data entry error when filling in the ballot (for example, marking two candidates as the first preference).

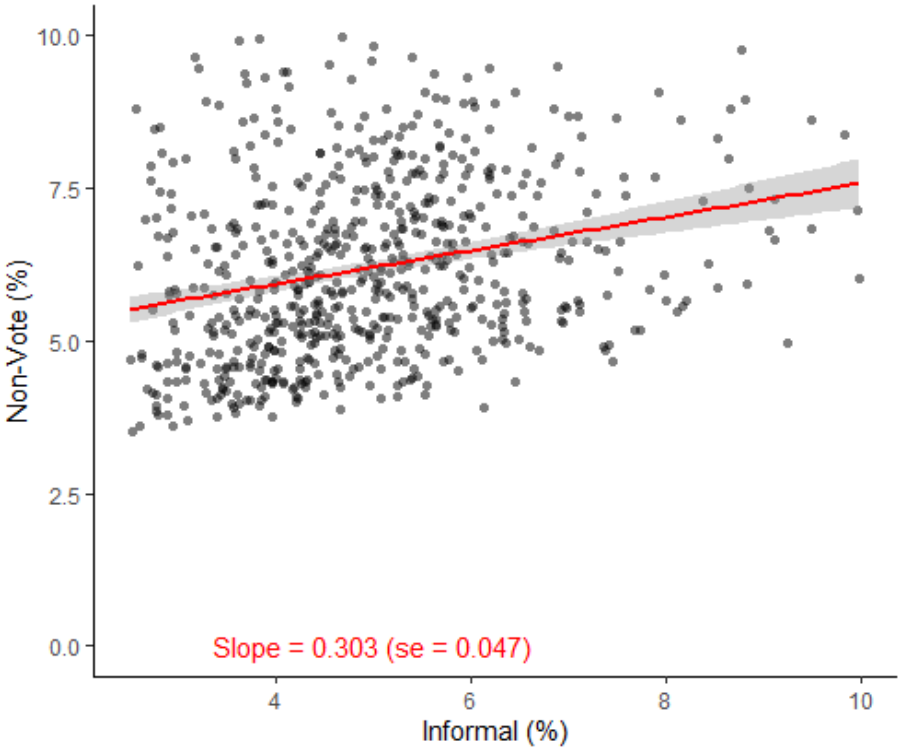

Figure 2. Relationship Between Informal Voting and Turnout

Note: Data is at the electorate level covering the years 2004-2016. Non-Vote is defined as 100 minus the turnout percentage recorded the Australian Electoral Commission. The coefficient of the slope remains statistically significant at the 5% level of significance when including fixed effects for year and electoral division.
Source: Authors' calculations based on multiple data sources from the Australian Electoral Commission

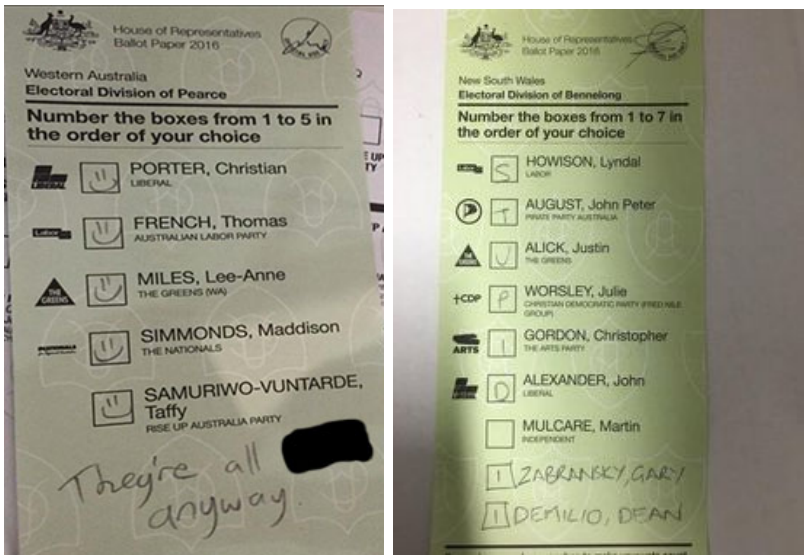

FIGURE 3. EXAMPLES OF INFORMAL VOTES

Source: Google Images

In recent elections, the rate of informal voting has been around 5-6% in the House and around 3-4% in the Senate. Informal votes can be broadly split between intentionally informal votes and unintentionally informal votes. While data on intentionality isn't systematically gathered or available at a Polling Place level, a recent review at the national level reveals that around 60% of informal votes in the House could be classed as unintentional while around 40% could be classed as intentional, with intentional informal votes trending upwards over time, this data is summarized in Table **??**.[2] Overall, the most common type of informal voting is where a voter only numbers one candidate on the ballot, accounting for an average of 28% of informal votes. The next most common type of informal vote, averaging 23.6% of informal votes, is a blank ballot, Table 2 shows data for 2007-2013 – the period over which detailed and comparable data is publicly available.

Australia's rate of informal voting is relatively high when compared to other OECD nations. For example, informal votes account for around 0.2% and 0.4% of all votes in the United Kingdom and United States 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. Among countries with compulsory voting, rates of informality tend to be higher and more comparable to rates in

[2]In this review, unintentional informal votes are defined as ballots with incomplete numbering, non-sequential numbering, ticks and crosses, and ballots in which the voter is identified; all other informal votes are classified as intentionally informal (Australian Electoral Commission, 2016*a*)

Table 1—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%

Note: In this review, unintentional informal votes are defined as ballots with incomplete numbering, non-sequential numbering, 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 for 2016 the AEC does not recommend comparison of 2016 figures with those of previous years (Australian Electoral Commission, 2016*b*).
Source: Authors' calculations based on data available in Australian Electoral Commission (2016*b*).

Table 2— Types of Informal Voting at the electoral division Level (2007-2013)

	Mean of number of votes	Mean of 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

Note: 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 data from the Australian Electoral Commission (Australian Electoral Commission (2009), Australian Electoral Commission (2011), Australian Electoral Commission (2016*a*)

Australia with examples being Singapore (2.1%), Argentina (4.13%), Belgium (5.8%), and Brazil (16%) (IDEA, n.d.).

National elections are administered by the Australian Electoral Commission (AEC). The AEC is a federal government agency that oversees organizing, conducting, and supervising federal elections and referendums. The AEC is also responsible for electoral division boundaries and redistributions, maintaining the electoral roll, publishing official records of election results, following up on voters who do not vote, monitoring the activities of registered political parties, and distributing public funding of political parties. The existence and role of the AEC is an important piece of the institutional framework for voting in Australia and is critical in the research design of this paper. The structure of the AEC means that many of the administrative aspects of voting are conducted independently of political considerations.

B. Changing electoral division boundaries

The AEC frequently undertakes adjustments of electoral division boundaries – referred to as redistributions. For example, in 2000–2017, there were 19 redistributions at the state level, and at least one redistribution occurred between every election. Redistributions are undertaken to ensure that each state and territory in Australia gains representation in the House in proportion to its population and so that each electoral division within a state or territory has similar numbers of voters.[3]

There is a clear and established process for conducting a redistribution. The first step is the establishment of a Redistribution Committee by the AEC for the state or territory where a redistribution has commenced. The committee then calculates the enrolment quota for each electoral division, which is essentially an estimate of the population divided by the number of seats in the House that that state is entitled to. The committee then divides the state into electoral divisions in order to ensure that the population in each electoral division is as close to the enrolment quota as possible. This division process is, legally, required to consider factors such as the economic, social and regional communities of interest, means of communication and travel, the physical features and area, and the existing boundaries. The public can comment on the proposed boundaries, but, once new boundaries have been set, the Parliament has no power to reject or amend the new boundaries. This process is conducted with minimal political input, implying that the redistribution is effectively exogenous to political processes and past political outcomes.

Following completion of each redistribution, the AEC must notify all voters who have been transferred between electoral divisions because of the creation of a new

[3]Less populous states and territories tend to experience greater variability in the average size of electorates than more populous states and territories.

electoral division, renaming of an existing electoral division, or because of a change in the boundaries of an existing electoral division. Notification is made through all national and state-wide newspapers as well as with letters to all affected voters.

Figure I.B shows examples of polling places changing electoral divisions. The main figure highlights the number of polling places nationally that changed electoral divisions between the 2010 and 2013 elections. The insets show a particular instance in western Melbourne where boundaries and polling places were exchanged between two neighboring electoral divisions between the 2010 and 2013 elections. In some cases, as seen in I.B a redistribution results in a Polling Place moving from one electoral division to another. In this case, the local voters who have changed electoral divisions will continue to vote at the same Polling Place but will now be voting in a different electoral division. This means that these voters experience an exogenous change in the number of candidates on the ballot, number of voters, and the margin.

The approach to redistributions in Australia can be contrasted with gerrymandering of electoral districts in parts of the US, which makes reallocation of a Polling Place an endogenous political decision. The issue of endogeneity of electoral division boundaries is given significant consideration Jones and Walsh (2018) with strong evidence of endogeneity presented in Carson and Crespin (2004). Stephanopoulos and McGhee (2015) propose a relevant statistic, named the efficiency gap, for measuring the degree of gerrymandering. The efficiency gap is a measure of the difference in the proportion of 'wasted votes' between two parties in an election; a wasted vote being defined as the number of votes for the winning party more than 50% plus all votes for the losing party. This statistic allows us to provide evidence for where the Australian system sits in the spectrum from fair to gerrymandered. If information is available on vote margin and the final number of seats won by a party then the approach in Stephanopoulos and McGhee (2015) can be simplified into the formula $e = s - (2 \times v)$ where e is the efficiency gap, s is the seat margin (which is defined as the share of seats won by a party minus 50%), and v is the vote margin (which is defined as the share of votes received by a party minus 50%). In gerrymandered divisions, the efficiency gap will be higher as the goal of gerrymandering is, basically, to waste votes for the opposition party. Applying the simplified efficiency gap formula to the results of recent Australian elections in the House provides the estimated efficiency gaps shown in Table 3.

These results indicate that, over this period, elections in the Australian House have seen an average efficiency gap of around 2.3% in favor of the Liberal National Coalition. To put this result into context, Jackman (2015) estimates that Wisconsin, a state where there is suspicion of gerrymandering, recorded efficiency gaps of between 10% and 13% in favor of the Republican party in recent years.

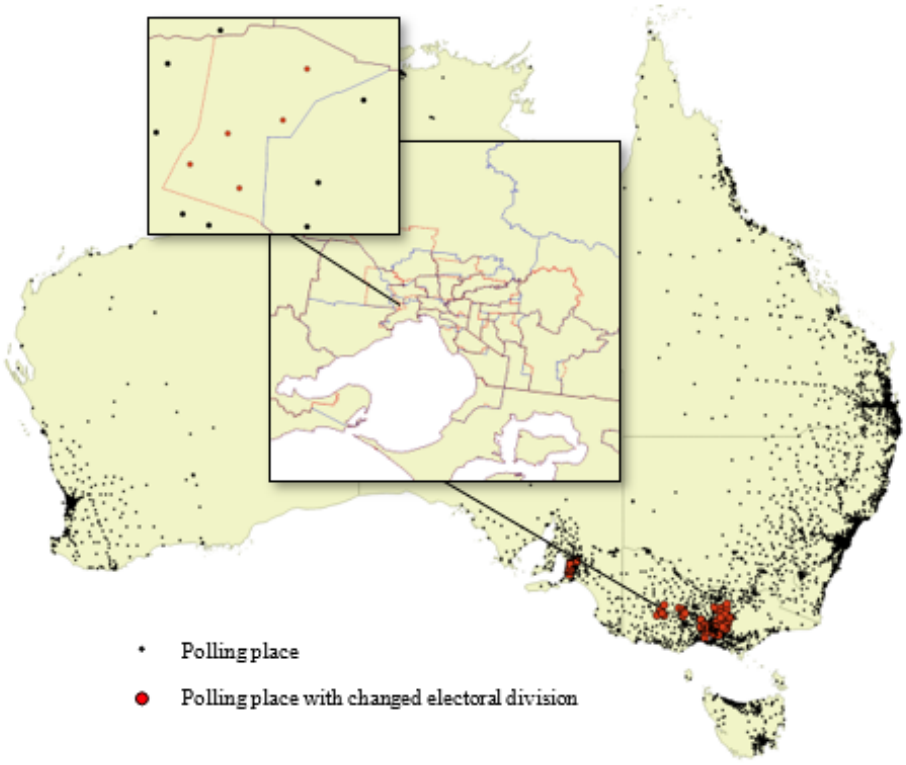

Figure 4. Polling Places in Australia with Insets of Boundary Changes

Note: The main map shows polling places with and without changed electoral divisions between the 2013 and 2016 elections. The insets show the electoral division boundaries in western Melbourne for 2010 in solid, blue lines and the boundaries for 2013 in dashed, red lines.
Source: Authors' calculations based on mapping data from Australian Electoral Commission (2018)

Table 3—Observed efficiency gap in recent Australian elections

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

Note: Efficiency Gap is defined as in Stephanopoulos and McGhee (2015), essentially a measure of the difference in the proportion of 'wasted votes' - those which don't help to elect a candidate.
Source: Authors' calculations based on data available on Wikipedia (2018).

II. Research Design and Econometric Specification

We test three hypotheses drawn from the literature: (H1) when the expected margin in an election is higher, informal voting will be higher; (H2) when there are more voters in an electoral division, informal voting will be higher; and (H3) when there are more candidates on the ballot, informal voting will be higher.

The basic relationships stated in these hypotheses are seen in the data. Figure II shows binned scatter-plots of informal voting percentages in the House at each polling place against the margin, the electoral division size, and the number of options on the ballot. In all three cases there is a positive, statistically significant relationship (slopes and standard errors are reported in brackets at the top of the chart).

The relationship seen for Margin is statistically significant, but its size does not appear to be practically meaningful with the estimated effect on informality being only 0.4 percentage points when moving from 0% margin to 25% margin. A similar weak relationship is reported by Kawai, Toyama and Watanabe (2020). This is an initial indication that support for H1 may be weak.

The margin is a standard indicator of the closeness of an election and is defined as the winning party's vote share minus 50%. In the Australian context, where instant runoff voting is used, the margin refers to the final round of voting that results in the selection of a winning candidate. We also considered an approach to defining the margin based on first preferences, rather than final shares, this alternative definition does not materially alter the results in this paper and also lack theoretical justification as margin based on first preferences does not directly determine the probability of being the pivotal voter in an election. As a result, the remainder of this paper focusses on the standard definition of the margin.

The positive, statistically significant relationships in Figure II, however, do not imply causal relationships. Consider the following simple model:

$$(1) \qquad Informal_{i,t} = \theta_1 Options_{i,t} + \theta_2 Voters_{i,t} + \theta_3 Margin_{i,t} + \mathbf{X}_{i,t}\beta + \epsilon_{i,t}$$

In this model specification, i refers to a specific Polling Place while t refers to the time period. $Informal_{i,t}$ is the percentage of informal votes recorded at Polling Place i in time period t, $Margin_{i,t}$ is the margin for Polling Place i in time period t, $Voters_{i,t}$ is the number of voters in the electoral division for Polling Place i in time period t, $Options_{i,t}$ is the number of options on the ballot for Polling Place i in time period t and $\mathbf{X}$ is a vector of covariates such as demographic characteristics. That is, this specification simply merges together the relationships seen in the three scatter plots above.

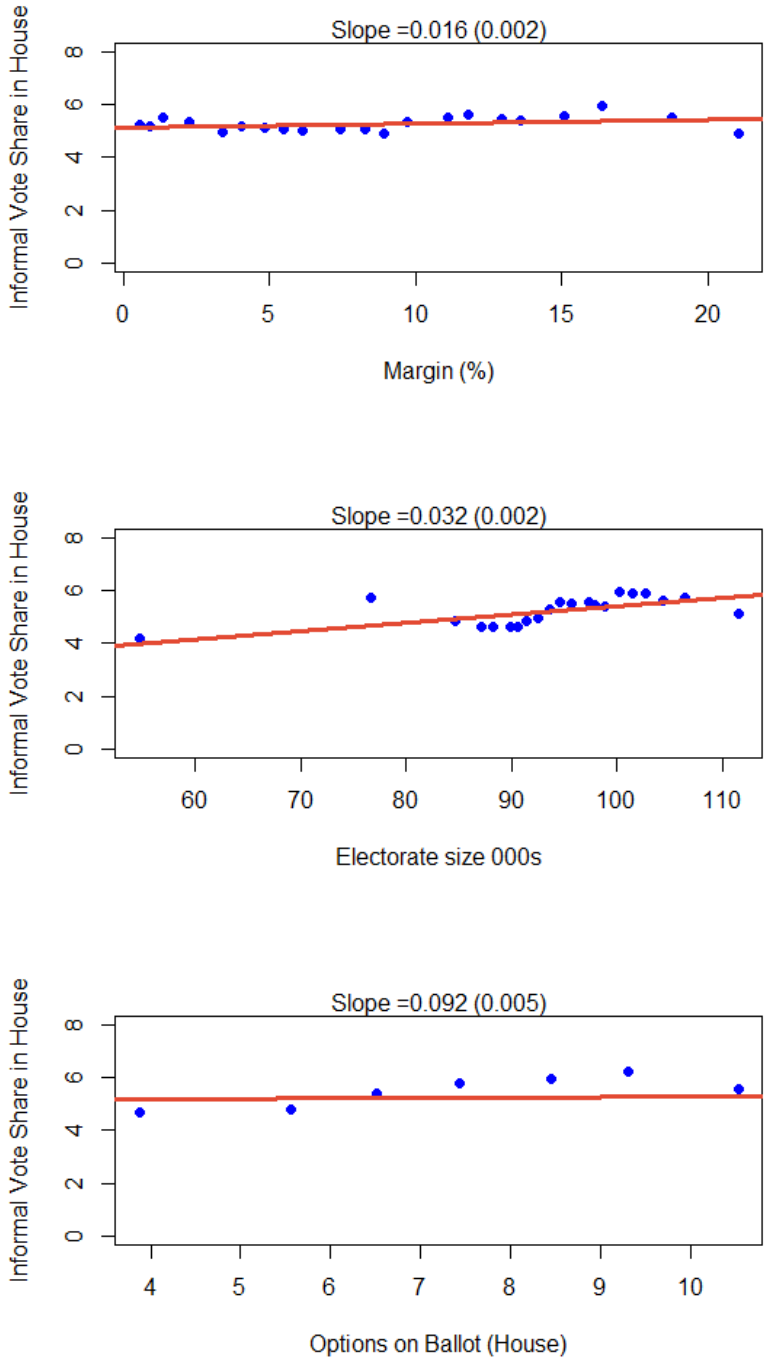

FIGURE 5. POSITIVE RELATIONSHIP BETWEEN MARGIN, ELECTORAL DIVISION SIZE AND OPTIONS ON BALLOT WITH INFORMAL VOTING

Note: Dots represent binned averages where each dot has the same number of observations. The fitted line is derived from unweighted OLS regression of the form $Informal\% = \beta_0 + \beta_1 Margin + \epsilon$ with the slope and standard error listed at the top of the figure.
Source: Authors' calculations based on data from Australian Electoral Commission (n.d.)

The parameter estimates for θ_1, θ_2, and θ_3 from running this regression model will not necessarily yield causal effects due to endogeneity problems present in the relationship. There is likely to be omitted variable bias, such as knowledge of and interest in politics, which systematically vary across electoral divisions. Also, electoral divisions with small numbers of voters are likely to be in rural areas, leading to systematic differences in informal voting due to unobservable characteristics that aren't captured in $\mathbf{X}$. The voting behavior of an electoral division can also affect which candidates are available in that electoral division and so can influence the margin.

This type of endogeneity issue is widely acknowledged and has been identified in several papers. For example, Carson and Crespin (2004), Cox and Katz (2002) and Jones and Walsh (2018) all discuss similar identification issues. One strategy that has been used to address this endogeneity issue is Difference-in-Differences (DID) using changes in electoral division boundaries. Several papers have used changes in US electoral division boundaries as a source of exogenous variation. Examples here include Ansolabehere, Snyder Jr and Stewart III (2000), Desposato and Petrocik (2003), Fraga (2016), and Jones and Walsh (2018). A potential concern with applications of this approach to US data is raised by Carson and Crespin (2004) who provide evidence that the method of deciding electoral division boundaries in the US affects competitiveness of elections even when courts and administrative organizations set the boundaries. This result is further reinforced by Henderson, Sekhon and Titiunik (2016) who show that electoral division boundary changes appear to be endogenous political decisions in a number of US jurisdictions.

Our causal analysis assumes the exogeneity of boundary changes. We argue this 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 (see Section I.B for details), the results show that the assumption is plausible in Australia. As a result, in Australia, when a redistribution results in a polling place moving from one electoral division to another, voters experience an exogenous change in the number of candidates on the ballot, number of voters, and the margin. This approach requires an assumption that there is no systematic bias for voters who change which polling place they vote at relative to voters who consistently vote at the same polling place.

In this research design, treated polling places are those that changed electoral division while the control group are polling places that did not change electoral division.[4] A potential concern is that these control polling places do experience

[4]As a practical example, the Polling Place located at the Panania Senior Citizens Centre in south Sydney has moved between the electoral divisions of Banks and Hughes several times. In the 2007 election, the Polling Place was in Banks but by 2010 the Polling Place was in Hughes (due to a redistribution in

some form of treatment as the boundaries of their electoral division do change. This concern is partially addressed by the fact that the level of change experienced by these control polling places is very small relative to that experienced by the treated polling places. Further, additional analysis will be undertaken to test whether the results are sensitive to inclusion or exclusions of these control polling places.

The structure of the Australian electoral system also provides another potential set of controls for the treated polling places that change electoral divisions. Australian elections require voters to submit ballots for both the House of Representatives and the Senate at the same time. In the House of Representatives voters are voting as part of their local electoral division. This means that the number of options, electoral division size, and expected margin varies between each electoral division. In contrast, in the Senate, voters are voting as part of their state. This means that, within a state, every electoral division has the same number of candidates, number of voters and expected margin. Redistributions don't create changes in the Senate, voting in the Senate can therefore form another control with the treatment group being voting in the House of Representatives.

Comparing the strengths and weaknesses of these two potential control groups, voting behavior in the House in similar Polling Places is likely to provide a better control group then voting behavior in the Senate. This is because there are significant differences between voting and electoral procedures in the House and the Senate. For this analysis, the most important difference is that, in the House, for a vote to be counted as formal, voters must completely rank all candidates available on the ballot, while, in the Senate, a voter can vote 'above the line' and not rank all options available. This can make voting in the Senate less costly in terms of time, research and decision-making effort than when voting in the House. As a result, informal voting tends to be systematically higher in the House than the Senate. For this reason, while results for both approaches are reported, nearby or similar Polling Places are our preferred control group while voting in the Senate acts as an additional robustness check.

For the remainder of this paper, the approach that uses voting in the House and compares polling places that changed electoral divisions with those that didn't is referred to as DID version 1 (DID v1), this is our preferred specification. The approach where voting in the House and Senate are compared for polling places that changed electoral division is referred to as DID version 2 (DID v2). While results for both approaches are reported, voting in the Senate acts as a robustness check.

2009). Then, following another redistribution in 2016, the Polling Place was back in Banks for the 2016 election. The nearby Polling Place at the Lugarno Public School remained in the electoral division of Banks throughout all these elections and so could form part of the control group for the Polling Place located at the Panania Senior Citizens Centre. Comparisons between these Polling Places informs the research design.

Using a DID approach is also beneficial as it controls for unobserved time-invariant factors associated with each Polling Place (such as unobserved socioeconomic characteristics) as well as time trends common across Polling Places (such as changes in the national political situation).

Because the treatments (change in number of options, change in electoral division size, and change in margin) are continuous variables, i.e. some polling places will move to an electoral division with a higher margin while other polling places will move to an electoral division with a lower margin. On average, the change in margin for the treatment group could be roughly zero. To account for this, Angrist and Pischke (2008) suggest a model specification that is based on Card (1992). Following this approach, for DID v1, we estimate variations on the following models:

$$\begin{aligned} Informal_{i,t} =& \delta_1 Change.in.Margin_{i,t} + \delta_2 Change.in.Voters_{i,t} \\ &+ \delta_3 Change.in.Options_{i,t} + \delta_4 Changed.Division_{i,t} \\ &+ \gamma_t + \alpha_i + X_{i,t}\beta + \epsilon_{i,t} \end{aligned} \tag{2}$$

Here i refers to a specific Polling Place and t refers to the time period. In DID v1, the sample is restricted to the House. $Informal_{i,t}$ is the percentage of informal votes in the House (potentially ranging from 0 to 100) recorded at Polling Place i in time period t. $Change.in.Margin_{i,t}$ is the change in margin caused by moving between electoral divisions for Polling Place i in the House, in time period t. For this variable, the margin is coded so that it potentially ranges from 0 to 1. $Change.in.Voters_{i,t}$ is the change in the natural log of the number of voters in the electoral division caused by moving between electoral divisions for Polling Place i in the House. $Change.in.Options_i, t$ is the change in the natural log of the number of options on the ballot caused by moving between electoral divisions for Polling Place i in time period t. $Changed.Division_{i,t}$ is a dummy variable indicating whether Polling Place i changed electoral Division since the previous election, in time period t. The inclusion of this variable reflects findings from Hayes and McKee (2009) that, after redistricting, voters are 3-7% more likely to suffer roll-off, where voters don't fully complete their ballot, in redistricted areas when compared to voters that were not affected by redistricting.

These particular scales and transformations have been selected in order to give the parameters a comparable and meaningful interpretation, which will be discussed in the results. For fixed effects, γ_t is a time fixed effect while α_i is a Polling Place fixed effect. $\mathbf{X}$ is a vector of covariates such as demographic characteristics of the electoral division. It is assumed that the error term $\epsilon_{i,t}$ $N(0, \sigma_i^2)$. We assume that errors can be correlated within a Polling Place which is addressed through the use of clustered standard errors (Bertrand, Duflo and Mullainathan, 2004).

In addition to the specification described above, we also estimate other specifications of DID v1 as robustness checks. These include additional variables showing the exogenous change in the tenure of the incumbent candidate and the exogenous change in the share of votes for progressive parties.

The approach for DID v2 is:

$$\begin{aligned} Informal_{i,j,t} &= \delta_1 Change.in.Margin_{i,t} + \delta_2 Change.in.Voters_{i,t} \\ &+\delta_3 Change.in.Options_{i,t} + \varphi_{i,t} + X_{i,t}\beta + \eta House_{i,j,t} + \epsilon_{i,t} \end{aligned} \tag{3}$$

For DID v2, the sample is restricted to polling places that moved between electoral divisions. The two main difference is that DID v2 includes both houses of Parliament (so that j can take a value of either House or Senate). The difference between the two Houses is controlled for using a dummy variable $House_{i,j,t}$, which takes a value of 1 if the observation is for the House and 0 otherwise. Further, this specification excludes the $Changed.Division$ variable as all observations in this specification have changed division. The other major change is that the fixed effects, $\varphi_{i,t}$, have been implemented as an interaction of polling place and time. This is because, in 2013, all the treated polling places only received treatment in that year -– which creates perfect multicollinearity between time period and polling place fixed effects for 2013.

The presence of the two alternative control groups and the two DID formulations allows for the application of a triple differences approach (DDD):

$$\begin{aligned} Informal_{i,j,t} &= \delta_1 Change.in.Margin_{i,j,t} + \delta_2 Change.in.Voters_{i,j,t} \\ &+\delta_3 Change.in.Options_{i,j,t} + \delta_4 Changed.Division_{i,j,t} \\ &+\gamma_t + \alpha_i + X_{i,t}\beta + \eta House_{i,j,t} + \epsilon_{i,j,t} \end{aligned} \tag{4}$$

The model is then estimated using both Houses of Parliament and Polling Places that both moved and did not move between electoral divisions.

Within all three of these models, δ_1, δ_2, and δ_3 measure the treatment effects of interest.

Considering the $Change.in.Margin_{i,t}$, $Change.in.Voters_{i,t,}$ and $Change.in.Options_{i,t}$ in more detail, these variables could affect voting behavior either immediately or with a lag. It has become common in the relevant literature to test the relationship between margin and a 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 as voters are forward looking and react rationally to information such as polling data (see Geys (2006) for further analysis of this approach). In general, in applications without a

clear source of exogenous variation, the ex-post approach poses some endogeneity concerns, but these concerns are avoided due to the proposed research design.

III. Data

Data on voting and elections was sourced from several administrative data sets compiled by the Australian Government and its agencies: the AEC for voting data (Australian Electoral Commission, n.d.); the Australian Bureau of Statistics for social, demographic, and economic data (ABS, 2016*b*); and the Australian Broadcasting Corporation for data on election outcomes (Australian Broadcasting Corporation, n.d.).

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. For these elections, the AEC makes Polling Place level data available on its website. Earlier elections only have electoral division level data available and so are not useful in this research design. The observation level in this data is the (Polling Place, House of Parliament) pair with a count of first preference votes given for each candidate including a count of informal votes.

There are currently around 8300 Polling Places in Australia in total, but this varies between elections. A number of restrictions are placed on the Polling Places that are included in the analysis. First, we only include Polling Places that are present in each election from 2004 to 2016. This reduces the number of Polling Places included to 6360 across 150 electoral divisions, the data remaining is referred to as 'All Data' in the following summary tables.

Next, electoral divisions which were created, nullified, or renamed are also excluded.[5] This ensures that a change in the electoral division name of a Polling Place was genuinely associated with a change in electoral division boundaries and not a result of other administrative changes. This results in a final set of around 6,000 Polling Places in each election across 143 electoral divisions.

Finally, data for 2004 is excluded as the available data does not allow us to determine whether a Polling Place changed electoral divisions for the 2004 election. Data for 2004 is used for determining whether a Polling Place changed electoral divisions for the 2007 election.

The data remaining after these exclusions is referred to as 'DID Data' in the following summary tables. A summary of these Polling Places is set out in the table below. In general, around 2-8% of Polling Places are identified as having

[5]Electoral divisions are not often created, destroyed or renamed (of the 65 electoral divisions included in the 1901 election, 34 are still in existence). For example, the electoral division of Charlton, located north of Sydney in the Hunter region of NSW, was eliminated prior to the 2013 election. The electoral division of Burt, located in the South-Western suburbs of Perth in Western Australia, was created for the 2013 election — reflecting relative population growth of Western Australia to the rest of Australia. Three electoral divisions were renamed over the period from 2001.

changed electoral division in any given election. For 2013, there are fewer Polling Places that changed electoral division because, between 2010 and 2013, only relatively minor redistributions occurred, and these were limited to Victoria and South Australia.

Table 4—Number of Polling Places and Electoral Divisions in the Sample

	2004	2007	2010	2013	2016
Panel A - All Data					
Polling Places	6,360	6,360	6,360	6,360	6,360
Changed electoral division	NA	623	661	156	591
Did not change	NA	5,737	5,699	6,204	5,769
Electoral divisions	150	150	150	150	150
Panel B - DID Data					
Polling Places		6,060	6,014	6,014	6,031
Changed electoral division		535	482	156	484
Did not change		5,480	5,532	5,858	5,547
Electoral divisions		143	143	143	143

Note: "All Data" includes all Polling Places present in each election. "DID data" excludes observations for electoral divisions that were created, nullified, or renamed.
Source: Authors' calculations based on data from Australian Electoral Commission (n.d.)

This data indicates that the level of informal voting varies between polling places, year and House of Parliament but, overall, in the House of Representatives, the informal vote share has a mean of 5.3%. In the Senate, the informal vote share has a mean of 3.6%. The data also shows that, although the AEC attempts to keep the population in each electoral division roughly equal (at around 100,000), there is still variation. The smallest electoral division included 54,725 registered voters while the largest included 143,231. The average electoral division contains around 95,195 voters. The number of candidates on the ballot in each polling place for both the House and Senate is also covered. The minimum number of candidates seen in the House is 3, the maximum is 19 with a mean of 7.6. The number of options on Senate ballots ranges from 3 up to 65 with a mean of 29.2.[6]

Data was gathered from the Australian Broadcasting Corporation (ABC) on the incumbent party in each electoral division, the winning party, and the margin for the winning party in each election. The average margin across electoral divisions was 9.3%, while the maximum margin was 26%.[7]

[6]There are generally more options on a Senate ballot as each state or territory elects multiple senators by proportional representation, making the Senate a more appealing target for minor party candidates.

[7]Frequently, betting market data is 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 electoral division level in Australia with the only historical data available being for which party will form government.

We also make additional calculations based on this data. We identify polling places that move between electoral divisions using name changes in successive elections. We transform the number of options and number of voters by taking log-differences. We calculate differences in levels for the margin. For use in additional robustness checks we also calculate the tenure (in years) of the incumbent candidate, also taking log-differences, and the share of votes for progressive parties (defined as first preference votes for the ALP and Green parties).

There are around 24,000 observations at the (polling place, year) level in the DID data, this creates around 48,000 observations at the (polling place, House of Parliament, year) level as each polling place generates an observation for the House and the Senate in each year. Around 1,657 observations are for polling places that changed electoral divisions. A summary of this data is provided in Table 5, below. In this table, the 'Change in' variables are always zero for polling places in the control group and also for the Senate, as it is only treated polling places and the House that are affected by the change in electoral division boundaries. The 'Change in' variables are also close to zero in the All Data and DID Data columns as the treated polling places only make up a small portion of the overall data.

Table 5—Summary Statistics For Key Variables

	All Data		DID Data		Treated		Control	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
House informal share	5.30	2.61	5.29	2.67	5.95	2.97	5.25	2.64
Senate informal share	3.58	1.86	3.52	1.84	3.91	2.23	3.49	1.81
Margin	0.093	0.061	0.091	0.061	0.089	0.062	0.091	0.061
Voters	95,195	9,939	97,081	9,545	98,107	8,983	97,005	9,581
Number of House Options	7.63	2.76	7.55	2.77	6.79	1.96	7.61	2.81
Number of Senate Options	29.18	10.38	31.17	10.38	30.60	8.15	31.21	10.53
Tenure	9.94	5.88	10.26	6.12	9.54	5.91	10.32	6.13
Progressive vote share	0.51	0.16	0.52	0.16	0.53	0.17	0.52	0.16
ln(Voters)	11.46	0.11	11.48	0.10	11.49	0.09	11.48	0.10
ln(N House Options)	1.97	0.34	1.96	0.34	1.87	0.29	1.97	0.35
ln(N Senate Options)	3.30	0.41	3.38	0.38	3.38	0.28	3.38	0.39
ln(Tenure)	2.06	1.02	2.12	0.69	2.04	0.68	2.12	0.69
Change in Margin	0.000	0.017	0.000	0.018	0.004	0.071	0.000	0.000
Change in ln(Voters)	0.000	0.010	0.000	0.011	0.002	0.045	0.000	0.000
Change in ln(N House Options)	−0.001	0.073	−0.001	0.078	−0.016	0.314	0.000	0.000
Change in ln(N Senate Options)	0	0	0	0	0	0	0	0
Change in ln(Tenure)	−0.036	0.767	−0.000	0.237	−0.006	0.966	0.000	0.000
Change in Progressive vote share	0.005	0.029	0.006	0.031	0.095	0.086	0.000	0.000
Observations		31,800		24,119		1,657		22,462

Note: All Data includes all Polling Places present in each election. DID data removes some observations for electoral divisions that entered, exited or changed names.
Source: Authors' calculations based on data from Australian Electoral Commission (n.d.) and Australian Broadcasting Corporation (n.d.)

This voting data is then merged with social, demographic and economic information. This supporting data is not available at the Polling Place level. As a

result, a GIS program was used to match the latitude and longitude coordinates for Polling Places to their respective Statistical Area Level 2 (SA2) as defined by the Australian Bureau of Statistics (ABS). There are around 2,200 SA2s in Australia with populations in the range of 3,000 - 25,000 and an average population of around 10,000. Each SA2 aims to represent a community that interacts together socially and economically (ABS, 2016*a*). A single SA2 is likely to contain multiple Polling Places and so this approach cannot provide social, demographic, and economic information at the level of the Polling Place. The SA2 was selected as the most appropriate area for this matching as it presents a reasonable tradeoff between a smaller geographic area that is better matched to a Polling Place and a larger geographic area that has better data available. After mapping Polling Places into SA2s, a range of social, demographic, and economic data from ABS's National Regional Profile data (ABS, 2016*b*) is matched to each Polling Place. The National Regional Profiles contain information such as population, median age, median income, population density, percentage of people with a tertiary degree, the unemployment rate, and housing costs. A summary of these demographic and economic characteristics is provided in columns 1-3 of Table 6. When used in regressions, data on Population Growth is split into two variables '– one for areas that experienced population growth and one for areas that experiences population experienced population decline (the absolute value is used in the regression).

Although the change in electoral division boundaries is administered in an apolitical way, there is still the possibility that the treatment and control polling places differ in meaningful ways. For example, in a given year, electoral division boundaries may change in Victoria and be unchanged in the Northern Territory. polling places in the Northern Territory may differ systematically in both observable and unobservable characteristics from those in Victoria and so may not present the ideal controls. Two different approaches to address this are used: propensity score matching and distance-based restrictions. When neither of these approaches are used, the output tables are called "Standard" outputs.

For the propensity score matching approach, the propensity score is estimated using a range of observable characteristics of the polling place that should not be affected by treatment:

$$\begin{aligned} Changed.Division_{i,t} &= \beta_0 + \beta_1 Median.Age_{i,t-1} + \beta_2 Average.Income_{i,t-1} \\ &+ \beta_3 Unemployment.Rate_{i,t-1} + \beta_4 Population.Density_{i,t-1} \\ &+ \beta_5 Population.Growth.Rate_{i,t-1} + \beta_6 New.House.Value_{i,t-1} \\ &+ \beta_7 ESL_{i,t-1} + \beta_8 Tertiary.Degree_{i,t-1} + \epsilon_{i,t} \end{aligned} \tag{5}$$

Where $Changed.Division_{i,t}$ takes the value of 1 if Polling Place i changed electoral division between time period $t-1$ and time period t. Other variables should be self-explanatory with the possible exception of $ESL_{i,t-1}$, which is the percent-

age of households where English is a second language. The propensity score is estimated using a logit distance function. Matching is done based on the nearest neighbor technique without replacement (Ho et al., 2007). The pre- and post-matching means for both the treated and control groups are presented in Table 6. In most cases the treated and control groups are similar in observable characteristics pre-matching. Post-matching, the absolute value of the standardized. mean difference is less than 0.1 which satisfies the rule of thumb provided in Flury and Riedwyl (1986) for matching to have successfully produced treated and control groups that are similar in observable characteristics.

Table 6—Pre and Post matching means for treated and control groups

	Pre Matching			Post Matching		
	All	Treated	Control	Treated	Control	Std. Mean Diff
Median Age	39.63	39.31	39.64	38.45	38.5	-0.01
	(5.17)	(5.42)	(5.15)	(5.42)	(4.58)	
Mean Income (000)	49.86	48.6	49.93	44.26	44.53	-0.02
	(15.58)	(13.68)	(15.68)	(13.68)	(14.15)	
Unemployment (%)	5.58	5.81	5.57	5.71	5.6	0.05
	(2.1)	(2.2)	(2.09)	(2.2)	(2.22)	
Population Density	1,171.14	1,097.12	1,175.4	1,015.68	990.6	0.02
	(1,589.28)	(1,706.6)	(1,582.17)	(1,706.6)	(1,274.88)	
Population Growth (%)	0.87	0.5	0.89	1.04	1.02	0.01
	(3.36)	(4.29)	(3.3)	(4.29)	(1.73)	
House Value (000)	452.4	400.62	455.39	332.18	337.75	-0.02
	(578.36)	(414.53)	(586.27)	(414.53)	(327.14)	
English 2nd Language (%)	14.14	16.64	14	16.04	14.66	0.07
	(16.14)	(20.93)	(15.8)	(20.93)	(17.13)	
Tertiary Degree (%)	17.23	14.96	17.36	14.42	14.65	-0.03
	(10.78)	(9.49)	(10.83)	(9.49)	(9.19)	

Note: Matching is done based on the nearest neighbor technique without replacement Ho et al. (2007). Standard errors shown 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).

For distance-based restrictions, the sample is restricted to Polling Places within 2.5 kilometers of treated Polling Places. To the extent that there are characteristics of treated Polling Places that are correlated with their physical location, this geographic restriction should help ensure that the treated and control Polling Places are as similar as possible to each other in terms of observable characteristics. An advantage of the distance-based restrictions over propensity score matching is that distance restriction may provide additional control for time varying unobservable characteristics that are associated with geographic location (such as regional political sentiment) while the propensity score matching approach can only adjust for observable characteristics. The distance-based restriction does results in the smallest sample size though.

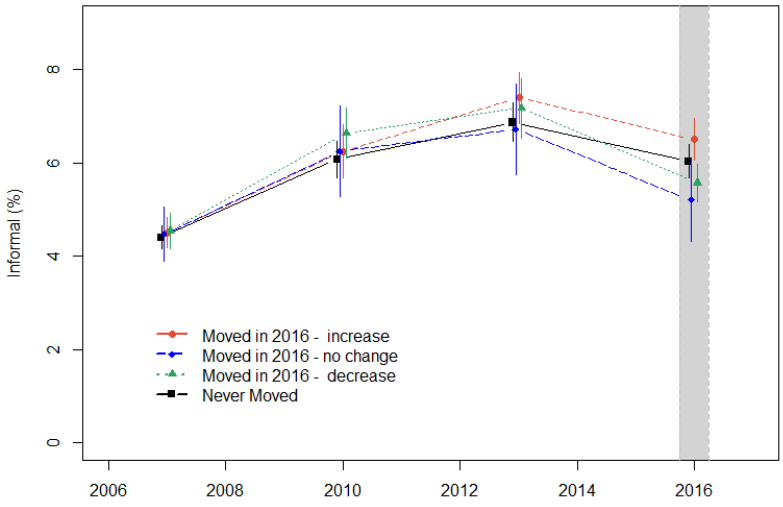

FIGURE 6. TIME TRENDS FOR GROUPS RECEIVING DIFFERENT CHANGES IN THE NUMBER OF OPTIONS ON THE BALLOT

Note: Polling places in the 'increase' group experienced an increase in the number of options on the ballot in 2016 – similarly for 'no change' and 'decrease' group. The 'Never Moved' group did not move electoral divisions 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 data from Australian Electoral Commission (n.d.)

Undertaking matching also allows for a visual test of whether the common trends assumption is met in the data. Figure III shows the rate of informal voting in the House over time and split into four groups: polling places that never moved; polling places that only moved in 2016 and experienced an increase in the number of options on the ballot; polling places that only moved in 2016 and experienced decrease in the number of options on the ballot; and polling places that only moved in 2016 but experienced no change in the number of options on the ballot. That is, Figure III isolates the effect of treatment in 2016. 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 above but applied at the polling place level and only to those in the 'Never Moved' group. For this figure, polling places were matched based on observable characteristics in 2007, 2010 and 2013.

Visually, it appears that each group broadly follows the same time trend with observations in 2007, 2010 and 2013 showing substantial overlap in confidence intervals between all four groups. In 2016, the group which recorded an increase

in the number of options appears to experience higher levels of informal voting while those in the no change and decrease groups record relatively lower levels of informal voting. This visual analysis supports the common trends assumption and also indicates some of the main findings in the following section.

IV. Results

A. Main results

There are several sets of results presented below with a range of robustness and specification tests also shown. Table 7 presents the main results, starting with some basic models for comparison and then focuses on DID v1. Column 1 shows the results of a simple linear regression comparable to the relationships that are shown in Figure II. In this simple analysis, the relationships between the number of options, voters and margin are all positive and statistically significant. Column 2 introduces a range of socioeconomic covariates to the simple regression, this results in a change of sign for the margin covariate, although it is no longer statistically significant at conventional levels. The same general findings carry over into Column 3, where fixed effects for polling place are also included.

Columns 4-6 show the results of DID v1, Column 4 uses exogenous changes in the margin, voters and number of options.[8] Column 5 shows the same model after propensity score matching, while Column 6 shows the distance limited results. The results in column 5 and 6 should be interpreted as robustness checks – and produce similar values for the statistically significant variable Change in ln(Options). Columns 5 and 6 also produce results for Change in ln(Voters) that are of the expected sign, although not statistically significant at conventional levels.

To confirm the three main hypotheses, the coefficients for Change in Margin, Change in ln(Voters) and Change in ln(Options) shown in columns 4, 5, and 6 should all be positive and statistically significant. In all cases, the parameter estimate for Change in ln(Options) is statistically significant and positive, the sign implied by the hypotheses. Across all three models, the parameter estimates for Change in Margin and Change in ln(Voters) are not statistically significant and vary in sign. This is a rejection of H1 and H2, and a confirmation of H3. These results are in line with the literature that focuses on the mental processing costs of voting as well as with Feddersen and Pesendorfer (1996), whose model suggests no causal relationship between pivot probabilities and abstention.

The parameter estimate of 1.319 for Change in ln(Options) indicates that, if there was a doubling of the number of options on the ballot, then the level of informal voting will increase by 1.3 percentage points. With an average level of informal

[8]The data set is voting in the House of Representatives with treated polling places being those that moved between electoral divisions and control polling places being those that did not move

TABLE 7—MAIN RESULTS

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. covariates	Fixed Effects	DID Standard	DID Propensity Score Matched	DID Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	2.136***	−0.130	−3.173***			
	(0.347)	(0.281)	(0.358)			
ln(Voters)	3.308***	1.418***	0.862**			
	(0.191)	(0.156)	(0.351)			
ln(N Options)	0.523***	0.864***	2.080***			
	(0.066)	(0.053)	(0.057)			
Change in Margin				0.448	0.029	−0.017
				(0.881)	(1.097)	(0.878)
Change in ln(Voters)				−0.186	−0.156	0.479
				(1.139)	(1.532)	(1.151)
Change in ln(N Options)				1.319***	1.414***	1.316***
				(0.213)	(0.292)	(0.209)
Changed Division				−0.249***	−0.384***	−0.145**
				(0.065)	(0.098)	(0.065)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Clusters (Polling Place)	6,184	6,184	6,184	6,184	5,463	2,056
Observations (Polling Place × Year)	24,119	23,476	23,476	23,276	12,969	7,457
Treated Observations	NA	NA	NA	1,435	1,177	1,435
Control Observations	NA	NA	NA	21,841	11,792	6,022
Adjusted R^2	0.027	0.430	0.701	0.663	0.679	0.770

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the polling place level

Note: 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; Column 3 introduces fixed effects for Polling Place. Columns 4-6 shows the results of DID v1. Column 4 presents the main output, which uses voting in the House with treated Polling Places being those that moved between electoral divisions and control Polling Places being those that did not move. Column 5 shows the same model but after propensity score matching to make the observable characteristics of treated and control groups similar has been applied; 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

voting of around 5.3%, this would be equivalent to a 25% increase in informality. As the number of options tends to be between 4 and 8, changes of 25% to 100% are realistic and this suggests the results are significance in practical terms.

Applying this estimate linearly indicates that, if the number of options were reduced by half in each electorate than the total number of observed informal votes in the data would reduce by 14.9% and the share of informal votes would fall from 5.4% in total to 4.6%.

The parameter estimate for Changed Division is negative and statistically significant across all specifications. This indicates that voters who experience a change in their electoral division are less likely to submit an informal vote than those that don't change division. This may be due to voters taking more interest in the 'new' candidates or issues in their new electorate and so being more engaged in the political process.

For all models, the results are presented using clustered standard errors with the cluster being defined at the polling place level (Bertrand, Duflo and Mullainathan, 2004). Clustering at the polling place level allows for the errors to be correlated within a polling place. This seems reasonable as voting behavior within a polling place is unlikely to be affected by or influence other polling places but there is likely to be some form of correlation over time within the same polling place.

We also analyzed subgroups within the data. We focused on subgroups defined by the level of margin, the size of electoral division, income levels, tertiary education levels and English as a second language levels. Investigation of these subgroups generally produced results similar to those above. We were generally not able to identify specifications or subgroups that support H1, H2, and H3 at the same time. Table 8 shows these results when the data is subset according to quartiles of tertiary education. For the highest education group (Quartile 4), informal votes are less prevalent, the sign for Change in ln(Options) and Change in Margin are in line with the implications of the hypotheses and are statistically significant.

These results indicate that, in better-educated areas voters react to the competitiveness of their electoral division as predicted by theory. Better educated voters may consider the strategic implications of their voting decisions more closely than other groups. A potential explanation is that those with high education levels may be more engaged in political issues (which may reduce their costs of acquiring information on candidates); may see greater personal benefits from having their preferred candidate elected; and may understand how the electoral system works in greater detail. These voters still respond to the number of options in the same way as other voters -– indicating that this result is not explained by differences in education.

The results in Table 8 provide mixed support for the implications of Feddersen and Pessendorfer's model of vote abstention, where uninformed voters abstain so that

Table 8—Subgroups analysis based on quartiles of Tertiary education levels

	Dependent variable: Informal %		
	Quartile 1 Lowest %	Quartile 2 and 3	Quartile 4 Highest %
Change in Margin	−0.096	−0.275	6.673***
	(1.309)	(1.482)	(1.636)
Change in ln(Voters)	−3.201	0.757	−0.027
	(2.806)	(1.490)	(2.032)
Change in ln(N Options)	1.236***	1.204***	2.393***
	(0.385)	(0.321)	(0.362)
Changed Division	−0.252*	−0.245***	0.024
	(0.135)	(0.089)	(0.123)
Covariates	✓	✓	✓
Time FE	✓	✓	✓
Polling Place FE	✓	✓	✓
Mean of Informal %	5.66	5.50	4.52
Observations (Polling Place × Year)	5,973	11,667	5,636
Adjusted R^2	0.543	0.689	0.697

Note:

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

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.

Note: 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. *p<0.1; **p<0.05; ***p<0.01

the election can be controlled by informed voters 1996. The general relationship between education and the level of informal voting is as implied by the theory, but the statistically significant parameter estimate for Change in Margin is counter to their theory. This suggests that the literature that focuses on the mental processing costs associated with correctly completing a ballot may provide the best explanation for the behavior seen here.

B. Robustness checks

Table 9, expands the main results by including the DID v2 and DDD approaches. In this table, Panel A shows the DID v1 results, Panel B shows the DID v2 results and Panel C shows the DDD results. Within each of these panels, Column 1 presents the standard results, Column 2 presents the propensity score matched results and Column 3 presents the distance limited results. This aligns with Columns 4, 5 and 6 of the main results table. The alternative approaches do not affect DID v2 and so only a single set of results is presented in Panel B. Across all models and data restrictions, the parameter estimates for *Change in ln(Options)* is positive and statistically significant. Further, the parameter estimates for *Change in ln(Voters)* and *Change in Margin* are generally not statistically significant and vary in sign. Again, this is a rejection of H1 and H2, and a confirmation of H3.

Turning to Column 4 and 5 of Table 9, a potential alternative approach is to change or restrict the definition of the control group. In the results above, treated Polling Places are those that changed electoral division while the control group are Polling Places that did not change electoral division. In this control group, there are Polling Places which are located in electoral divisions that have a boundary change. An argument could be made that these control Polling Places do experience some form of treatment, although not as extreme as Polling Places that change electoral divisions. To address this issue, Column 4, in the results below, removes these Polling Places from the control group. That is, in Column 4, the control group is made up of only Polling Places that are located in electoral divisions that did not experience any boundary change for that year. In contrast, Column 5 restricts the control group to be only Polling Places in electoral divisions that had boundary changes. Although these two approaches use disjoint control groups, they provide results which are very similar in nature to those seen in the main results. Across all models and data restrictions, the parameter estimates for Change in ln(Options) is positive and statistically significant while the parameter estimates for Change in Margin and Change in ln(Voters) are generally not statistically significant and vary in sign. Propensity score matching is not used in Column 4 or 5 of Table 9.

Table 10 presents the results of a placebo test where the dependent variable is changed to variables where there is not expected to be a genuine treatment effect. The variables that have been selected are related to the election process but should not, theoretically, be affected by the number of options available on

Table 9—Alternative model specifications and control groups

	Dependent variable: Informal %				
	Standard	Propensity Score Matched	Distance Limited	Controls are non-treated divisions	Treated Divisions only
	(1)	(2)	(3)	(4)	(5)
Panel A - DID (House, treated and control Polling Places)					
Change in Margin	0.448	0.029	−0.017	0.441	−1.373
	(0.881)	(1.097)	(0.878)	(0.994)	(1.413)
Change in ln(Voters)	−0.186	−0.156	0.479	−0.664	−1.101
	(1.139)	(1.532)	(1.151)	(1.303)	(2.137)
Change in ln(N Options)	1.319***	1.414***	1.316***	1.239***	1.332***
	(0.213)	(0.292)	(0.209)	(0.239)	(0.342)
Changed Division	−0.249***	−0.384***	−0.145**	−0.424***	0.022
	(0.065)	(0.098)	(0.065)	(0.078)	(0.129)
Observations (Polling Place × Year)	23,276	12,969	7,457	16,229	8,482
Adjusted R^2	0.663	0.679	0.770	0.686	0.646
Panel B - DID (House and Senate, treated Polling Places only)					
Change in Margin			1.077		
			(0.944)		
Change in ln(Voters)			0.009		
			(1.246)		
Change in ln(N Options)			1.705***		
			(0.217)		
Observations (Polling Place × Year × House)			2,870		
Adjusted R^2			0.820		
Panel C - DDD (House and Senate, treated and control Polling Places)					
Change in Margin	1.164	1.018	0.660	1.276*	0.300
	(0.739)	(0.818)	(0.726)	(0.751)	(0.860)
Change in ln(Voters)	−0.192	−0.019	0.050	−0.310	−0.778
	(0.911)	(1.028)	(0.927)	(0.945)	(1.083)
Change in ln(N Options)	1.484***	1.540***	1.518***	1.458***	1.596***
	(0.187)	(0.209)	(0.181)	(0.187)	(0.220)
Changed Division	−0.102**	−0.182***	−0.034	−0.196***	0.002
	(0.043)	(0.058)	(0.043)	(0.047)	(0.072)
Observations (Polling Place × Year × House)	46,552	29,018	14,914	32,458	16,964
Adjusted R^2	0.650	0.662	0.719	0.630	0.731

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the Polling Place level.
Models include covariates and Polling Place fixed effects. Time period fixed effects are included for models in Panel A and Panel C.
Models shown in Panel B produce the same results across all specifications.

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the Polling Place level. Models include covariates and Polling Place fixed effects. Panel A shows the DID v1 results, Panel B shows the DID v2 results and Panel C shows the DDD results. Time period fixed effects are included for models in Panel A and Panel C. Models shown in Panel B produce the same results across all specifications and do not include a Changed Division variable. Column 1 presents the standard results, Column 2 presents the propensity score matched results and Column 3 presents the distance limited results. In Column 4, the control group is made up of only Polling Places in electoral divisions that did not experience a boundary change. In Column 5, the control group is restricted to be only Polling Places in electoral divisions that had boundary changes. Propensity score matching is not applied to columns 1, 3, 4 or 5. *p<0.1; **p<0.05; ***p<0.01

the ballot, the number of voters in the electoral division or the margin in the electoral division. Column 1 and Column 2 report the results of a model where the dependent variables are the total number of votes recorded in the House and Senate respectively — due to compulsory voting in Australia, this shouldn't be directly affected by political conditions. Columns 3, 4, and 5 focus on outcomes in the Senate. Column 3 focuses on the percent of informal votes in the Senate, Column 4 focuses on the percent of Donkey votes in the Senate while Column 5 looks at the share of votes for non-major parties (i.e. not Liberal, National, Labor or the Greens). In this case, Donkey voting is defined as when a voter votes for the first party on the ballot as their first preference. Each of these outcomes should not be affected by changes in the number of options, voters or margin in the House.

The treatment is not found to be statistically significant at conventional levels in 27 of the 30 parameters. A statistically significant result is found for the relationship between change in the number of voters and Donkey voting in the Senate as well as voting for non-major parties in the Senate. There does not appear to be a ready theoretical explanation for this behavior and the estimated parameter values are fairly small in magnitude. Overall, this set of placebo tests provides supporting evidence that the treatment effect estimated in the main results is a genuine effect and not a chance result of noise in the data.

TABLE 10—PLACEBO TEST

	Dependent variable:				
	Total House Votes	Total Senate Votes	Senate Informal %	Senate Donkey %	Senate Other %
	(1)	(2)	(3)	(4)	(5)
Panel A - Standard					
Change in Margin	−16.705	−20.030	0.006	−0.002	−0.013
	(119.104)	(118.748)	(0.007)	(0.006)	(0.021)
Change in ln(Voters)	−61.683	−62.620	−0.005	0.034***	0.075***
	(177.595)	(177.078)	(0.008)	(0.009)	(0.029)
Change in ln(N Options)	17.587	17.198	−0.001	0.0005	−0.003
	(24.027)	(24.002)	(0.002)	(0.001)	(0.004)
Changed Division	6.753	6.928	0.0004	−0.0002	−0.012***
	(10.294)	(10.297)	(0.0004)	(0.0005)	(0.002)
Covariates	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	23,276	23,276	23,276	23,276	23,276
Adjusted R^2	0.948	0.948	0.688	0.443	0.782
Panel B - Propensity Score Matched					
Change in Margin	41.315	41.922	−0.001	0.002	0.004
	(150.220)	(149.466)	(0.009)	(0.008)	(0.024)
Change in ln(Voters)	217.067	205.293	−0.006	0.025*	0.111***
	(254.236)	(254.234)	(0.012)	(0.014)	(0.038)
Change in ln(N Options)	0.786	0.331	−0.001	0.002	0.006
	(32.084)	(32.084)	(0.002)	(0.002)	(0.006)
Changed Division	9.329	9.571	0.0003	−0.0004	−0.013***
	(14.186)	(14.179)	(0.001)	(0.001)	(0.002)
Covariates	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	12,969	12,969	12,969	12,969	12,969
Adjusted R^2	0.946	0.946	0.725	0.389	0.804

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the Polling Place level. The dependent variable changes between each column, for Column 1 it is the total number of votes recorded in the House; Column 2 is the total number of votes recorded in the Senate; Column 3 is the percent of informal votes in the Senate; Column 4 is the percent of Donkey votes in the Senate; Column 5 looks at the share of votes for non-major parties (i.e. not Liberal, National, Labor or the Greens). Donkey voting is defined as when a voter votes for the first party on the ballot as their first preference. *p<0.1; **p<0.05; ***p<0.01

C. Alternative stories

This section considers some alternative stories that could explain the results seen in the previous sections. First, there is the possibility that there are other variables not included in the main results that are causing the observed effect for the change in number of options. One possibility is that a strong incumbent candidate may also affect the likelihood of submitting an informal vote and incumbency may be correlated with aspects such as margin and number of options.

The role of incumbency advantage in elections is frequently discussed in the literature. Papers such as Ansolabehere, Snyder Jr and Stewart III (2000), Friedman and Holden (2009), Desposato and Petrocik (2003) all analyze incumbency advantage with reference to electorate boundaries while papers such as Lee (2008) and Carson, Engstrom and Roberts (2007) consider the source of incumbency advantage more broadly.

To account for the potential role of incumbency, the results in Table 11 include variables *ln(Tenure)* and *Change in ln(Tenure)*. Tenure is defined as the number of years that the incumbent has been representing the electoral division at the time of the election. This variable is logged and differences applied in order to make its interpretation align with other variables focused on in the previous section.

Another variable added in the results shown in Table 11 is the share of votes for progressive parties (defined as first preference votes for the ALP and Green parties). The potential effect of the strength of progressive parties on voting has been tested by Hill and Jones (2017), who show that progressive parties spend more on minorities relative to conservative parties, and Pettersson-Lidbom (2008) who find that progressive parties spend and tax 2-3% more than right wing parties. These findings are important as Bechtel and Hainmueller (2011) show that voters respond to increases in expenditure by increasing their vote for incumbents for at least 2 rounds of future elections. As a result, high progressive share of the vote may be associated with strong preferences for certain groups that experience the benefits of expenditure increases under progressive incumbents and this may affect the level of informal voting.

The results do not show statistical significance for either the *Change in ln(Tenure)* or the *Change in Progressive Vote Share* variables. The results for *Change in Margin*, *Change in ln(Voters)* and *Change in ln(N Options)* also remain similar to those in the main results. This suggests that the possible stories described above do not affect the results presented earlier.

An additional alternative story is related to the complexity of defining and understanding the margin in an Instant Runoff Voting system – as is used in the House in Australia. In this system, the margin is defined based on the share of votes in the final round of voting not on the share of first preference votes. This

Table 11—Main results – Including Tenure and Progressive Share

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. covariates	Fixed Effects	DID Standard	DID Propensity Score Matched	DID Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	0.744**	0.048	−3.530***			
	(0.292)	(0.284)	(0.363)			
ln(Voters)	1.925***	1.429***	0.898**			
	(0.147)	(0.156)	(0.352)			
ln(N Options)	0.789***	0.876***	2.157***			
	(0.051)	(0.054)	(0.057)			
ln(Tenure)	−0.086***	−0.062***	−0.056***			
	(0.021)	(0.021)	(0.021)			
Progressive Vote Share (%)	0.620***	0.229**	1.252***			
	(0.112)	(0.114)	(0.123)			
Change in Margin				0.533	−0.173	0.066
				(0.904)	(1.176)	(0.912)
Change in ln(Voters)				−0.273	−0.165	0.394
				(1.145)	(1.525)	(1.165)
Change in ln(N Options)				1.336***	1.385***	1.315***
				(0.213)	(0.289)	(0.211)
Change in ln(Tenure)				−0.027	−0.0002	−0.050
				(0.068)	(0.094)	(0.068)
Change in Progressive Vote Share				0.280	1.608	0.904
				(0.950)	(1.383)	(0.974)
Changed Division				−0.284***	−0.538***	−0.239**
				(0.105)	(0.153)	(0.105)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Clusters	6,184	6,184	6,184	6,184	5,455	2,187
Treated Observations	NA	NA	NA	1,435	1,177	1,435
Control Observations	NA	NA	NA	21,841	11,760	6,219
Observations	23,476	23,476	23,476	23,263	12,937	7,641
Adjusted R^2	0.371	0.430	0.703	0.663	0.678	0.769

Note: 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; Column 3 introduces fixed effects for Polling Place. Columns 4-6 shows the results of DID v1. Column 4 presents the main output, which uses voting in the House with treated Polling Places being those that moved between electoral divisions and control Polling Places being those that did not move. Column 5 shows the same model but after propensity score matching to make the observable characteristics of treated and control groups similar has been applied; Column 6 shows results where the sample is limited to Polling Places within 2.5 kilometers of treated Polling Places. A variation was tried where the change in progressive vote share was split into an increase and a decrease variable, this did not materially affect the results presented above. *p<0.1; **p<0.05; ***p<0.01

may make it more difficult for voters to understand and respond to the expected margin when placing their vote (as it requires some calculation of the flow of preferences throughout the runoff process). To address this, Table 12 presents a version of the main results where the margin is defined as the margin measured on first preferences (*i.e.* the number one preference on the ballots).

The results in Table 12 are similar to the main results with the exception that Change in Margin is statistically significant at the 10% level of significance in one specification. This result does not, however, confirm H1 as the first preference margin is not related to the chance of a voter being pivotal and so, in theory, shouldn't affect decisions around the benefits of casting a vote.

Finally, there is the possibility that, voters on the margin between a decision to turnout or vote informally may be responsible for the results. The analysis in Table 13 reproduce the approach of the Main Results but increases the rate of non-voting by 1 percentage point on the assumption that all of this increase comes from voters who submit an informal ballot. The results are similar to those seen 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 suggests that this potential explanation is not responsible for the findings.

The results in this section confirm that the main results are retained even when considering a range of other potential explanations and variables that could affect decisions around informal voting.

V. Additional Results: types of informal voting

There are a number of ways in which a ballot can result in an informal vote. For example, both a completely blank ballot and a ballot where the voter writes their name are recorded as informal. In general, 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.

These categories can be broadly split between intentionally informal votes and unintentionally informal votes. Unintentional informal votes are defined as ballots with incomplete numbering, non-sequential numbering, ticks and crosses, and ballots in which the voter is identified; all other informal votes are classified as intentionally informal (Australian Electoral Commission, 2016*b*).

Data on the type of informal vote isn't systematically gathered or available at a Polling Place level and so causal analysis isn't possible given the framework set out in Section III.[9] The AEC has, however, undertaken reviews of informal

[9]Communication with the AEC indicated that this data is not available even in an unpublished format.

Table 12—Main results – Margin Defined on First Preferences

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. covariates	Fixed Effects	DID Standard	DID Propensity Score Matched	DID Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	0.942***	0.617***	0.232			
	(0.157)	(0.124)	(0.163)			
ln(Voters)	3.276***	1.243***	0.592*			
	(0.192)	(0.158)	(0.349)			
ln(N Options)	0.478***	0.890***	2.177***			
	(0.066)	(0.053)	(0.057)			
Change in Margin				0.724*	0.658	0.519
				(0.401)	(0.559)	(0.405)
Change in ln(Voters)				−0.136	−0.417	0.518
				(1.137)	(1.533)	(1.160)
Change in ln(N Options)				1.360***	1.444***	1.365***
				(0.209)	(0.293)	(0.207)
Changed Division				−0.258***	−0.400***	−0.153**
				(0.065)	(0.097)	(0.065)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Clusters	6,184	6,184	6,184	6,184	5,455	2,187
Treated Observations	NA	NA	NA	1,440	1,177	1,440
Control Observations	NA	NA	NA	21,841	11,760	6,219
Observations	24,119	23,476	23,476	23,281	12,937	7,659
Adjusted R^2	0.027	0.431	0.698	0.663	0.678	0.769

Note: 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; Column 3 introduces fixed effects for Polling Place. Columns 4-6 shows the results of DID v1. Column 4 presents the main output, which uses voting in the House with treated Polling Places being those that moved between electoral divisions and control Polling Places being those that did not move. Column 5 shows the same model but after propensity score matching to make the observable characteristics of treated and control groups similar has been applied; 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 13—SENSITIVITY ANALYSIS – INCREASING NON-VOTING BY 1% FROM INFORMAL VOTERS

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. covariates	Fixed Effects	DID Standard	DID Propensity Score Matched	DID Distance Limited
Margin	2.168***	−0.120	−3.220***			
	(0.349)	(0.284)	(0.309)			
ln(Voters)	3.340***	1.428***	0.905***			
	(0.193)	(0.157)	(0.303)			
ln(N Options)	0.524***	0.868***	2.092***			
	(0.067)	(0.053)	(0.049)			
Change in Margin				0.483	0.054	0.012
				(0.762)	(0.839)	(0.757)
Change in ln(Voters)				−0.172	−0.239	0.497
				(0.989)	(1.167)	(0.996)
Change in ln(N Options)				1.327***	1.426***	1.323***
				(0.184)	(0.225)	(0.180)
Changed Division				−0.248***	−0.393***	−0.143**
				(0.057)	(0.075)	(0.056)
Covariates		✓	✓	✓	✓	✓
Time FE		✓	✓	✓	✓	✓
Polling Place FE			✓	✓	✓	✓
Clusters	6,184	6,184	6,184	6,184	5,455	2,187
Treated Observations	NA	NA	NA	1,435	1,177	1,435
Control Observations	NA	NA	NA	21,841	11,760	6,219
Observations	24,119	23,476	23,476	23,276	12,937	7,654
Adjusted R^2	0.027	0.431	0.702	0.664	0.678	0.769

Note: 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; Column 3 introduces fixed effects for Polling Place. Columns 4-6 shows the results of DID v1. Column 4 presents the main output, which uses voting in the House with treated Polling Places being those that moved between electoral divisions and control Polling Places being those that did not move. Column 5 shows the same model but after propensity score matching to make the observable characteristics of treated and control groups similar has been applied; 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

voting at the electoral division level for a number of recent elections and this data provides additional insight into how socioeconomic characteristics correlate with informal voting.

Table **??** provides results of regressions where the dependent variable is the percentage (0-100) of different types of informal votes. In this analysis the observation level is the electoral division and data is included for the elections of 2007, 2010, and 2013. The analysis aggregates 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.

Of note in these results is that more options on the ballot is associated with higher levels of blank, non-sequential and incomplete informal votes, indicating the challenges and costs of completing larger ballots. A higher margin is also associated with higher levels of blank ballots, a type of intentional informal voting.

Turning to socioeconomic factors, higher levels of income are associated with more blank ballots, potentially relating to the opportunity cost of completing a ballot. Areas with higher levels of English as a second language tend to have higher levels of informality across the board but with a notably strong effect on increasing the number of blank, one-only, non-sequential and non-numeric ballots -– likely reflecting errors or misunderstandings in completing the ballot. This is similar in nature 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, but the effect is strongest on informal votes where there is only a one on the ballot, where numbering is non-numeric and where there is a blank ballot. Education levels appear to have a weaker relationship with non-sequential numbering on ballots and incomplete numbering. As non-sequential and incomplete numbering tend to represent unintentional informal voting, these results support the idea that better educated voters may consider the strategic implications of their voting decisions more closely than other groups. This interpretation is similar to results seen in Driscoll and Nelson (2014) and Cohen (2018) who both find evidence that, in Latin America, voters that are high in knowledge protest poor government performance by submitting blank ballots in compulsory voting elections.

While the results in this section do not have a causal interpretation, they provide further support for the main findings. In particular, these results show that more options on a ballot are associated with higher levels of blank, non-sequential and incomplete informal votes and that voters in areas with higher levels of education are less likely to unintentionally vote informally, indicating more strategic voting

TABLE 14—CONTRIBUTORS TO TYPES OF INFORMAL VOTING (AS PERCENTAGE OF ALL VOTES IN ELECTORAL DIVISION)

	Dependent variable: Informal type share of total votes							
	Intentional		Unintentional					
	Blank	Scribble	One only	Non-sequential	Non-numeric	Incomplete	Other	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Margin (%)	1.268*	0.467	−0.578	0.598	0.425	−0.559	−0.037	1.583
	(0.679)	(0.489)	(1.879)	(0.898)	(0.670)	(0.436)	(0.176)	(3.259)
ln(Voters)	−0.021	0.111***	−0.133	−0.007	−0.079***	0.013	−0.0004	−0.117
	(0.060)	(0.036)	(0.082)	(0.044)	(0.026)	(0.017)	(0.010)	(0.171)
ln(N Options)	0.829***	−0.047	−0.565**	1.771***	−0.208*	0.582***	0.117***	2.479***
	(0.137)	(0.076)	(0.271)	(0.241)	(0.121)	(0.090)	(0.029)	(0.579)
Median Age	0.011	0.012	−0.037	−0.017	0.0001	−0.028***	−0.001	−0.060
	(0.020)	(0.013)	(0.045)	(0.024)	(0.015)	(0.010)	(0.004)	(0.085)
ln(Mean Income (000))	1.851***	0.051	2.812**	−0.794	2.067***	−0.177	−0.224*	5.585***
	(0.419)	(0.246)	(1.257)	(0.634)	(0.468)	(0.241)	(0.127)	(1.994)
Unemployment (%)	0.017	−0.004	0.322***	−0.039	0.106**	0.031	−0.006	0.428**
	(0.041)	(0.024)	(0.106)	(0.052)	(0.041)	(0.023)	(0.009)	(0.198)
Population Density	0.0002**	0.00003	0.0005***	−0.00004	0.0002***	0.00001	0.00002*	0.001***
	(0.0001)	(0.00004)	(0.0002)	(0.0001)	(0.0001)	(0.00004)	(0.00001)	(0.0003)
Population Growth (%)	−0.060	0.130***	−0.328***	0.037	−0.141***	0.013	0.043***	−0.305*
	(0.040)	(0.032)	(0.108)	(0.059)	(0.041)	(0.023)	(0.011)	(0.185)
Population Decline (%)	0.377	0.768**	−2.007**	0.838	−0.487	−0.059	0.230***	−0.340
	(0.509)	(0.342)	(0.935)	(0.597)	(0.320)	(0.225)	(0.088)	(1.729)
ln(House Value (000))	−0.322*	−0.130	0.971**	0.313	−0.129	0.199*	−0.028	0.874
	(0.175)	(0.098)	(0.420)	(0.230)	(0.172)	(0.102)	(0.038)	(0.864)
English 2nd Language (%)	0.039***	0.017***	0.056***	0.025***	0.027***	0.009***	0.003**	0.176***
	(0.005)	(0.002)	(0.014)	(0.005)	(0.005)	(0.003)	(0.001)	(0.025)
Tertiary Degree (%)	−0.093***	−0.021***	−0.172***	−0.019	−0.074***	−0.010*	0.002	−0.386***
	(0.010)	(0.006)	(0.028)	(0.012)	(0.011)	(0.005)	(0.003)	(0.050)
Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Share of informal (%)	23.6	16.1	28.1	13.9	11.3	4.5	2.6	100.0
Observations	429	429	429	429	429	429	429	429
Adjusted R^2	0.740	0.552	0.501	0.450	0.575	0.483	0.293	0.677

Note: *p<0.1; **p<0.05; ***p<0.01

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the electoral division level. 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. The dependent variable is defined as the number of a specific type of informal vote divided by the total number of votes (both valid and informal). Time period fixed effects use 2007 as the reference year. *p<0.1; **p<0.05; ***p<0.01

behavior.

VI. Conclusion

We find support for the hypothesis that, when there are more options available on the ballot, informal voting will be higher (H3) but we do not find support for the hypotheses that, when the expected margin in an election is higher, informal voting will be higher (H1); nor that, when there are more voters in an electoral division, informal voting will be higher (H2).

Once a voter has incurred travel and time costs to arrive at the polling place and is considering whether to vote or not, their decision is largely driven by the costs of thinking about and ranking their options rather than factors that may affect 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 complexity costs. The choice not to vote seems to be primarily affected by cost of voting not the potential benefit.

The subgroup analysis for voters with higher levels of education indicates that their response to the number of candidates is similar to those with lower levels of education but their response to competitiveness is more aligned with theory, this suggests that better educated voters may consider the strategic implications of their voting decisions more closely than other groups but that the behavior around the number of options on a ballot isn't explained by differences in levels of education. This is also supported by non-causal analysis which shows that voters in better educated areas are less likely to make involuntary informal votes.

The lack of support for H1 and H2 is in contrast to the theoretical literature, where an increase in the margin or the number of voters should lead to an increase in informal voting. These findings are also contrary to those in recent papers, such as Lyytikäinen and Tukiainen (2019). The result that the number of options available on the ballot leads to an increase in informal voting, is also contrary to recent findings reported in Nagler (2015).

One potential reason that the results in this paper are contrary to those seen in other recent papers is that no previous paper has tested these contributors to voting behavior in the one model. Further, many tests of both these theories haven't focused strongly on a research design that makes use of exogenous variation in the key parameters of interest.

These results are more in line with those seen in Iyengar and Kamenica (2010) or Iyengar and Lepper (2000) which show that, in general cases, larger choice sets can lead decision-makers to abstain from making a decision. These findings are also similar to those in Augenblick and Nicholson (2016) who show that more decisions on the same ballot can increase abstention. The results show mixed support for the implications of Feddersen and Pessendorfer's theory of voter abstention,

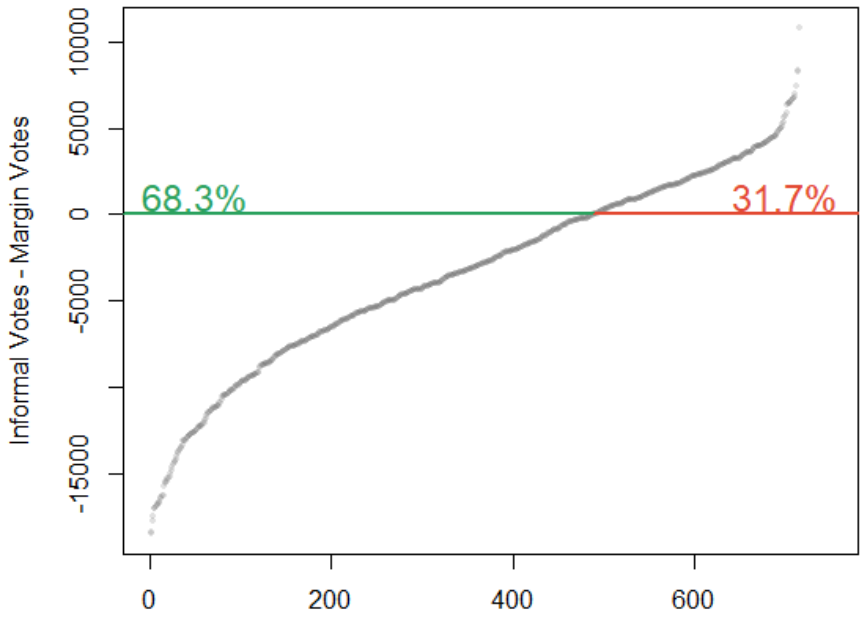

FIGURE 7. DIFFERENCE BETWEEN NUMBER OF INFORMAL VOTES AND MARGIN

Note: Data is at the electorate level covering the years 2004-2016.

Source: Authors' calculations based on multiple data sources from the Australian Electoral Commission and Australian Broadcasting Corporation (n.d.).

where uninformed voters abstain to allow outcome of the election to be controlled informed voters 1996. The general relationship between education and the level of informal voting is as implied by their theory, whereby more informed voters are less likely to abstain from voting, but the statistically significant parameter estimate for Change in Margin for highly educated voters is counter to their theory.

Going beyond the implications for theory, the results are also informative from a practical point of view. This is because, in some electoral divisions in Australia, the number of informal votes can be greater than the margin. For example, in the 2016 election, the seat of Gilmore had a margin of around 400 votes with almost 4,000 informal votes while, in the seat of Hindmarsh, the margin was around 430 votes and the informal vote count was around 3,000. Figure 10 shows that, in total, around 32% of contests have more informal votes than the margin.

The frequency of informal voting exceeding margin suggests that informal voting may have practical consequences for outcomes of elections. Policies which affect the level of informal voting may, therefore, affect the final composition of Parliament and the economic and social policies that are ultimately implemented by the government. The findings in this paper suggest that increasing the rate of formal voting could be achieved by strategies that make it easier for voters to research, understand and rank the candidates on the ballot.

One potential approach to simplify voting in Australua would be to allow a non-exhaustive ranking of candidates in the House, as is done in the Senate and in some Australian states. While data on the full preference listing for ballots is not available, some data on the flow of preferences in counting is made available by the AEC, this allows for a reconstruction of how often preferences are used in counting votes. Analysis of this data shows that, of the roughly 13.5 million ballots counted in the 2016 election, approximately 10.5 million ended up being counted against their first preference. Further, around 12.5 million voters (92% of all voters) had to rank five or more candidates on their ballot while fifth or higher preferences were only used in around 10,500 ballots (less than 0.1% of all ballots). This suggests that the requirement to exhaustively rank all candidates on the ballot is not often used when counting votes and may contribute significantly (in both statistical and practical terms) to the level of informal voting. Moving to a simplified process for stating preferences in the House could generate benefits in terms of reducing informal voting without significant costs.

Finally, considering paths for future research in this area, an alternative explanation of the results is that the presence of compulsory voting in Australia, by eliminating the role of travel and time costs in the voting decision, minimizes the role of margin and number of voters in the voting decision while leaving choice costs unchanged. That is, under compulsory voting, voters no longer consider how likely they are to affect the results of the election but only focus on their choice costs. To address whether this explanation is correct would require applying a similar research design to elections held under a voluntary voting regime, such as in the United States. Assigning a causal interpretation to any differences between two countries would be problematic as the design of electoral systems and voting behavior are likely to be jointly determined.

REFERENCES

ABS. 2016*a*. "1270.0.55.001 - Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016." *http://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.001*, (Accessed on 11/09/2016).

ABS. 2016*b*. "1379.0.55.001 - National Regional Profile, 2010-14." *http://www.abs.gov.au/ausstats/abs@.nsf/mf/1379.0.55.001*, (Accessed on 11/09/2016).

Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

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.

Augenblick, Ned, and Scott Nicholson. 2016. "Ballot position, choice fatigue, and voter behaviour." *The Review of Economic Studies*, 83(2): 460–480.

Australian Broadcasting Corporation. n.d.. "Elections." *Available at http://www.abc.net.au/news/elections/ (2019-02-26).*

Australian Electoral Commission. 2009. "Research Report 11 - Analysis of Informal Voting (House of Representatives 2007 Election) - Australian Electoral Commission." *https://www.aec.gov.au/About_AEC/research/paper_11/*, (Accessed on 07/13/2020).

Australian Electoral Commission. 2011. "Analysis of Informal Voting, House of Representatives, 2010 federal election - Australian Electoral Commission." *https://www.aec.gov.au/About_AEC/research/paper12/*, (Accessed on 07/13/2020).

Australian Electoral Commission. 2016*a*. "Analysis of informal voting - 2013 House of Representatives elections - Australian Electoral Commission." *https://www.aec.gov.au/About_AEC/research/paper13.htm*, (Accessed on 07/13/2020).

Australian Electoral Commission. 2016*b*. "Analysis of informal voting (House of Representatives 2016 Election) - Australian Electoral Commission." *https://www.aec.gov.au/About_AEC/research/analysis-informal-voting-2016-election.htm*, (Accessed on 07/13/2020).

Australian Electoral Commission. 2016*c*. "How to vote - Australian Electoral Commission." *https://www.aec.gov.au/Voting/How_to_Vote/*, (Accessed on 07/13/2020).

Australian Electoral Commission. 2017. "Voter turnout: 2016 House of Representatives and Senate elections." Australian Electoral Commission.

Australian Electoral Commission. 2018. "Federal electoral boundary GIS data for free download - Australian Electoral Commission." *https://www.aec.gov.au/Electorates/gis/gis_datadownload.htm*, (Accessed on 07/13/2020).

Australian Electoral Commission. 2019. "Direct Enrolment and Update - Australian Electoral Commission." *https://www.aec.gov.au/Enrolling_to_vote/About_Electoral_Roll/direct.htm*, (Accessed on 07/13/2020).

Australian Electoral Commission. n.d.. "Size of the electoral roll and enrolment rate 2019 - Australian Electoral Commission." *https://www.aec.gov.au/Enrolling_to_vote/Enrolment_stats/national/2019.htm*, (Accessed on 07/13/2020).

Australian Electoral Commission. n.d.. "Tally room archive - Australian Electoral Commission." *https://results.aec.gov.au/*, (Accessed on 07/13/2020).

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

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.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics*, 119(1): 249–275.

Blais, André. 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.

Card, David. 1992. "Using regional variation in wages to measure the effects of the federal minimum wage." *ILR Review*, 46(1): 22–37.

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(4): 455–469.

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*, 289–301.

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."

De Paola, Maria, and Vincenzo Scoppa. 2014. "The impact of closeness on electoral participation exploiting the Italian double 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(1): 18–32.

Downs, Anthony. 1957. "An economic theory of political action in a democracy." *Journal of Political Economy*, 65(2): 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, Timothy J, and Wolfgang Pesendorfer. 1996. "The swing voter's curse." *The American economic review*, 408–424.

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

Fraga, Bernard L. 2016. "Redistricting and the causal impact of race on voter turnout." *The Journal of Politics*, 78(1): 19–34.

Friedman, John N, and Richard T Holden. 2009. "The rising incumbent reelection rate: What's gerrymandering got to do with it?" *The Journal of Politics*, 71(2): 593–611.

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.

Geys, Benny. 2006. "Explaining voter turnout: A review of aggregate-level research." *Electoral studies*, 25(4): 637–663.

Gimpel, James G, and Jason E Schuknecht. 2003. "Political participation and the accessibility of the ballot box." *Political Geography*, 22(5): 471–488.

Godefroy, Raphael, and Emeric Henry. 2016. "Voter turnout and fiscal policy." *European Economic Review*, 89: 389–406.

Haspel, Moshe, and H Gibbs Knotts. 2005. "Location, location, location: Precinct placement and the costs of voting." *The Journal of Politics*, 67(2): 560–573.

Hayes, Danny, and Seth 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(3): 404–412.

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.

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

Ho, Daniel E, Kosuke Imai, Gary King, and Elizabeth A Stuart. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political analysis*, 15(3): 199–236.

Hoffman, Mitchell, Gianmarco León, and María Lombardi. 2017. "Compulsory voting, turnout, and government spending: Evidence from Austria." *Journal of Public Economics*, 145: 103–115.

Horiuchi, Yusaku, and Alexandra Lange. 2019. "Randomized Ballot Order Can Increase Invalid Votes: Evidence from Australia." *Available at SSRN 2817549*.

IDEA, International. n.d.. "Invalid votes — International IDEA." *https://www.idea.int/data-tools/question-view/443*, (Accessed on 07/13/2020).

Indridason, Indridi H. 2008. "Competition & turnout: the majority run-off as a natural experiment." *Electoral Studies*, 27(4): 699–710.

Iyengar, Sheena S, and Emir Kamenica. 2010. "Choice proliferation, simplicity seeking, and asset allocation." *Journal of Public Economics*, 94(7-8): 530–539.

Iyengar, Sheena S, and Mark R Lepper. 2000. "When choice is demotivating: Can one desire too much of a good thing?" *Journal of personality and social psychology*, 79(6): 995.

Jackman, Simon. 2015. "Assessing the Current Wisconsin State Legislative Districting Plan."

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. 2020. "Voter turnout and preference aggregation." *American Economic Journal: Microeconomics*.

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

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

Levine, David K, and Thomas R Palfrey. 2007. "The paradox of voter participation? A laboratory study." *American political science Review*, 101(1): 143–158.

Lijphart, Arend. 1998. "The problem of low and unequal voter turnout-and what we can do about it."

Lundie, Rob. n.d.. "Electronic voting at federal elections – Parliament of Australia." *https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook45p/ElectronicVoting*, (Accessed on 07/13/2020).

Lyytikäinen, Teemu, and Janne Tukiainen. 2019. "Are voters rational?" *European Journal of Political Economy*, 59: 230–242.

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

Muller, Damon. n.d.. "The 2016 federal election – Parliament of Australia." *https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook45p/FederalElection2016*, (Accessed on 07/13/2020).

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(1): 62–78.

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

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

Power, Timothy J, and J 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(1): 25–42.

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

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

Stockemer, Daniel. 2017. "What affects voter turnout? A review article/meta-analysis of aggregate research." *Government and Opposition*, 52(4): 698–722.

Wikipedia. 2018. "List of Australian federal elections - Wikipedia." *https://en.wikipedia.org/wiki/List_of_Australian_federal_elections*, (Accessed on 07/13/2020).

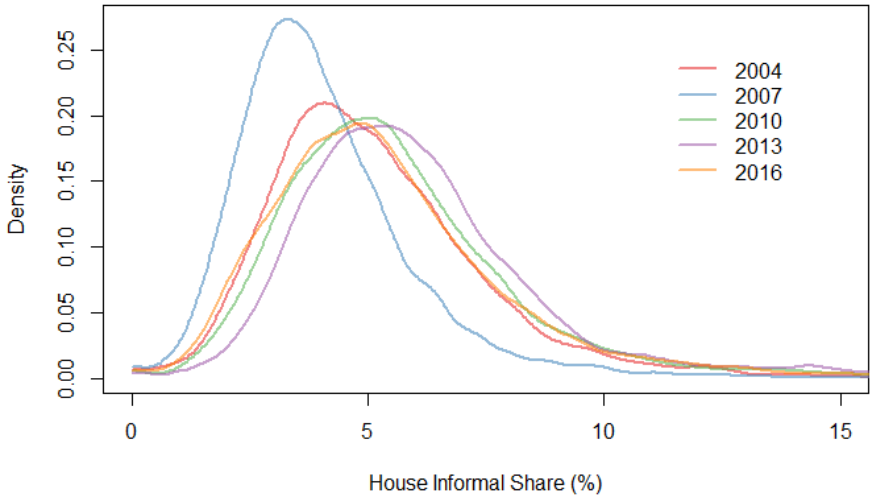

Figure A1. Informal Vote Share in House

Appendix A - Distributions of some variables of interest

A number of kernel density plots for the level of informal voting are shown in Figure A, below. The lower level of informal voting in 2007 is noteworthy, a potential explanation for this is that the 2007 election saw a long serving government replaced by a new government and so voter's may have been more interested and engaged in the 2007 election than other elections, however the 2013 election also saw a change in government without an associated reduction in informality.

The distribution of electoral division sizes over time in shown in Figure A. The figure shows that the size of electoral divisions has been increasing over time, associated with population growth.

The distribution of the number of options in the House of Representatives is shown in Figure A.

Kernel density estimates for the margin are shown in Figure A, these demonstrate a consistency in the distribution of margin over time.

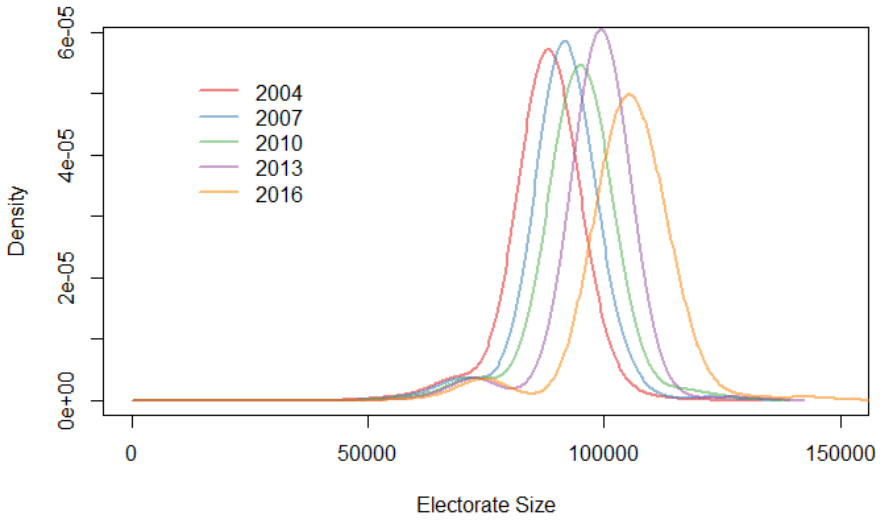

FIGURE A2. ELECTORAL DIVISION SIZE

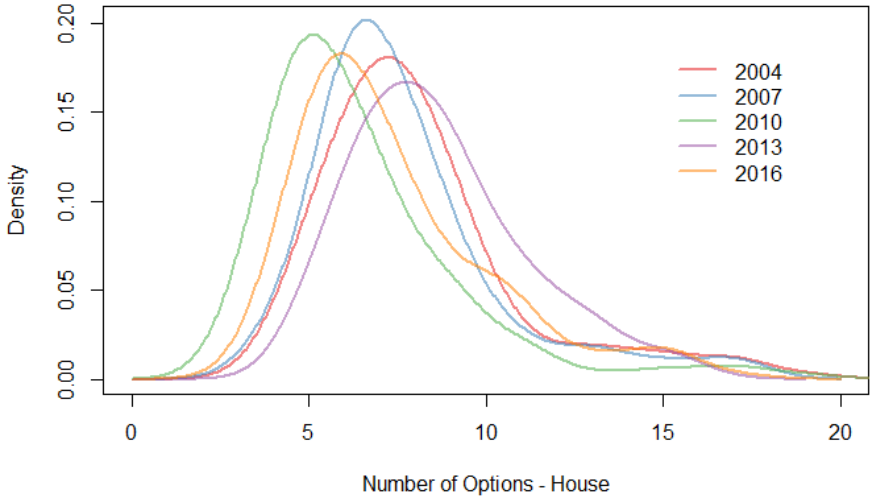

FIGURE A3. NUMBER OF OPTIONS IN HOUSE

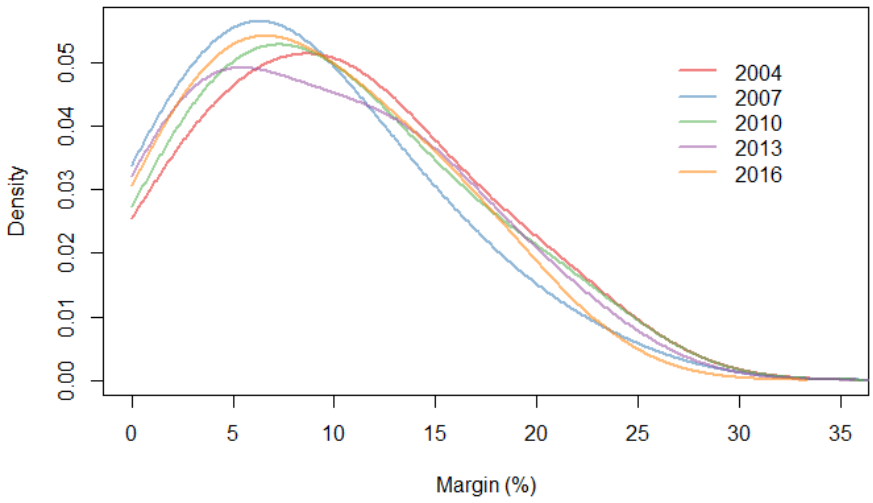

FIGURE A4. MARGIN

Appendix B - Detailed model outputs

This appendix includes full regression results for all results reported in the main body of the text.

Table B1, sets out the results of a first stage regression where, in Column 1, the dependent variable is whether a polling place changed division and, in Column 2, whether the polling place is in an electorate that changed division. If this treatment indicator could be easily predicted by observable characteristics then the research design may be problematic. The results indicate that observable characteristics are often not strongly associated with treatment and that, overall, the regression has low explanatory power, as measured by adjusted R^2. This supports use of the proposed research design and treating the change in electorate boundaries as a exogenous change.

Table B2 presents the main results of the analysis but also includes parameters estimates for all covariates except Polling Place fixed effects. The main result in this table is that the parameter estimates for *Change in Margin* and *Change in ln(Voters)* are not generally statistically significant and also change in sign while the parameter estimate for *Change in ln(N Options)* is consistently positive and statistically significant at conventional levels. This indicates that more options on the ballot is associated with higher levels of informal voting.

Tables B3-B5 show alternative versions of the main results that both include alternative control group (Columns 4 and 5) and different model specifications (Tables B4 and B5). The alternative control groups are used to test whether the fact that control booths in the main result receive some form of treatment has any effect on the overall results. The consistency of findings across all columns in Table B3 indicates that this aspect of selection of the control group does not strongly determine the findings of the analysis. Table B4 and B5 present alternative model specifications that include voting in the Senate as a form of control group. Senate voting is not affected by changes in electorate boundaries and so can potentially form a control for voting in the House of Representatives. However there are important differences between voting in both Houses that makes the Senate less preferable control group. Despite this, the findings are broadly consistent across all columns in Tables B3-B5.

Table B6 and B7 present the results of further robustness checks – a placebo test – where the dependent variable is changed to variables where there is not expected to be a genuine treatment effect. The variables that have been selected are related to the election process but should not, theoretically, be affected by the number of options available on the ballot, the number of voters in the electoral division or the margin in the electoral division. In particular, Column 1 and Column 2 report the results of a model where the dependent variables are the total number of votes recorded in the House of Representatives and Senate respectively – due to compulsory voting in Australia, this shouldn't be directly affected by political

Table B1—First Stage Regressions

	Dependent variable:	
	Polling Place Changed Division	Division had Polling Place change
	(1)	(2)
Median Age	−0.001**	−0.006***
	(0.0004)	(0.001)
Mean Income (000)	−0.00000	−0.0001
	(0.0002)	(0.0003)
Unemployment (%)	−0.002*	0.010***
	(0.001)	(0.002)
Population Density	0.00000	0.00000
	(0.00000)	(0.00000)
Population Growth (%)	0.0004	0.006***
	(0.001)	(0.002)
Population Decline (%)	0.002***	0.004***
	(0.001)	(0.001)
House value (000)	0.00000	0.00002***
	(0.00000)	(0.00001)
English 2nd Language (%)	0.001***	0.0004
	(0.0001)	(0.0003)
Tertiary degree (%)	−0.002***	−0.003***
	(0.0003)	(0.001)
Constant	0.141***	0.562***
	(0.019)	(0.036)
Observations	23,476	23,476
Adjusted R^2	0.008	0.011

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the Polling Place level. Results in the first column for 'Polling Place Changed Division' are for a linear probability model where the dependent variable is equal to one if that Polling Place changed division in a given year. Results in the second column for 'Electoral division had Polling Place change' are for a linear probability model where the dependent variable is equal to one if that Polling Place was in an electoral division where any Polling Places changed division in a given year. *p<0.1; **p<0.05; ***p<0.01

Table B2—Main Results – Full Version

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w. covariates	Fixed Effects	DID Standard	DID Propensity Score Matched	DID Distance Limited
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	2.136***	−0.130	−3.173***			
	(0.347)	(0.281)	(0.358)			
ln(Voters)	3.308***	1.418***	0.862**			
	(0.191)	(0.156)	(0.351)			
ln(N Options)	0.523***	0.864***	2.080***			
	(0.066)	(0.053)	(0.057)			
Change in Margin				0.448	0.029	−0.017
				(0.881)	(1.097)	(0.878)
Change in ln(Voters)				−0.186	−0.156	0.479
				(1.139)	(1.532)	(1.151)
Change in ln(N Options)				1.319***	1.414***	1.316***
				(0.213)	(0.292)	(0.209)
Changed Division				−0.249***	−0.384***	−0.145**
				(0.065)	(0.098)	(0.065)
2010		1.598***	2.093***	1.700***	2.013***	2.054***
		(0.029)	(0.039)	(0.037)	(0.071)	(0.072)
2013		1.582***	2.045***	2.478***	2.790***	3.149***
		(0.041)	(0.073)	(0.073)	(0.138)	(0.137)
2016		0.772***	1.762***	2.007***	2.206***	2.250***
		(0.051)	(0.099)	(0.094)	(0.173)	(0.167)
Median Age		−0.031***	−0.011	−0.015	0.004	0.068**
		(0.004)	(0.015)	(0.016)	(0.028)	(0.028)
Mean Income (000)		0.021***	−0.016***	−0.027***	−0.032***	−0.048***
		(0.002)	(0.004)	(0.004)	(0.008)	(0.007)
Unemployment (%)		0.127***	−0.104	−0.242	−0.999***	−0.426
		(0.012)	(0.188)	(0.208)	(0.326)	(0.438)
Population Density		0.0002***	0.00003	0.0001	0.0005*	0.0005***
		(0.00003)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Population Growth (%)		−0.025***	0.0002	0.018	0.046***	0.040**
		(0.009)	(0.011)	(0.012)	(0.017)	(0.018)
Population Decline (%)		0.014**	−0.005	−0.006	0.024*	0.016
		(0.006)	(0.008)	(0.008)	(0.015)	(0.013)
House Value (000)		0.0001***	−0.00004	−0.00002	0.0001	−0.00004
		(0.00003)	(0.00002)	(0.00003)	(0.0001)	(0.00003)
English 2nd Language (%)		0.071***	−0.063***	−0.106***	−0.113***	−0.108***
		(0.003)	(0.013)	(0.015)	(0.024)	(0.020)
Tertiary Degree (%)		−0.130***	0.060	−0.058	−0.095	0.058
		(0.003)	(0.076)	(0.082)	(0.123)	(0.104)
Constant	−33.900***	−13.267***				
	(2.151)	(1.760)				
Polling Place FE			✓	✓	✓	✓
Clusters (Polling Places)	6,184	6,184	6,184	6,184	5,463	2,056
Observations (Polling Place × Year)	24,119	23,476	23,476	23,276	12,969	7,457
Treated Observations	NA	NA	NA	1,435	1,177	1,435
Control Observations	NA	NA	NA	21,841	11,792	6,022
Adjusted R^2	0.027	0.430	0.701	0.663	0.679	0.770

Note: 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 B3—Alternative Model Specifications – Panel A DID v1 (House, Treated and Control Polling Places)

	Dependent variable: Informal %				
	Model specifications				
	Standard	Propensity Score Matched	Distance Limited	Controls are non-treated divisions	Treated divisions only
	(1)	(2)	(3)	(4)	(5)
Change in Margin	0.448	0.029	−0.017	0.441	−1.373
	(0.881)	(1.097)	(0.878)	(0.994)	(1.413)
Change in ln(Voters)	−0.186	−0.156	0.479	−0.664	−1.101
	(1.139)	(1.532)	(1.151)	(1.303)	(2.137)
Change in ln(N Options)	1.319***	1.414***	1.316***	1.239***	1.332***
	(0.213)	(0.292)	(0.209)	(0.239)	(0.342)
Changed Division	−0.249***	−0.384***	−0.145**	−0.424***	0.022
	(0.065)	(0.098)	(0.065)	(0.078)	(0.129)
2010	1.700***	2.013***	2.054***	1.589***	2.035***
	(0.037)	(0.071)	(0.072)	(0.050)	(0.096)
2013	2.478***	2.790***	3.149***	2.365***	
	(0.073)	(0.138)	(0.137)	(0.102)	(0.000)
2016	2.007***	2.206***	2.250***	2.072***	1.717***
	(0.094)	(0.173)	(0.167)	(0.131)	(0.251)
Median Age	−0.015	0.004	0.068**	0.016	−0.069
	(0.016)	(0.028)	(0.028)	(0.021)	(0.053)
Mean Income (000)	−0.027***	−0.032***	−0.048***	−0.026***	−0.005
	(0.004)	(0.008)	(0.007)	(0.005)	(0.010)
Unemployment (%)	−0.242	−0.999***	−0.426	−0.134	−0.552
	(0.208)	(0.326)	(0.438)	(0.265)	(0.587)
Population Density	0.0001	0.0005*	0.0005***	−0.00002	0.0002
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0002)
Population Growth (%)	0.018	0.046***	0.040**	0.008	0.004
	(0.012)	(0.017)	(0.018)	(0.018)	(0.028)
Population Decline (%)	−0.006	0.024*	0.016	−0.003	0.006
	(0.008)	(0.015)	(0.013)	(0.012)	(0.016)
House Value (000)	−0.00002	0.0001	−0.00004	−0.00005	0.00001
	(0.00003)	(0.0001)	(0.00003)	(0.00004)	(0.0001)
English 2nd Language (%)	−0.106***	−0.113***	−0.108***	−0.119***	−0.032
	(0.015)	(0.024)	(0.020)	(0.021)	(0.033)
Tertiary Degree (%)	−0.058	−0.095	0.058	0.084	−0.597***
	(0.082)	(0.123)	(0.104)	(0.111)	(0.223)
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	23,276	12,969	7,457	16,229	8,482
Adjusted R^2	0.663	0.679	0.770	0.686	0.646

Note: 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 not applied to columns 1, 3, 4 or 5. *p<0.1; **p<0.05; ***p<0.01

Table B4—Alternative Model Specifications – PANEL B – DID V2 (House and Senate, Treated Polling Places Only)

	Dependent variable: Informal %				
	Model specifications				
	Standard	Propensity Score Matched	Distance Limited	Controls are non-treated divisions	Treated divisions only
	(1)	(2)	(3)	(4)	(5)
Change in Margin			1.077		
			(0.944)		
Change in ln(Voters)			0.009		
			(1.246)		
Change in ln(N Options)			1.705***		
			(0.217)		
Polling Place × Year FE			✓		
Observations (Polling Place × Year × House)			2,870		
Adjusted R^2			0.820		

Note: 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

conditions. Columns 3, 4, and 5 focus on outcomes in the Senate. Column 3 focuses on the percent of informal votes in the Senate, Column 4 focuses on the percent of Donkey votes in the Senate while Column 5 looks at the share of votes for non-major parties (i.e. not Liberal, National, Labor or the Greens). In this case, Donkey voting is defined as when a voter votes for the first party on the ballot as their first preference. Each of these outcomes should not be affected by changes in the number of options, voters or margin in the House of Representatives.

The treatment is not found to be statistically significant at conventional levels in 27 of the 30 parameters. A statistically significant result is found for the relationship between change in the number of voters and Donkey voting in the Senate as well as voting for non-major parties in the Senate. There does not appear to be a ready theoretical explanation for this behavior and the estimated parameter values are fairly small in magnitude. Overall, this set of placebo tests provides supporting evidence that the treatment effect estimated in the main results is a genuine effect and not a chance result of noise in the data.

Table B8 and B9 present results when the data is subset according to quartiles of tertiary education. For the highest education group (Quartile 4), informal votes are less prevalent, the sign for *Change in ln(Options)* and *Change in Margin* are in line with the implications of the hypotheses, and are statistically significant.

These results indicate that, in these better-educated areas voters are more strategic, reacting to the competitiveness of their electoral division as predicted by theory. Better educated voters may therefore consider the strategic implications of their voting decisions more closely than other groups. Surprisingly, voters in these better educated areas still respond to the number of options in the same

TABLE B5—ALTERNATIVE MODEL SPECIFICATIONS – PANEL C – DDD (HOUSE ANS SENATE, TREATED AND CONTROL POLLING PLACES)

	Dependent variable: Informal %				
	Model specifications				
	Standard	Propensity Score Matched	Distance Limited	Controls are non-treated divisions	Treated divisions only
	(1)	(2)	(3)	(4)	(5)
Change in Margin	1.164	1.018	0.660	1.276*	0.300
	(0.739)	(0.818)	(0.726)	(0.751)	(0.860)
Change in ln(Voters)	−0.192	−0.019	0.050	−0.310	−0.778
	(0.911)	(1.028)	(0.927)	(0.945)	(1.083)
Change in ln(N Options)	1.484***	1.540***	1.518***	1.458***	1.596***
	(0.187)	(0.209)	(0.181)	(0.187)	(0.220)
Changed Division	−0.102**	−0.182***	−0.034	−0.196***	0.002
	(0.043)	(0.058)	(0.043)	(0.047)	(0.072)
House	1.757***	1.842***	2.176***	1.723***	1.866***
	(0.018)	(0.026)	(0.039)	(0.022)	(0.024)
2010	1.523***	1.823***	1.903***	1.314***	2.104***
	(0.026)	(0.046)	(0.052)	(0.033)	(0.056)
2013	1.583***	1.827***	2.169***	1.343***	
	(0.047)	(0.084)	(0.089)	(0.065)	(0.000)
2016	1.897***	2.155***	2.258***	1.716***	2.391***
	(0.060)	(0.105)	(0.111)	(0.081)	(0.138)
Median Age	0.005	0.004	0.068***	0.027**	−0.058**
	(0.011)	(0.017)	(0.019)	(0.013)	(0.029)
Mean Income (000)	−0.019***	−0.022***	−0.039***	−0.013***	−0.018***
	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)
Unemployment (%)	−0.131	−0.464**	0.124	−0.100	−0.219
	(0.137)	(0.212)	(0.281)	(0.162)	(0.336)
Population Density	−0.00003	0.0002	0.0002**	−0.0001	0.00001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Population Growth (%)	0.022***	0.034***	0.044***	0.015	0.009
	(0.007)	(0.010)	(0.011)	(0.011)	(0.014)
Population Decline (%)	−0.002	0.017*	0.016*	−0.0003	0.006
	(0.005)	(0.009)	(0.008)	(0.007)	(0.009)
House Value (000)	−0.00003**	0.00004	−0.00004**	−0.0001**	−0.0001*
	(0.00002)	(0.0001)	(0.00002)	(0.00003)	(0.0001)
English 2nd Language (%)	−0.054***	−0.061***	−0.057***	−0.062***	−0.016
	(0.009)	(0.014)	(0.012)	(0.012)	(0.017)
Tertiary Degree (%)	−0.048	−0.066	0.047	0.031	−0.330***
	(0.051)	(0.072)	(0.072)	(0.068)	(0.127)
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year × House)	46,552	25,894	14,914	32,458	16,964
Adjusted R^2	0.650	0.666	0.719	0.630	0.731

Note: 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 not applied to columns 1, 3, 4 or 5. *p<0.1; **p<0.05; ***p<0.01

Table B6—Placebo Test - Panel A - Standard

	Dependent variable:				
	Total House Votes	Total Senate Votes	Senate Informal %	Senate Donkey %	Senate Other %
	(1)	(2)	(3)	(4)	(5)
Change in Margin	−16.705	−20.030	0.006	−0.002	−0.013
	(119.104)	(118.748)	(0.007)	(0.006)	(0.021)
Change in ln(Voters)	−61.683	−62.620	−0.005	0.034***	0.075***
	(177.595)	(177.078)	(0.008)	(0.009)	(0.029)
Change in ln(N Options)	17.587	17.198	−0.001	0.0005	−0.003
	(24.027)	(24.002)	(0.002)	(0.001)	(0.004)
Changed Division	6.753	6.928	0.0004	−0.0002	−0.012***
	(10.294)	(10.297)	(0.0004)	(0.0005)	(0.002)
2010	−94.341***	−94.911***	0.013***	0.006***	0.021***
	(6.745)	(6.764)	(0.0003)	(0.0004)	(0.001)
2013	−261.046***	−261.558***	0.007***	0.011***	0.143***
	(15.774)	(15.800)	(0.0005)	(0.001)	(0.002)
2016	−261.962***	−262.106***	0.018***	0.003*	0.182***
	(19.945)	(19.979)	(0.001)	(0.001)	(0.002)
Median Age	9.463***	9.495***	0.0003**	−0.002***	0.005***
	(3.079)	(3.085)	(0.0001)	(0.0003)	(0.0004)
Mean Income (000)	5.085***	5.070***	−0.0001***	−0.0004***	−0.002***
	(0.865)	(0.867)	(0.00003)	(0.0001)	(0.0001)
Unemployment (%)	−9.435	−9.082	−0.0002	0.004*	−0.007
	(24.767)	(24.786)	(0.001)	(0.003)	(0.005)
Population Density	0.025*	0.025*	−0.00000***	−0.00000**	−0.00001***
	(0.014)	(0.014)	(0.00000)	(0.00000)	(0.00000)
Population Growth (%)	5.249*	5.339*	0.0003***	0.001***	0.001***
	(2.795)	(2.804)	(0.0001)	(0.0002)	(0.0003)
Population Decline (%)	−1.918	−1.886	0.00002	0.001***	−0.001***
	(1.817)	(1.816)	(0.00005)	(0.0001)	(0.0002)
House Value (000)	−0.007	−0.007	−0.00000**	−0.00000***	−0.00001***
	(0.007)	(0.007)	(0.00000)	(0.00000)	(0.00000)
English 2nd Language (%)	−8.146***	−8.175***	−0.00002	−0.001**	−0.004***
	(2.693)	(2.695)	(0.0001)	(0.0002)	(0.0003)
Tertiary Degree (%)	4.388	4.697	−0.0004	−0.0004	0.006**
	(17.245)	(17.265)	(0.001)	(0.001)	(0.002)
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	23,276	23,276	23,276	23,276	23,276
Adjusted R^2	0.948	0.948	0.688	0.443	0.782

Note: 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 B7—Placebo Test - Panel B - Propensity Score Matched

	Dependent variable:				
	Total House Votes	Total Senate Votes	Senate Informal %	Senate Donkey %	Senate Other %
	(1)	(2)	(3)	(4)	(5)
Change in Margin	41.315	41.922	−0.001	0.002	0.004
	(150.220)	(149.466)	(0.009)	(0.008)	(0.024)
Change in ln(Voters)	217.067	205.293	−0.006	0.025*	0.111***
	(254.236)	(254.234)	(0.012)	(0.014)	(0.038)
Change in ln(N Options)	0.786	0.331	−0.001	0.002	0.006
	(32.084)	(32.084)	(0.002)	(0.002)	(0.006)
Changed Division	9.329	9.571	0.0003	−0.0004	−0.013***
	(14.186)	(14.179)	(0.001)	(0.001)	(0.002)
2010	−109.193***	−109.452***	0.016***	0.007***	0.026***
	(13.194)	(13.214)	(0.001)	(0.001)	(0.002)
2013	−278.025***	−278.022***	0.009***	0.013***	0.137***
	(26.373)	(26.377)	(0.001)	(0.001)	(0.003)
2016	−277.695***	−277.274***	0.021***	0.011***	0.184***
	(32.928)	(32.946)	(0.001)	(0.002)	(0.004)
Median Age	2.070	2.092	0.00000	−0.002***	0.004***
	(5.136)	(5.140)	(0.0002)	(0.0004)	(0.001)
Mean Income (000)	6.591***	6.552***	−0.0001**	−0.001***	−0.001***
	(1.343)	(1.345)	(0.00005)	(0.0001)	(0.0002)
Unemployment (%)	−10.952	−10.622	0.001	−0.008**	−0.001
	(60.696)	(60.810)	(0.002)	(0.004)	(0.010)
Population Density	0.015	0.015	−0.00000	−0.00000	−0.00001*
	(0.052)	(0.052)	(0.00000)	(0.00000)	(0.00001)
Population Growth (%)	7.544	7.480	0.0002**	0.001***	0.002***
	(5.197)	(5.209)	(0.0001)	(0.0003)	(0.001)
Population Decline (%)	−1.699	−1.635	0.0001	0.001**	−0.001***
	(4.237)	(4.245)	(0.0001)	(0.0002)	(0.0004)
House Value (000)	−0.014	−0.013	0.000	−0.00000	−0.00001***
	(0.022)	(0.022)	(0.00000)	(0.00000)	(0.00000)
English 2nd Language (%)	−10.208**	−10.225**	−0.00004	−0.0005	−0.005***
	(4.583)	(4.588)	(0.0001)	(0.0003)	(0.0005)
Tertiary Degree (%)	−18.675	−17.598	−0.0003	0.003*	0.011***
	(25.782)	(25.851)	(0.001)	(0.001)	(0.004)
Polling Place FE	✓	✓	✓	✓	✓
Observations (Polling Place × Year)	12,969	12,969	12,969	12,969	12,969
Adjusted R^2	0.946	0.946	0.725	0.389	0.804

Note: 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

way as other voters – indicating that this result is not explained by differences in education. A potential explanation for why areas with high levels of education respond to the Change in Margin as predicted by the hypothesis is that those with high education levels may be more engaged in political issues (which may reduce their costs of acquiring information on candidates); may see greater personal benefits from having their preferred candidate elected; and may understand how the electoral system works in greater detail.

Table B8—Quartiles of Percentage with Tertiary Degree - Panel A - Standard

	Dependent variable: Informal %		
	Quartile 1 Lowest %	Quartile 2 and 3	Quartile 4 Highest %
Change in Margin	−0.096	−0.275	6.673***
	(1.309)	(1.482)	(1.636)
Change in ln(Voters)	−3.201	0.757	−0.027
	(2.806)	(1.490)	(2.032)
Change in ln(N Options)	1.236***	1.204***	2.393***
	(0.385)	(0.321)	(0.362)
Changed Division	−0.252*	−0.245***	0.024
	(0.135)	(0.089)	(0.123)
2010	1.783***	1.814***	1.230***
	(0.092)	(0.055)	(0.057)
2013	2.335***	2.570***	2.075***
	(0.181)	(0.117)	(0.126)
2016	1.918***	2.169***	1.344***
	(0.238)	(0.146)	(0.156)
Median Age	−0.083**	0.008	−0.041
	(0.037)	(0.023)	(0.029)
Mean Income (000)	−0.011	−0.031***	−0.018***
	(0.011)	(0.007)	(0.005)
Unemployment (%)	−0.546	−0.383	1.056**
	(0.475)	(0.246)	(0.515)
Population Density	−0.001	0.001***	0.00004
	(0.001)	(0.0003)	(0.0001)
Population Growth (%)	0.018	0.029*	−0.007
	(0.022)	(0.016)	(0.027)
Population Decline (%)	0.038	0.039***	−0.007
	(0.038)	(0.013)	(0.015)
House Value (000)	0.001***	−0.00004	0.00002
	(0.0004)	(0.0002)	(0.00003)
English 2nd Language (%)	−0.062	−0.137***	−0.089***
	(0.044)	(0.023)	(0.020)
Tertiary Degree (%)	−0.106	−0.305**	0.222*
	(0.529)	(0.144)	(0.133)
Polling Place FE	✓	✓	✓
Observations (Polling Place × Year)	5,973	11,667	5,636
Adjusted R^2	0.543	0.689	0.697

Note: 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 B9—QUARTILES OF PERCENTAGE WITH TERTIARY DEGREE - PANEL B - PROPENSITY SCORE MATCHED

	Dependent variable: Informal %		
	Quartile 1 Lowest %	Quartile 2 and 3	Quartile 4 Highest %
Change in Margin	−1.555	0.171	5.318
	(1.655)	(1.680)	(3.237)
Change in ln(Voters)	−4.969	0.180	1.223
	(3.687)	(1.929)	(3.887)
Change in ln(N Options)	1.690***	1.187***	2.352***
	(0.563)	(0.437)	(0.503)
Changed Division	−0.363*	−0.431***	−0.263
	(0.200)	(0.137)	(0.167)
2010	2.066***	2.094***	1.599***
	(0.157)	(0.102)	(0.138)
2013	2.508***	2.843***	2.713***
	(0.289)	(0.206)	(0.291)
2016	2.173***	2.309***	1.617***
	(0.372)	(0.248)	(0.369)
Median Age	−0.100*	0.040	0.024
	(0.053)	(0.038)	(0.067)
Mean Income (000)	−0.015	−0.034***	−0.028*
	(0.017)	(0.013)	(0.015)
Unemployment (%)	−1.978***	−1.252***	1.126
	(0.531)	(0.438)	(0.964)
Population Density	0.001	0.001***	0.0001
	(0.001)	(0.0004)	(0.0004)
Population Growth (%)	0.082*	0.094***	0.021
	(0.048)	(0.024)	(0.041)
Population Decline (%)	0.206***	0.059***	−0.005
	(0.067)	(0.020)	(0.021)
House Value (000)	0.001**	0.0002	0.0001
	(0.001)	(0.0004)	(0.0001)
English 2nd Language (%)	−0.204***	−0.115***	−0.044
	(0.077)	(0.032)	(0.039)
Tertiary Degree (%)	−1.040*	−0.349*	0.099
	(0.556)	(0.208)	(0.229)
Polling Place FE	✓	✓	✓
Observations (Polling Place × Year)	3,399	6,332	3,232
Adjusted R^2	0.582	0.708	0.714

Note: 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 B10 present a version of the main results where two additional variables that could potentially affect informal voting are included. These are the tenure (in years) of the incumbent candidate and the share of votes for progressive parties (defined as first preference votes for the ALP and Green parties). Similarly as for other variables, logs and differences are then applied to the tenure variable. These variables could potentially affect informal voting as, for example, voters who are moved into an electoral division where there is a strong incumbent candidate may not see value in voting while voters that move into an area with a strong (or weak) progressive voter base may also not see much value in casting a valid vote. The parameters for these variables are not found to be statistically significant in determining the rate of informal voting.

Table B11 presents a version of the main results where the margin is defined as the margin measure on first preferences (that is the number one preference on the ballots when submitted). This is important because of the complexity of defining and understanding the margin in an Instant Runoff Voting system – as is used in the House of Representatives in Australia. In this system, the margin is defined based on the share of votes in the final round of voting not on the share of first preference votes. This may make it more difficult for voters to understand and respond to the expected margin when placing their vote (as it requires some calculation of the flow of preferences throughout the runoff process). The results are similar to the main results with the exception that Change in Margin is statistically significant at the 10% level of significance in one specification. This result does not, however, confirm H1 as the first preference margin is not related to the chance of a voter being pivotal and so, in theory, shouldn't affect decisions around the benefits of casting a vote.

TABLE B10—MAIN RESULTS - INCLUDING TENURE AND PROGRESSIVE SHARE

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w cov	FE	DID Std	DID PSM	DID DL
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	0.744**	0.048	−3.530***			
	(0.292)	(0.284)	(0.363)			
ln(Voters)	1.925***	1.429***	0.898**			
	(0.147)	(0.156)	(0.352)			
ln(N Options)	0.789***	0.876***	2.157***			
	(0.051)	(0.054)	(0.057)			
ln(Tenure)	−0.086***	−0.062***	−0.056***			
	(0.021)	(0.021)	(0.021)			
Progressive Vote Share (%)	0.620***	0.229**	1.252***			
	(0.112)	(0.114)	(0.123)			
Change in Margin				0.533	−0.173	0.066
				(0.904)	(1.176)	(0.912)
Change in ln(Voters)				−0.273	−0.165	0.394
				(1.145)	(1.525)	(1.165)
Change in ln(N Options)				1.336***	1.385***	1.315***
				(0.213)	(0.289)	(0.211)
Change in ln(Tenure)				−0.027	−0.0002	−0.050
				(0.068)	(0.094)	(0.068)
Change in Progressive Vote Share				0.280	1.608	0.904
				(0.950)	(1.383)	(0.974)
Changed Division				−0.284***	−0.538***	−0.239**
				(0.105)	(0.153)	(0.105)
2010		1.573***	2.008***	1.701***	2.023***	2.062***
		(0.031)	(0.039)	(0.037)	(0.071)	(0.072)
2013		1.585***	2.097***	2.479***	2.829***	3.150***
		(0.040)	(0.073)	(0.073)	(0.138)	(0.136)
2016		0.759***	1.760***	2.007***	2.251***	2.239***
		(0.051)	(0.099)	(0.094)	(0.174)	(0.167)
Median Age	−0.019***	−0.029***	−0.005	−0.015	−0.004	0.068**
	(0.004)	(0.005)	(0.015)	(0.016)	(0.028)	(0.028)
Mean Income (000)	0.037***	0.022***	−0.015***	−0.027***	−0.033***	−0.048***
	(0.002)	(0.002)	(0.004)	(0.004)	(0.008)	(0.007)
Unemployment (%)	0.135***	0.123***	−0.138	−0.244	−1.031***	−0.448
	(0.012)	(0.012)	(0.187)	(0.209)	(0.329)	(0.445)
Population Density	0.0002***	0.0002***	0.00002	0.0001	0.0004*	0.0005***
	(0.00003)	(0.00003)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Population Growth (%)	−0.043***	−0.025***	0.007	0.018	0.047***	0.034**
	(0.009)	(0.009)	(0.011)	(0.012)	(0.017)	(0.017)
Population Decline (%)	−0.005	0.013**	−0.006	−0.006	0.022	0.015
	(0.006)	(0.006)	(0.008)	(0.008)	(0.015)	(0.013)
House Value (000)	0.00004	0.0001***	−0.00004*	−0.00002	0.0001	−0.00003
	(0.00003)	(0.00003)	(0.00002)	(0.00003)	(0.0001)	(0.00003)
English 2nd Language (%)	0.072***	0.071***	−0.058***	−0.106***	−0.114***	−0.106***
	(0.003)	(0.003)	(0.013)	(0.015)	(0.024)	(0.020)
Tertiary Degree (%)	−0.144***	−0.130***	0.061	−0.059	−0.094	0.056
	(0.003)	(0.003)	(0.074)	(0.082)	(0.128)	(0.105)
Constant	−19.195***	−13.456***				
	(1.624)	(1.767)				
Polling Place FE			✓	✓	✓	✓
Clusters	6,184	6,184	6,184	6,184	5,455	2,187
Treated Observations	NA	NA	NA	1,435	1,177	1,435
Control Observations	NA	NA	NA	21,841	11,760	6,219
Observations	23,476	23,476	23,476	23,263	12,937	7,641
Adjusted R^2	0.371	0.430	0.703	0.663	0.678	0.769

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the Polling Place level. Share of progressive vote is defined as votes for Labor and Greens and the calculation excludes all votes for other minor parties. Change in progressive vote share is the absolute value of the exogenous change in votes for Labor and Green candidates due to a polling place changing electoral divisions. *p<0.1; **p<0.05; ***p<0.01

Table B11—Main Results - Margin Defined on First Preferences

	Dependent variable: Informal %					
	Model specifications					
	OLS	OLS w cov	FE	DID Std	DID PSM	DID DL
	(1)	(2)	(3)	(4)	(5)	(6)
Margin	0.942***	0.617***	0.232			
	(0.157)	(0.124)	(0.163)			
ln(Voters)	3.276***	1.243***	0.592*			
	(0.192)	(0.158)	(0.349)			
ln(N Options)	0.478***	0.890***	2.177***			
	(0.066)	(0.053)	(0.057)			
Change in Margin				0.724*	0.658	0.519
				(0.401)	(0.559)	(0.405)
Change in ln(Voters)				−0.136	−0.417	0.518
				(1.137)	(1.533)	(1.160)
Change in ln(N Options)				1.360***	1.444***	1.365***
				(0.209)	(0.293)	(0.207)
Changed Division				−0.258***	−0.400***	−0.153**
				(0.065)	(0.097)	(0.065)
2010		1.588***	2.086***	1.697***	2.014***	2.042***
		(0.030)	(0.040)	(0.037)	(0.072)	(0.072)
2013		1.573***	2.066***	2.473***	2.823***	3.128***
		(0.041)	(0.074)	(0.073)	(0.139)	(0.137)
2016		0.792***	1.870***	2.003***	2.253***	2.233***
		(0.051)	(0.098)	(0.094)	(0.174)	(0.166)
Median Age		−0.033***	−0.026*	−0.015	−0.004	0.068**
		(0.004)	(0.015)	(0.016)	(0.028)	(0.028)
Mean Income (000)		0.021***	−0.019***	−0.027***	−0.033***	−0.047***
		(0.002)	(0.004)	(0.004)	(0.008)	(0.007)
Unemployment (%)		0.133***	−0.100	−0.243	−1.040***	−0.464
		(0.012)	(0.183)	(0.208)	(0.327)	(0.445)
Population Density		0.0002***	0.0001	0.0001	0.0004*	0.0005***
		(0.00003)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Population Growth (%)		−0.022**	−0.0004	0.018	0.047***	0.033*
		(0.009)	(0.011)	(0.012)	(0.017)	(0.017)
Population Decline (%)		0.012*	−0.004	−0.006	0.022	0.015
		(0.006)	(0.008)	(0.008)	(0.015)	(0.013)
House Value (000)		0.0001***	−0.00003	−0.00001	0.0001	−0.00003
		(0.00003)	(0.00002)	(0.00002)	(0.0001)	(0.00003)
English 2nd Language (%)		0.071***	−0.078***	−0.107***	−0.114***	−0.108***
		(0.003)	(0.013)	(0.015)	(0.024)	(0.020)
Tertiary Degree (%)		−0.128***	0.066	−0.057	−0.098	0.058
		(0.003)	(0.076)	(0.082)	(0.126)	(0.104)
Constant	−33.425***	−11.405***				
	(2.154)	(1.782)				
Polling Place FE			✓	✓	✓	✓
Clusters	6,184	6,184	6,184	6,184	5,455	2,187
Treated Observations	NA	NA	NA	1,440	1,177	1,440
Control Observations	NA	NA	NA	21,841	11,760	6,219
Observations	24,119	23,476	23,476	23,281	12,937	7,659
Adjusted R^2	0.027	0.431	0.698	0.663	0.678	0.769

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the Polling Place level. Share of progressive vote is defined as votes for Labor and Greens and the calculation excludes all votes for other minor parties. Change in progressive vote share is the absolute value of the exogenous change in votes for Labor and Green candidates due to a polling place changing electoral divisions. *p<0.1; **p<0.05; ***p<0.01