

The State of the Literature

Deciding to Vote

Why Turnout Matters

Turnout is the most commonly used measure for participation. It is important because it signifies the level of engagement of the population with the state, the level of incorporation of different subgroups of the population into democratic processes, and the legitimacy of elected officials. Turnout is a metric for how widespread democratic participation is; it is one of the best quantitative measures of the strength of the democratic franchise, alongside qualitative metrics such as voter education and information. Turnout for an election can be calculated or predicted, the difference being that in the former case we use data post-election that is reflective of the actual number of voters, while in the latter we use a series of individual and community covariates to infer the levels of turnout.

To calculate turnout, we simply divide the number of ballots cast by the potential voting population, as in the following equation:

$$\% \text{ Turnout} = \frac{\text{Total Ballots Cast}}{\text{Measure of Total Voting Population}} \times 100\%$$

The choice of numerator is fairly obvious and universal; the denominator, however, is a different story. The three main statistics used are the total voting age population, voting eligible population, and the number of registered voters in a certain geographical location. The total voting age population (all individuals over 18 years of age) can be measured using data from the US Census. However, such an interpretation of voting age population positively counts individuals of age that are not allowed to vote, like people with severe mental illnesses or felons, and does not count overseas voters or military personnel. Michael McDonald offers an alternative to voting age population he calls “voting eligible population”, which corrects for such individuals [mcdonald_what_nodate].

Counts of registered voters are also a useful tool for calculation of turnout, as they usually require no estimation. These counts can simply be extracted from voter registration files. Using registered voters, however, also brings with it two problems. First, voter registration files many times can include discrepancies like deceased voters, voters included in multiple counties, or individual voters included multiple times. Furthermore, the total amount of actual voters among registered voters can be misrepresentative of democratic participation; consider that if a certain minority community has historically low registration rates they are not included at all in calculations of the turnout statistic.

The punch line here is that how the turnout statistic is calculated is not a clear choice, and will have an impact on how studies are set up. To give one example, consider Oregon’s Motor Voter program, that automatically registers voters when they interact with the DMV. It is conceivable that this reform will *decrease* turnout when measured as a percentage of the total registered voter count, but *increase* turnout when measured against total population. This happens if more people register to vote, but do not actually do so—in other words, both number of registrants and number of ballots cast are increasing, but the former increases at a larger rate than the latter. I will specify how I calculate turnout in the next chapter.

Statistical models of turnout can be constructed at either the individual or community level. At the individual level, a model is built to predict the probability of voting for every member of a group, and then sum over the members to create an estimate for turnout. Probit or Logit models are preferred. At the community level, researchers first choose a geographical level at which to calculate, which then constitutes the individual observation in the data that is used to create the model.

Both these models include a standard set of societal variables at the individual and aggregate level, policy variables (whether the district does Postal Voting, whether Voter ID requirements are particularly strict), election-specific variables (closeness of election or campaign expenditure) and sometimes time-series data, like previous levels of turnout. This type of analysis is not exclusively used to predict turnout but also to,

as will be later shown, draw inferences on the effects that certain explanatory variables have on electoral participation.

Through meta-analyses on studies of turnout, it is possible to get a clear picture on what variables effect individual and collective choices to turn out. Three such studies are conducted by Geys (2006), Geys and Cancela (2016), and Smets (2013). Geys includes 83 studies of national US elections in his initial meta-analysis [Geys explaining 2006], later increasing that number to 185 (Geys and Cancela, 2016) and adding local elections. On aggregate-level models for national elections they conclude that competitiveness, campaign financing, and registration policy have the most pronounced effects, while on the sub-national level there are more pronounced effects for societal variables and characteristics of election administration (spending, voting policy, etc.). Smets and Van Ham (2013) examine individual-level predictors for turnout in a similar meta-analysis, and conclude that “age and age squared, education, residential mobility, region, media exposure, mobilization (partisan and nonpartisan), vote in previous election, party identification, political interest, and political knowledge” [Smets embarrassment 2013] are the most significant explanatory variables for turnout, along with income and race. I will specify the model I will use for turnout in the second chapter.

Theories of Voting

Here I take one step back from turnout, and examine the theories surrounding individual choices to vote or abstain. There are three main theories outlined in the literature on why individuals chose to vote. While there is some overlap, the following are mostly distinct:

- *Decision “at the margins”*: In his 1993 study, Aldrich posits that voting is a low cost-low benefit behavior. Therefore, he continues, voting is a decision that individuals make “at the margins”; in most people, the urge to vote is not overwhelmingly strong, and therefore individuals will vote when it is convenient to them, when they are motivated by a competitive race, when policies are put in place to help them, and when they are subjected to GOTV (Get Out the Vote) efforts. For Aldrich, this is corroborated by the fact that most turnout models present consistent, yet weak, relational variables; if decisions are made “at the margins”, then no single predictor would have an overwhelming result. This is also supported by Matsusaka (1997), and Burden & Neiheisel (2012). Matsusaka expresses support for a more “random” process of voting, where turnout models are ambiguous because of the difficulty that predicting “at the margins” entails [Matsusaka voter 1999]. Burden & Neiheisel (2013) also demonstrate support for Aldrich’s thesis by using data from Wisconsin to calculate a net *negative* effect of 2% on turnout following the expansion of early voting access in the state. [Aldrich rational 1993; Neiheisel impact 2012]
- *Habitual Voting*: While Aldrich supports that there is no single overwhelming predictor of turnout, Fowler (2006) posits that future voting behavior can be strongly predicted using individual voting history. This leads to the conclusion that individuals are set to either be habitual voters, or habitual non-voters [Plutzer becoming 2002] by their upbringing and social circumstances, locking them into distinct groups. [Fowler habitual 2006]
- *Social/Structural Voting*: Close to habitual voting are those that support a model of social and structural voting; these researchers claim that the decision to vote or not is deeply rooted in socioeconomic factors, which means that the divide between traditionally voting and non-voting groups can only be bridged by directly dealing with the socioeconomic divide between them [Edlin voting 2007; Berinsky perverse 2005]. Their reasoning is that “at the margins” voting only addresses groups that do not face significant burdens against voting (like the working poor, or marginalized racial groups) and are usually already registered. Similarly, they address habitual voting claims by arguing that they are too short-sighted; individuals themselves might be habitually voting, but their decision to do so is rooted in strong societal and policy factors.
- *Resources and Organization*: To some extent growing from structural theories of voting, resources and organizations theory emphasizes the interaction of personal political and societal characteristics of voters, and actions taken by politicians to mobilize participation. This theory is very broad in the inputs it assesses for voter participation, ranging from practical issues of access and resources (how easy it is for

someone to vote if they so choose), to public policy feedback effects and signaling (how the government’s policies affect the people and how they react), to how political parties and groups choose to mobilize and approach voters [rosenstone_mobilization_2003]. Apart from Rosenstone and Hansen’s work (2013), there have been several studies examining voter participation based on resources and organizations theory, a lot of which come from the public policy side of political science. Some examples are Chen’s study of how distributive benefits like federal emergency aid affect participation among recipients, after controlling for partisan characteristics (2012), or Mettler and Stonecash’s examination of correlation between welfare program participation and political mobilization (2008), or Campbell’s analysis of social security recipients and their voting patterns (2002). The punchline in all these studies is that public policy is correlated with trends in participation, either because recipients of benefits wish to protect such programs, or because of the interaction between partisanship and government support, or because of access related to resources and voting laws [mettler_government_2008; chen_voter_2013; campbell_self-interest_2002].

From Theory to Policy

Voting Methods

I have already flagged in my introduction the reason why theories behind voting choice matter: each construct an image of the electorate that reacts differently to policy change around voting. They are all an answer to the fact that elections policy, and how we conduct elections, is not value neutral but has implications for turnout, which in turn has implications on the franchise of democracy.

In trying to respond to the issues set up by theoretical paradigms, different states—both in the global and US contexts—have adapted to different ways of conducting elections. In the US, voting styles can be simplified into three categories:

- *In-Person Election Day*, for which all individuals are required to vote at a polling place, on a single election day. There can be some leeway for overseas voters, or excused absentee voters, but the vast majority of people will have to be present to vote in a particular time frame.
- *In-Person Early Voting*, for which all individuals must vote in person at a polling place or vote center, but the timeframe for voting extends for around two weeks, not a single day.
- *Vote-By-Mail, Absentee Early Voting*, for which individuals have a clear, no-excuse-necessary option for not being present when they vote, or for filling in a mailed ballot and dropping it off at designated locations.

For the purposes of this thesis I will examine the latter category, and specifically Vote-By-Mail. The reason behind this is that the model of in-person, election day voting is usually seen as the baseline, the “vanilla” way of conducting elections if you will. Therefore it has been of interest for researchers to examine if other systems can outperform that baseline. Specifically, it is most interesting to examine voting styles that are heralded for their expansion of turnout, to see whether popular beliefs on their benefits and drawbacks hold; if they are different from the base model of conducting American elections, or if they present new challenges and unique selling points. Vote-By-Mail is particularly interesting because it is quickly taking the form of a trend in state elections, as more and more states are enforcing more open models of VBM. In the next section, I will more closely examine the particulars of Vote-By-Mail.

What is VBM?

Vote-By-Mail is a process by which voters receive a ballot delivered by mail to their homes. Voters then have a variety of options on how to return these ballots, ranging from dropping them off at pre-designated locations, to mailing them in, to bringing them to a polling place. The two first options are most commonly implemented, with a very small number of states still operating polling places for mail ballots. This varies across states that have implemented VBM. Some common forms of the VBM policy are:

- *Postal Voting*: All voters receive a ballot by mail, which can then be returned to a pre-designated location or mailed in to be counted. All-mail elections currently occur in Oregon, Washington, and Colorado.
- *No-Excuse Absentee*: Voters can choose to register as absentee voters without giving any reason related to disability, health, distance to polling place etc. This is the case in 27 states and the District of Columbia.
- *Permanent No-Excuse Absentee*: This is similar to the previous system, but allows voters to register as absentees indefinitely, without having to renew their registration each year; they become de facto all-mail voters. This is in place in a very large number of the no-excuse absentee voting states like Utah, California, Montana, Arizona, New Jersey, and others.
- *Hybrid or Transitional Systems*: In hybrid systems, voters receive a mail ballot but can choose to disregard it and vote conventionally. This is the case in Colorado. Transitional systems exist in states that have chosen to eventually conduct all elections by postal voting, but have given counties an adjustment period during which this shift is not mandatory, or mandatory only for certain elections. This is the case in California, Utah, and Montana.

Vote-By-Mail is also commonly considered a type of early voting, since voters receive their ballots around two weeks in advance of election day; they are also able to return that ballot whenever they wish within that time-frame. This means that Vote-By-Mail can be counted as a “convenience voting” reform [gronke_convenience_2008]. These are usually implemented by state and local governments with the argument that they either expand the democratic franchise by bringing in new voters, or by making it more likely that current registered voters participate in the electoral process [national_council_of_state_legislatures_absentee_nodate].

How Theories Apply to VBM

Under Aldrich’s paradigm, vote by mail would not effect significant change in voting behavior. The whole concept of a decision “at the margins” is that the forces at play when an individual decides to vote are overwhelmingly strong both ways, so any effect that policy can have will minimally shift these margins. If, for example, we take a presidential election the forces at play include the media, national committees, social effects etc. In this environment, some added convenience does not significantly add to an individual’s decision to turn out. However, this would indicate that at a local level, where national and media effects are less strong, the effect of VBM on turnout might be more significant. The effect would be present for all groups, not only those currently registered, since voting would be easier uniformly.

If we assume habitual voting, the conclusion on VBM would differ significantly. In this case, the effect to be considered is how VBM impacts already formed habits around voting. It could be argued that VBM has no effect, which follows if we assume that voting habits formed do not shift if the mode of voting changes. It could also be argued that VBM might have a negative effect on turnout in the short term, because it disrupts the habit of election day for a readjustment period, before people settle into new groups of habitual voters and non-voters, adapted to the new policy context.

Under social and structural voting contexts, VBM retains rather than stimulates new voters [berinsky_perverse_2005]. This means that already registered and semi-active voters are more likely to participate, but there is no significant change in the amount of new voters entering the franchise. This would mean that traditional forms of voting policy that emphasize access to the polls will do nothing to bring in disenfranchised people, and potentially hide the problem under an inflated turnout statistic calculated on registered voters. Berinsky in particular emphasizes the need for a shift towards voter education, rather than early voting or VBM policies [berinsky_perverse_2005].

Vote-by-Mail is obviously not a welfare or spending program, but it does expand individual resources in terms of voting capacity. A ballot delivered to your home means that less resources need to be expended in the act of voting, which in turn has both a practical effect—building capacity—and a more behavioral effect—a feeling of inclusion, an interaction with the process of voting that comes through a ballot at your doorstep

that would not exist if you had not gone to a polling place [schneider_behavioral_1990]. Under a resources and organizations framework, both these effects are most likely to be net positive to political participation, and as such would predict a strong, positive effect of VBM on turnout.

General Results on VBM

I will start with studies that show a negative effect on turnout. Bergman (2011) uses a series of logit models of individual voting probability in California, during a period where part of the state conducted VBM elections, while others maintained traditional voting. This is called a “quasi-experiment”, and it is common throughout the literature. Bergman’s results show a statistically significant drop in voting probability in VBM counties [bergman_changing_2011]. Using a similar method, Keele (2018) takes a single city in Colorado, Basalt City, which is divided into two different voting districts using different voting systems. The conclusion is, again, a 2-4% drop in turnout along the VBM part of the city [keele_geographic_2017]. Burden et al. (2014) takes a different approach, using country-wide election data from 2004 and 2008 presidential elections, and compares districts based on early voting practices. Their results show a significant drop in turnout, which can be associated to VBM as well due to its closeness to EV [burden_election_2014].

In contrast, Gerber et al. (2013), applying both individual and county-level models for the state of Washington, reach the conclusion that VBM increases turnout by around 2-4%; they use the same quasi-experimental model that offers itself to researchers in states that are under transitional systems [gerber_identifying_2013]. R.M. Stein also reaches a similar conclusion when examining Colorado’s practice of “vote centers”, which are non-precinct attached polling places, which can service multiple counties [stein_engaging_2008]. I include this paper here due to the link that voting centers have with VBM, as they serve as drop-off points for mail-in ballots. Richey (2008) examines the effects that Oregon’s VBM program has on turnout by using past elections data, concluding a 10% positive trend associated with the policy [richey_sean_voting_2008]. This effect is studied again by Gronke et al. (2012) who find a similar positive effect with much lower magnitude, which might point to a novelty effect: the existence of diminishing returns in turnout after the implementation of this policy [gronke_voting_2012]. Gronke et al. (2017), again studying Oregon but focusing on Oregon’s Motor Voter program, find evidence of positive association to turnout. I include these effects due to Oregon being an exclusively VBM state, and because this paper uses a “synthetic control group” model, a particularly interesting statistical technique. Lastly, I include a study conducted by Pantheon Analytics on Colorado, which compares actual turnout to predicted levels for VBM counties in Colorado. The results show a positive effect of approximately 3.3% due to VBM [edelman_analysis_2018].

The conclusion to be drawn from this section is that results on VBM vary significantly. There are multiple studies, using multiple methods, on multiple states, with multiple results. This only adds to the importance of being careful when constructing models and hypotheses to test VBM’s effects on turnout, as assumptions made in the process can critically impact the results.

Voter Registration Files as Data Sources

Before concluding this chapter, I want to briefly discuss some background research into the use of voter registration files for the purpose of elections research. This may seem like an abrupt shift from the previous section but, as I mentioned in my introduction, access to such files is what motivated and facilitated a lot of the aforementioned studies in the first place. I will not directly go into the structure of such files; such a section will be included later on in this thesis.

Inaccuracy of Survey Data

Apart from Voter Registration Files, the main source of data on the American electorate is national surveys, like the American National Election Studies’ survey (ANES), or the Cooperative Congressional Elections Study (CCES). These are post-election surveys, distributed to voters, which include fields associated directly with voting—participation, precinct, which party you voted for—and indirectly, through questions

on individual characteristics like race, income, or gender. On the surface these seem like a better source of data, since no record linkage or ecological inference need be made to connect individual voters with an extensive list of covariates. There is, however, a significant problem with these data: survey misreporting [Burden_voter_2000].

A cursory glance at the CCES and ANES estimates of turnout reveals the existence of a problem right off the bat: turnout calculated through surveys is usually higher than reported figures. When looking at surveys a bit closer, using either private, extensive data files like Catalyst [Ansolabehere_validation: 2012] or validated voter files from the late 20th century [Deufel_race_2010], the results show consistent misreporting among certain groups that tend to either be politically engaged non-voters or minorities and low socioeconomic status individuals. This gap, according to Deufel et al. (2010), has served to propagate societal stereotypes and class entrenchment into studies on turnout, which in turn negatively effect policy, since research using the ANES and CCES are widely used to study turnout among the groups that are consistently misreporting. Admittedly, the fact that misreporting happens among specific groups does open the way for statistical methods to compensate for the bias introduced, but for the purpose of my thesis I will prefer the use of Voter Registration Files.

The Importance of VRF

As mentioned in my introduction, access to voter registration files has provided researchers with unique insight into the voting process. Quantitative research has expanded significantly, for three key reasons. First, VRF data exists in a consolidated, state-wide format at least for national elections. This means that the process of data collection involves interaction with significantly fewer government agencies, and a data wrangling process that can be quickly adapted to a set format. This is, of course, not to say that the process of data collection and handling doesn't still pose a significant challenge, as will become apparent in my second chapter. Second, there is a huge benefit attached to the fact that VRF data describes the whole population, rather than a sample. As mentioned in the previous section, survey data might give more insight into variables not included in VRF, but that comes at a steep cost for accuracy. Using VRF, the problem of self-reporting bias is eliminated for some studies, and transformed into a problem of record linkage and ecological inference for others [Ansolabehere_adgn: 2017, Burden_new_1998]. Third, wide public access means reproducible and accessibility, which translates into greater accountability for researchers. This effect is important, even if mitigated somewhat by private data companies and access fees.