A Population Model of Voter Registration and Deadwood

Stephen Pettigrew
Harvard University

Charles Stewart III
MIT

Draft of April 3, 2016

Prepared for presentation at the Annual Meeting of the Midwest Political Science Association, April 7, 2016. A previous version of the paper was presented at the Conference on New Research in Election Administration and Reform, MIT, June 8, 2015. The authors thank the Pew Charitable Trusts for financial support in writing this paper. Pew bears no responsibility for the contents of this paper.

A Population Model of Voter Registration and Deadwood

Stephen Pettigrew
Charles Stewart III

The accuracy of voter registration lists has been at the center of debates over improving election administration in the United States for over a decade. Inaccurate voter registration lists pose a problem for both those concerned about the convenient access of voters to the polls and for those concerned about voter fraud. Concern about list accuracy has led academics and advocates alike to scrutinize voter registration rolls to ferret out "deadwood" (Ansolabehere and Hersh 2013). (Deadwood in this case is defined as obsolete records, usually due to a person moving or dying. See Shaw, Ansolabehere, and Stewart 2015, p. 30.) When voter rolls include more names than eligible voters, a local jurisdiction is an easy target for those concerned about list accuracy.[1]

Despite the great concern with identifying voter registration lists with excessive deadwood, and the negative optics associated with voter registration lists that contain more names than eligible voters, there has been little systematic attention given to the core questions, "how big *should* voter registration lists be?", and, "when is a voter registration list "too big"?" Related questions follow from the core questions: If there are too many names on a voter roll, why? How much is due to the operation of federal and state laws that limit how often list maintenance might be done? How much is due to the inability of local officials to ascertain when a registered voter has moved or died, either due to inadequate administrative practices or inattention to list maintenance altogether?

---

[1] See Judicial Watch, "Judicial Watch Warns Iowa, Colorado, DC of Potential Election Integrity Lawsuits," http://www.judicialwatch.org/press-room/press-releases/judicial-watch-warns-iowa-colorado-dc-of-potential-election-integrity-lawsuits/, March 24, 2014.

The goal of this paper is to answer the core question starting from first principles, particularly the basic population patterns that are the drivers of registration. The number of people on a state's voter registration rolls should derive from a limited set of basic underlying population dynamics. New voters come onto the rolls through three processes:[2] (1) the registration of young people who age into the electorate, (2) the registration of older people who move into a state and register, and (3) the registration of older residents who had been previously unregistered at the previous unit of time.[3] The lion's share of voters leave the rolls through two main processes: (1) the death of registered voters and (2) the out-migration of registered voters. Smaller numbers may leave the registration rolls because of felon and other disfranchisement laws, as yet others may leave because they lose interest in voting, topics we leave aside for future work.[4]

Using the insights that come from this simplified understanding of what causes "churning" within voter registration rolls, this paper develops a model of voter registration list size that draws from population ecology models, and thus is termed "a population model of registration and deadwood." Although the core of this paper is the specification of a core formal model of voter registration list size and its calibration, there are major empirical issues that must be attended to on the way to assessing the utility of this model.

[2] Here and throughout the paper, we ignore the obvious overlaps among these populations, such as young people who move into a state at 18 and then register soon thereafter.

[3] Throughout this paper we consider the natural unit of time in which to model voter registration list dynamics to be the four-year presidential election cycle. This is because of the clear pattern of registrations that sees new registrations peak during campaign periods, especially in the days before presidential elections (or at least the book closings associated with presidential elections in those states without Election Day Registration (EDR).) Modeling a shorter period of time, especially a two-year window, may be useful in some contexts, but we consider shorter time periods to require attention to second-order considerations.

[4] In the gray area are voters who fall between the cracks because they move within a state. For now, we treat the "reregistration" of these voters as a change of status that does not affect the number of people on the rolls. However, practically speaking, intrastate movers might affect voter roll churning in two ways, first, by creating the possibility that a voter might be on a state's voter rolls more than once, despite efforts to prevent duplicates, and second, by creating a hurdle to future voting, which might then cause a voter to fall off the rolls because of the failure to vote.

The first empirical issue provides further motivation to this effort, by comparing different estimates of the number of registered voters on the rolls across the states. We know that states vary considerably in the fraction of their eligible populations who vote. Therefore, we should expect that states would also vary considerably in the fraction of their eligible populations who are registered. However, although we can establish state turnout rates fairly precisely through the use of official vote reports and Census estimates of eligible populations,[5] we know of no similarly precise way to estimate registration rates. Official registration statistics clearly overstate the "actual" registration rates of eligible citizens. One sign of how inaccurate the official registration statistics are is the fact that survey-based estimates of state voter registration rates are often *lower* than the rates calculated using official records. If we assume that the answer to the survey research question about whether a respondent is registered to vote is prone to the same social desirability bias that besets questions about whether a respondent voted, then survey-based estimates of the voter registration rate must also be over-estimates of the actual rates. That survey-based estimates of registration rates are often *lower* than official estimates should give one pause.

The second empirical issue relates to the quality of "workflow" statistics reported by the states about their list maintenance activities. If states accurately reported the critical quantities identified in the theory that follows, then the administrative efforts could be easily assessed by examining how many voters had been removed because of death or because they had moved out-of-state. As we show below, the reality is far from expectations. First, barely half the states provide an accounting of why they remove voters from the rolls that is adequate for assessing

[5] We acknowledge uncertainties in estimating the size of each state's eligible populations and note that not all states even report the number of people who vote in their elections. However, as we show below, the measurement uncertainty created due to these factors pale in comparison to the measurement uncertainty created by the difficulties in establishing just how many eligibles are registered to vote.

whether an appropriate number of voters are removed during any given period of time. Many states either do not record (or report) why voters are removed from the rolls, or report quantities that are wildly inconsistent with underlying population dynamics.

This has two implications, one methodological and the other substantive. Methodologically, it is impossible to use official statistics to reliably assess the quality of many states' list maintenance regimens. Substantively, to the degree that the truth of these list maintenance programs is revealed (however imprecisely) by official statistics, it appears that too few voters are being removed from the voter lists, which probably has set into place a systematic ballooning of voter registration lists in many states.

This paper proceeds as follows. In Section I we illustrate the variability of the deadwood problem through a comparison of official voter registration statistics with survey-based estimates of registration rates. In Section II we describe our conceptual model of voter registration list composition and establish certain important definitions. In Section III we describe our population model in mathematical terms and illustrate its dynamics through a series of stylized simulations. In Section IV we apply the theoretical model to a set of states in order to test the accuracy of its predictions. In Section V we discuss the results of the simulation runs of the models. We conclude in Section VI.

I. Comparing Official Registration Rates with Survey-Based Estimates

We illustrate the problem we seek to address with Figure 1, which plots two measures of voter registration rates in 2012 at the state level against each other: the "official" registration rate reported by the states for the EAC's biennial NVRA report, and the registration rate calculated using responses to the Voting and Registration Study of the Current Population Survey (VRS).[6] This graph contains two surprises, one obvious, and the other more subtle.

[Figure 1 about here]

The obvious surprise is that Alaska and the District of Columbia report registration rates greater than 100%. Another three states (Iowa, Maine, and Michigan) report rates that approach 100%.[7] The presence of registration rates greater than 100% is old news to election administration aficionados, but is fodder for those who suspect that registration rolls are bloated and prone to fraud. Furthermore, the fact that some states have registration rates above 100% raises the possibility to some that the states with official registration rates below 100% are bloated, although to a lesser degree than the "100%+ Club."

The subtle surprise is that 29 states report official registration rates that are greater than the rates calculated using responses to the VRS.[8] (These are the states above the diagonal line in Figure 1.) This is a surprise because the registration rate calculated using the VRS is almost certainly an over-estimate, given that we know social desirability bias systematically inflates estimates of voting rates. The VRS voter registration estimate is based on answers to two questions. The first is whether the respondent voted in the most recent federal election. The second question is asked of respondents who reported they did not vote, and it asks whether the respondent is registered. The first question produced a 26 percentage point over-estimate of

[6] Attention has recently been drawn to the fact that the Census Bureau has long employed an unorthodox approach to calculating the registration rate in VRS reports (Hur and Achen 2013). The percentages reported here are calculated using more conventional methods, that is, dividing the number who report themselves as registered by the voting eligible respondents to the two questions in the VRS that determine registration status.

[7] North Dakota is included on the graph with a 100% official registration rate because it has no voter registration, and thus in a sense all its eligible voters are registered. In any event, it is instructive to see how many North Dakota voters say they are registered in that state.

[8] Because of the large number of observations for each state, the 95% confidence interval for the registration rates calculated for each state using the VRS is in the 2-3 percentage point range. As a consequence, virtually all the data tokens in Figure 1 that do not touch the diagonal line are outside this confidence interval.

turnout in the 2012 election.[9] While there has not been research into the reliability of self-reports of voter *registration*, there are good reasons to suspect that the second question is also prone to social desirability bias, and thus produces an over-estimate of voter registration among those who report they did not vote.

If the registration rate calculated using responses to the VRS are likely over-estimates, then it is even more likely that most — and perhaps all — state voter registration rates calculated using official registration records are even more inflated. The question is, by how much? A related question is whether some states' voter registration rates are more inflated than others.

Ansolabehere and Hersh (2013) provide one answer to these questions, by reporting the extent of "deadwood" on voter registration rolls as of 2012, using the Catalist deadwood model. Overall, approximately 4% of voter registration records were either "likely" or "probably" deadwood in 2012, but the cross-state variability was significant. The deadwood rate was less than 1% in twenty of the fifty-one states (including D.C.). At the other extreme, over 12% of the records of Colorado and Arkansas were either likely or probably deadwood. When we replicated Ansolabehere and Hersh's analysis using the 2013 Catalist 1% sample, we found a substantially higher nationwide percentage of "probably" or "very likely" deadwood (19.7%), ranging from 35.0% in Colorado down to 10.0% in Vermont.

The variability in the Catalist deadwood rate across time suggests that the quality of the registration lists is subject to considerable temporal flux. It also underscores the need to estimate baseline state registration rates that are divorced as much as possible from the vagaries of administrative practices in the states.

[9] Using responses to the 2012 VRS, we calculate a turnout rate of 80.0%. Michael McDonald's reported nationwide turnout, using the "VAP highest office" metric is 53.6%. The difference is 26.4 percentage points

II. A Conceptual Model of Voter Registration Statistics

One way to assess the degree to which administrative voter registration records contain too many names is to consider how many people would be on a list in a perfect world in which changes to one's voter status — particularly those that take someone off the list of eligible voters — were instantaneously recorded on the voter rolls. With such a baseline model, it would then be possible to layer on the administrative requirements of the state and federal laws that regulate list maintenance activities, along with standard operating procedures associated with those activities.

For instance, the NVRA gives states latitude in adopting procedures for removing voters because of their death, and HAVA requires states to coordinate their list maintenance activities with other state databases, such as vital statistics. One common procedure is to match the voter registration list with the state's death registry on an annual basis, and to do the maintenance after a federal election. If list maintenance to remove dead voters occurs annually, then the average voter who has died remains on the voter rolls for half a year.[10] If the mortality rate of registered voters is 1%, then we would expect official registration statistics to overstate the number of registered and eligible voters in an election by approximately half a percentage point due to administrative practices related to death removals. Using a similar logic, it would be possible to add to this the over-statement of voter registration rolls due to natural lags in accounting for voters moving from a state.[11]

---

[10] Only about 3% of Americans die outside their state of residence, and therefore do not appear on their state's death registry. Therefore, we ignore this complicating detail.

[11] We account for this in more detail below. For the moment note that the annual interstate mobility rate in the U.S. is approximately 2%. Considering how these voters would be removed from the "sending" state's voter rolls is complicated. Most likely, the voter would be removed when she or he registered in the "receiving" state. If voters who move are like voters generally, the lion's share will register in the new state when an election is coming up, most probably the presidential election. Thus, it is not unreasonable to start with the assumption that the move of the average registered voter who moves out-of-state is not reported to the original state for an average of two years (i.e., half of the four-year presidential election cycle). Thus, the voter registration rolls of the average state following standard practices would be inflated by 4 percentage points due to out-of-state moving.

To measure the degree to which the official registration rates overstate the actual registration rate, we need a baseline model of what the registration rolls "should" look like. It is the purpose of this section to specify such a model in conceptual terms. We do so primarily by defining terms that are often left vague in the literature and specifying a set of accounting identities that should hold in a dynamic formal model of voter registration lists.

Some definitions

We define the *eligible electorate* as the age-eligible citizens with residency in a state who also meet the state's other voting requirements. For now, we put aside most of the fine points of this definition, such as what it takes to establish residency, whether a state has felon disfranchisement, etc.

The *registered eligible electorate* in a state is a subset of the eligible electorate, consisting of the eligible electorate that is on a state's voter registration rolls.[12] The complement of this set is the *unregistered eligible electorate,* which is the subset of the eligible electorate not on that state's voter registration rolls. By definition, the following accounting identity holds for state *s*:

$$\textit{eligible electorate}_s = \textit{registered eligible electorate}_s + \textit{unregistered eligible electorate}_s$$

Next, we define the set of *registered voters.* Registered voters are individuals who are on the voter rolls of a state. We can divide the set of registered voters into three subsets: (1) the registered eligible electorate (as defined above), (2) registered ineligible voters, and (3) deadwood. Viewed this way, we see that we can think of the registered eligible electorate as individuals who are properly on the voting rolls and could cast a ballot without violating any election laws in the state. *Registered ineligible voters* are living residents who are on the rolls

[12] As above, for now we put aside the classification of voters into active and inactive, as is done in some states.

but who do not meet eligibility requirements, such as a non-citizen who may have registered or those who have not met residency requirements. Finally, *deadwood* consists of people who were previously registered eligible voters, but are no longer, either because they have died or because they have moved.[13]

With these definitions related to the voter registration rolls, the following accounting identity holds for state *s*:

$$\textit{registered voters}_s = \textit{registered eligible voters}_s + \textit{registered ineligible voters}_s + \textit{deadwood}_s$$

Because the number of non-citizens and felons who are on voter rolls are likely to be very small,[14] for the rest of this paper we assume that the number of registered ineligible voters is zero, and thus,

$$\textit{registered voters}_s = \textit{registered eligible voters}_s + \textit{deadwood}_s$$

Figure 2 illustrates these two accounting identities in terms of a series of sets. The major sets are (1) the eligible electorate and (2) registered voters. The intersection of these sets is the registered eligible electorate. The relative complement of the eligible electorate in the set of registered voters is partitioned into two subsets: (1) deadwood and (2) the registered ineligible electorate. Following the discussion above, we treat the subset of the registered ineligible electorate as empty for the rest of this paper.

[Figure 2 about here]

---

[13] We recognize that there is ambiguity about what changes an intra-state mover from the registered eligible voter category to the deadwood category. In some states, a voter who moves within a state remains on the rolls and could vote in the next election, either in the current or past precinct, without re-registering. In other states, intra-state movement triggers a re-registration requirement. Further complicating this is the fact that some states will allow intra-state movers to vote at their prior residence within a certain time window, after which time re-registration is necessary. There is no ambiguity that interstate movers who remain on a state's voter rolls are deadwood.

[14] There is a substantial literature on the voting propensity of ex-felons (see Meredith and Morse 2015 and citations therein), but that is a different issue from the number of felons who remain on the rolls after conviction, in those states where they are disfranchised. Similarly, there has been some scholarly attention to the question of non-citizen voting that, despite the controversial nature of the findings (c.f. Richman, Chattha, and Earnest 2014 and Ansolabehere, Luks, and Schaffner 2015), confirms that the likely number of non-citizens on the registration rolls nationwide is likely at barely detectable levels.

Referring to Figure 2, it is useful to define two more quantities that will come up repeatedly in this paper, the *true registration rate* and the *official registration rate.* We define the true registration rate for a state as

$$\frac{Registered\ eligible\ electorate}{Eligible\ electorate}$$

and the official registration rate as

$$\frac{Registered\ voters}{Eligible\ electorate}.$$

From Figure 2 is it clear that the official registration rate will be greater than the true registration rate.

Registration dynamics

The discussion thus far has been in terms of a voter registration file at a single point in time. However, the topic we are interested in arises because of the dynamic nature of the eligible electorate and its interaction with the list of registered voters. Deadwood occurs because some of the registered eligible electorate move or die during each period. This movement depletes the size of the *eligible* electorate over time without (necessarily) depleting the size of the *registered* electorate. Of course, the eligible electorate is also renewed each period because of the natural increase of the resident population and movement into the state of eligible adults.

Thus, we complicate the conceptual model sketched out in Figure 2 by adding a temporal dynamic. But, before doing that, we must introduce two new sets of eligible voters, the *newly eligible electorate* and the *new immigrant electorate.* We define the newly eligible electorate as resident citizens who turn 18 and thus become eligible to vote. We define the *new immigrant electorate* as otherwise eligible citizens who have moved into a state from a different state. Both the newly eligible electorate and the new immigrant electorate constitute a pool of individuals who replenish the set of the eligible electorate that is diminished due to death and out-migration.

The dynamics of the system from time t to time $t+1$ are illustrated in Figure 3.

[Figure 3 about here]

The arrows describe the transitions that individuals can make from one period to the next. In the middle of the figure, the two curved arrows indicate that both registered and unregistered eligibles can remain registered or unregistered, respectively, across the two periods. The arrows pointing between *R* and *U* indicate that there is flow from one period to the next among eligible voters who move between the registered and unregistered states. In general, we expect that most of the traffic in the system will move from the unregistered eligible electorate to the registered eligible electorate. However, there will be some movement from the registered to unregistered state, primarily due to mobility.

The upper part of the diagram depicts the sorting of new members of the eligible electorate into the registered and unregistered states. The lower part of the diagram depicts the exit of eligibles, either through death or out-migration.[15]

III. The Mathematical Model

In this section, we move from conceptual to mathematical models.

We draw our inspiration from two literatures. The first is research by Ansolabehere, Hersh, and Shepsle (2012) (hereafter AHS), who modeled the rate of registration among age cohorts in a vein similar to ours. The second is age-structure population growth models which

[15] To clarify one detail of Figure 3, despite the fact that the entire lower half if labeled "deadwood," only the flows out of registered eligibles have the possibility of contributing to deadwood. The flow from the unregistered eligible electorate is included for the sake of completeness.

generally fall under the category of Leslie-Lefkovitch models within the field of population ecology (Leslie 1945; Lefkovich 1965; Tuljapurkar and Caswell 1997).

There are two major points of similarity between this paper and that of AHS. The first is that we also seek to derive basic voter registration statistics, such as registration rates and deadwood rates, from underlying population dynamics. Second, we also conceive of the problem in terms of an iterated two-period process, in which the distribution of individuals in various registration states in time t are estimated from the distribution in time t-1, plus a series of transition probabilities that describe movement between states.[16]

However, our model departs from AHS in one important way, owing to the different aims of our analyses. AHS assume a fixed population. Theirs is a model in which an eligible voter is either a "mover" or "stayer" between periods, but the number of eligible voters never changes. In contrast, our attention to deadwood requires us to consider a fluid population: the flow of registered voters out of the state, either through death or out-migration, creates the possibility of deadwood if they stay on the rolls. With a flow out, we also have to consider how the eligible voter population is replenished, and the rate at which the eligible voter population grows because of migration and new entrants due to (lagged) births.

Our own model consists of six equations which predict the number of individuals in each of the states described in Figure 3. We describe each equation in turn.

Registered eligible voters (R)

The first equation pertains to the number of registered eligible voters at time t:

$$R_t = a_{rr}R_{t-1} + a_{ru}U_{t-1} + a_{rb}B_{t-1} + a_{ri}I_{t-1} \quad (1)$$

[16] There is one important empirical finding from AHS that we rely on as well: the core finding from AHS is that it a useful approximation of registration rates of different age cohorts can be estimated using a simple model in which the probability of becoming registered at time t+1, conditional on not being registered at time t, is constant for all eligible citizens. We return to this in the next section.

(For the moment, we suppress subscripts for individual states, although the application we have in mind requires us generally to use different coefficients for different states.) In words, the number of registered eligible voters in one period is equal to a fraction of registered eligible voters who remain registered from the previous period ("registration persistence") plus the fraction of formerly unregistered eligible voters who have become newly registered *and* the fraction of the new eligible electorate (either due to natural increase or immigration) who register immediately upon entering the eligible electorate. Following AHS, it is likely best to assume that $\alpha_{ru} = \alpha_{rb}$, and even perhaps $\alpha_{ru} = \alpha_{rb} = \alpha_{ri}$. It also goes without saying that all of the weights in Equation (1) are less than or equal to 1.

In introducing the coefficient α_{ru}, we find it useful to introduce two additional definitions. The coefficient α_{ru} can properly be referred to as a type of registration rate, that is, the registration rate of the unregistered, or alternatively, the registration rate conditional on being unregistered. This is in contrast to what is typically called the registration rate in popular discourse and in academic research. What is typically called the registration rate is simply the number of people on the registration rolls divided by the number eligible. However, it is clear that we need to make reference not only to the fraction of eligible voters who are already registered, we need to have a clear way of referring to the rate at which people who are unregistered go onto the registration rolls. For that reason, we find it useful to distinguish between the *conditional registration rate*, that is, the registration rate conditional on being unregistered (i.e., α_{ru}) and the *registration rate*, that is, the proportion of the entire eligible population that is registered.

Unregistered eligible voters (U)

The expression for the number of unregistered eligible voters is similar:

$$U_t = a_{ur}R_{t-1} + a_{uu}U_{t-1} + a_{ub}B_{t-1} + a_{ui}I_{t-1} \quad (2)$$

The coefficient a_{ur} describes the fraction of registered eligible voters in one period who cease to be registered (but still eligible) in the next. The other three coefficients are the complements of the corresponding coefficients in Equation (1). That is, a_{ub} = 1-a_{rb} and a_{ui} = 1-a_{ri}. Thus, all newly eligible 18-year-olds and new eligible voters who have moved into the state are allocated either to R_t or to U_t.

Dead formerly eligible voters (D)

The number of newly dead former eligible voters is defined as

$$D_t = a_{dr}R_{t-1} + a_{du}U_{t-1} \quad (3)$$

The two coefficients, a_{dr} and a_{du}, can be thought of as the mortality rates of registered and unregistered eligible voters, respectively. Below, we will adopt the simplifying assumption that $a_{dr} = a_{du}$, in other words, that registered eligibles die at the same rate as unregistered eligibles.

It should be noted that of the two parts of the right-hand side of Equation (3), only $a_{dr}R_{t-1}$ contributes to deadwood. In addition, the contribution of deaths to deadwood depends on the period over which registered voters who have died linger on the voter registration rolls after their deaths. As discussed in the introduction to Section II, for instance, if the mortality rate is 1% and deceased registrants are removed as part of an annual maintenance cycle, then we would expect that the amount of the list's deadwood that was attributable to deaths would approximate half a percentage point, even for a "well-scrubbed" list.

Moved formerly eligible voters (M)

Finally, the number of recently moved, formerly eligible voters is defined as

$$M_t = a_{mr}R_{t-1} + a_{mu}U_{t-1} \quad (4)$$

As with deaths, a_{mr} and a_{mu} can be interpreted naturally as the outbound migration rates of registered and unregistered voters. It is reasonable to assume that $a_{mr} = a_{mu}$, that is, that registered and unregistered voters move out of a state at equal rates. In addition, the contribution that these movers make to deadwood consists only of $a_{mr}R_{t-1}$.

As with deaths, the contribution of the $a_{mr}R_{t-1}$ portion of movers to deadwood depends on the period of time that former residents of a state linger on voter rolls after they move. As noted above, the out-migration rate of voters in most states usually exceeds the mortality rate by a factor of two, at around 2% in any given year. Because the process of capturing voters who move out of a state is less comprehensive than the process of accounting for deaths, it is likely that the bulk of out-of-state movers are removed from voter registration lists because of the failure to vote in two consecutive federal elections, not because election officials have been notified by the "receiving" state that the voter has re-registered elsewhere.[17] In a typical state with an out-migration rate of 2%, if these registrants were removed *only* because of non-voting, the voter registration list would be nearly 8% greater than what would obtain if out-of-state movers were removed instantaneously.

Replenishing the population

Because our model requires us to replenish the pool of eligible voters, we need to account for the number of new 18-year-olds and new eligible in-migrating residents explicitly. We do so using simple growth models, as follows:

$$B_t = (1 + a_{bb})B_{t-1}$$

[17] Failure to vote is the most common reason given for removing a voter from the voter rolls, at 30.1% of all removals in the 2012 NVRA report. In contrast, only 27.1% of removals were due to "moved from jurisdiction." (Equally telling, the percentage of registrations removed due to death, 21.6%, is barely less than the percentage removed due to moving, despite the fact that twice as many voters move out-of-state as die.) Furthermore, if we exclude the five states that generally fail to classify reasons for removal (Alabama, D.C., Florida, Indiana and Utah) the correlation between the fraction of removals due to non-voting and the fraction of removals due to moving is -.73 (after weighting for the number of registered voters), which suggests that one is a substitute for the other.

and

$$I_t = (1 + a_{ii})I_{t-1}$$

The coefficient a_{bb} can be easily interpreted as the rate of population growth due to births, or at least the birth rate net of the number of deaths and out-migration of minors. The coefficient a_{ii} can similarly be interpreted as the in-migration rate of a state.

Matrix expression of the model

In analyzing this system formally, it is useful to express it in matrix terms, as follows:

$$\begin{pmatrix} R \\ U \\ D \\ M \\ B \\ I \end{pmatrix}_t = \begin{bmatrix} a_{rr} & a_{ru} & 0 & 0 & a_{rb} & a_{ri} \\ a_{ur} & a_{uu} & 0 & 0 & a_{ub} & a_{ui} \\ a_{dr} & a_{du} & 0 & 0 & 0 & 0 \\ a_{mr} & a_{mu} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 + a_{bb} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 + a_{ii} \end{bmatrix} \begin{pmatrix} R \\ U \\ D \\ M \\ B \\ I \end{pmatrix}_{t-1}$$

or

$$\boldsymbol{C}_t = \boldsymbol{A}\boldsymbol{C}_{t-1}$$

If we expressed the starting parameters at time 0 as $\boldsymbol{C}_0$, the statistics of the system at time *t* are expressed as

$$\boldsymbol{C}_t = \boldsymbol{A}^{\boldsymbol{t}}\boldsymbol{C}_0$$

It is possible to seed $\mathbf{C_0}$ with starting values for the various quantities and then to run the simulations a large number of times and then examine the values remaining after many dozens (or hundreds) of iterations. However, note that all the coefficients except those accounting for population growth are generally less than 1, whereas population growth is assumed to be positive. Thus, the first thing to notice is that eventually the values within **C** that are of the most interest will be dominated by the ratio of the starting values of *B* and *I* (in other words, the ratio of births to in-migration at the start of the simulation) and the associated growth rates. The second thing to notice is that in general, although the values estimated in **C** are integers, we are really interested in various ratios, such as the ratio of registrants to the number of eligibles (i.e. $R/[R+U]$). As a general matter these ratios will quickly approach asymptotic values, and therefore relevant solutions can be derived after just a small number (10–20) of iterations of the model.

To illustrate the basic dynamics of the model, we begin with a simple simulation, with initial parameters given in Table 1.[18] In Figure 4 we show the simulated values that are related to the partition of the voting-age population into those who are registered and unregistered, with the raw values in Figure 4a and the registration rate in Figure 4b. Note that the registration rate converges very quickly to an asymptotic value of 76.0%.

[Figure 4 about here]

By construction, the death and moving rates are set in the simulation. In this case, the birth rate was set to 0.7% and the moving rate was set to 1.2%. Therefore, the potential deadwood rate will be 1.9% in each round.

The degree to which deaths and out-of-state moves result in deadwood accumulating on the voter rolls depends on the frequency and efficiency with which list maintenance procedures are conducted. If everyone who dies and moves out-of-state is removed from the rolls immediately, then there is no deadwood. If there are delays, or if only a fraction of deaths and out-of-state moves are removed, then deadwood will grow; indeed, it can become an increasingly large part of the voter rolls.

[18] The transition parameters (weights) in Table 1 are chosen because they approximate nationwide population averages. The starting population values were chosen to allow for the number of in-migrants to be slightly greater than the number of births, which is generally true in states. The starting population value places all voting eligible residents in the unregistered pool. These choices of initial population values do not have any effect on the steady-state dynamics of the model that we explore below.

This is illustrated in Figure 5, where we show the results after we have altered this simulation to examine two different list maintenance regimes. In Regime 1, deaths and out-of-state moves remain on the voting list forever. In Regime 2, deaths remain for one period while out-of-state moves remain for two.[19]

[Figure 5 about here]

In Regime 1, with no list maintenance, the voter list eventually becomes dominated by deadwood. However, deadwood on the voter rolls grows gradually, and even after 100 iterations of the model, deadwood is only 55.5% of the official registration list. Recall that population growth due to births and migration occurs at a faster rate than deaths and out-migration. Thus, the size of the "true" registered population almost keeps pace with the size of the deadwood component of the voter registration rolls. The operative word is "almost," of course, and thus deadwood does continue to grow as a fraction of the official registration list over time.

The fact that deadwood grows gradually even under this extreme model of no list maintenance may have implications for whether election administrators or others can informally judge whether a state's voter rolls contain an overabundance of deadwood. In the real world, there is considerable fluctuation in the number of new voters who come onto the rolls from year-to-year or election cycle-to-election cycle. It is quite likely that these dramatic shifts in the size of the rolls due to new voters registering in anticipation of a hotly contested election could mask the degree to which the rolls are growing more steadily because of a failure to completely account for deadwood.

Figure 5 also illustrates the results of the model when deadwood is removed after a period of latency. In this case, where deaths are allowed to stay on the rolls for one period and

[19] Having the two sources of deadwood persist at different rates is intended to illustrate a case in which deaths can be removed through an annual matching with the state voter list, whereas out-of-state movers may not be removed until they have failed to vote in two election cycles.

out-of-state moves for two periods, the deadwood percentage stabilizes at around 2.9%. This percentage is slightly below what we would have calculated if we just relied on the mortality and mobility rates alone. That is, the simulation assumes a mortality rate of 0.7% and a mobility rate of 1.2%. Without taking the full range of dynamics into the model, we would be tempted to assume that under Regime 2, the deadwood rate would approximate 3.1% (i.e., 0.7% + 2 × 1.2%). However, in the simulation, deadwood is produced by deaths and moves that are lagged one and two periods, respectively, while the number of eligible registrants is growing as the voting-age population grows. However, as this simulation shows, basing expectations about the amount of deadwood using a simple calculation like this makes for a reasonable rule-of-thumb.

IV. Application of the Model

Having now explored a simple version of the simulation, we implement the simulation using data from each state. To implement the model for each state, we must choose the transition probabilities in $\boldsymbol{A}$ and the starting values in $\mathbf{C}_0$. As discussed previously, most of the transition probabilities can be thought of either in terms of vital statistics (i.e., birth and death rates) or census estimates of population movements. The transition probabilities that are the most problematic are the registration persistence rate (a_{rr}) and the registration rate of unregistered voters (a_{ru}). Thus, we begin this section with a discussion of the easier parameters —the transition probabilities that are derived from vital statistics and census estimates — before discussing the more problematic measures.

Birth rate (a_{bb})

The birth rate is used to increment the number of 18-year-olds who enter the pool of eligible voters in each iteration of the model. Strictly speaking, the birth rate is only an approximation of the growth in the cohort of 18-year-olds in a state, since the growth in this cohort only reflects births 18 years prior, and is affected by population dynamics across these 18 years (through deaths and migration patterns of minors). However, as a starting point, the birth rate would seem to be a serviceable approximation of how many new voters are "ageing-into" the eligible population.

The contribution that newly eligible voters will have on the registration rate, and thus on the deadwood rate, will depend on three things: (1) the registration rate of unregistered voters (see below), (2) the birth rate, and (3) the fraction of voters in the youngest cohort. It has frequently been noted in the literature that the youngest eligible voters have the lowest registration percentages of any age cohort (Strate *et al* 1989; Timpone 1998; Highton 2000; Ansolabehere, Hersh, and Shepsle 2012). Thus, holding everything else constant, states with a larger fraction of its eligible population in the youngest eligible cohort should have lower registration rates. That said, the fraction of the eligible population that is 18-years-old in any state is relatively small. Based on counts from the 2010 Census, only about 2% of the population is 18-years-old, and thus the depressive effect of having a relatively young eligible voter population on the registration rate of a state may be difficult to detect statistically.

In-migration growth rate (a_{ii})

The in-migration growth rate, a_{ii}, is used to account for the growth in a state's voter-eligible population due to in-state migration. It should be distinguished from the conditional rate of registration among in-state migrants (a_{ri}), which is used to increase the number of registered voters on the rolls.[20] For this moment, we set a_{ii} equal to the rate of in-migration from other states.[21]

Death rate (a_{dr}, a_{du})

The death rate is used to calculate the number of registered and unregistered eligible voters who have died during the time period reflected in each iteration of the model. The death rate was calculated for each state using data from the Wide-ranging Online Data for Epidemiologic Research (WONDER) system of the Centers for Disease Control and Prevention (CDC).[22] The CDC WONDER system allows for the fine-tuning of mortality rates by age. Therefore, the mortality rates we used for each state described deaths per 100 residents 18 years-old and older. In calculating mortality, we used the most recent data, from 2013.

It is unclear whether death rates, like birth rates, should have an influence on conditional registration rates. If registration rates were strictly monotonic with age, then the greater mortality of older voters would mean that death would have the effect of removing disproportionately registered voters from the rolls. However, the registration rate is a curvilinear function of age. Around age 70, registration rates tend to decline, to the point that the oldest age cohorts have registration rates that approach those of the youngest cohorts. (See Figure 6.)

[Figure 6 about here]

While the mortality rate of the adult population monotonically increases as a function of age, compared to the rapid acceleration of the mortality rate after age 75, the increase in the mortality rate among the bulk of voters is relatively flat. (See Figure 7.) On the other hand,

---

[20] And similarly, it should be distinguished between a_{ui} = 1-a_{ri}, which is the fraction of new in-state residents who go unregistered.

[21] The data source for interstate mobility is the American Community Survey, 1-year estimates, for the relevant years.

[22] http://wonder.cdc.gov/

because of increasing rates of mortality beginning around age 75, the number of the oldest voters begins to decline rapidly. It is likely that these complex dynamics cancel each other out, leaving no relationship between the death rate and registration rates.

[Figure 7 about here]

Out-migration rate (a_{mr}, a_{mu})

The out-migration rates are used to decrement the counts of registered and unregistered eligible due to the flow of eligible voters out-of-state. The source of out-migration rates used in this paper is the American Community Survey's estimates of interstate residential mobility.

Registration persistence rate (a_{rr})

The purpose of the registration persistence rate (a_{rr}) is to account for registered voters who remain in the state between time *t* and time *t*+1 and who also remain registered. Estimating this parameter is tricky, and we have not settled on a robust method to determine it. For now, we assume that a registered voter who does not move remains registered from one period to the next.[23] This leaves only those who move intrastate who are available to transition into the pool of unregistered eligible voters.[24] (Recall that registered voters who move interstate are already accounted for with the a_{mr} parameter.) If we denote the fraction of registered voters who move between times *t* and *t*+1 as *m*, then the registration persistence rate a_{rr} is bounded between 1-*m* (i.e., no interstate movers re-register between elections) and 1 (i.e., all interstate movers re-register between elections). Using the one-year residential mobility data from the 2013 ACS, we

[23] The main consequence of this assumption is to set aside registered voters who stop voting — because they have lost interest, have become infirm, etc. — while retaining an unchanged residence.

[24] In considering this point, we have stumbled upon a fine point of considering the composition of deadwood that seems to have been unconsidered in most discussions of the quality of voter registration lists — the status of intrastate movers. In most states, if a registered voter moves from one jurisdiction to another, she or he must re-register. However, because intrastate moves are no more likely to be caught by election administrators than interstate moves, such individuals are likely to remain on the rolls, even though they are simultaneously unregistered but still eligible to register.

see that intrastate mobility ranged from 7.6% in New Jersey to 16.2% in Nevada, which would set the lower bound of a_{rr} between 0.838 (Nevada) and 0.924 (New Jersey) for a one-year period. If we use these numbers to estimate registration persistence across two- and four-year periods, then the lower bound of a_{rr} is much lower — between .493 in Nevada and .730 in New Jersey.[25] While these fractions seem intuitively much too low,[26] for the present we use the intrastate residential mobility rate to calculate a_{rr}, and leave fine tuning this parameter to future drafts of the paper. In particular, for each state, we let a_{rr} = 1-*m*, which means we let a_{ru} = *m*.

Conditional registration rate (a_{ru})

The conditional registration rate, a_{ru}, is the probability that an eligible voter who is unregistered at time *t* becomes registered at time *t*+1. In our model, the conditional registration rate is primarily used to move a fraction of unregistered voters into the registered voter pool. In addition, for this initial round of model development, we set two other parameters, a_{rb} (the fraction of 18-year-olds who register) and a_{ri} (the fraction of in-migrants who register) equal to a_{ru}.[27]

[25] This assumes that intrastate movers move only once during the period and that moving in one period is independent of moving in another.
[26] Among other things, it appears that state and local officials receive more change-of-address forms during an election cycle than new registrations. For instance, officials reported that of the 62.5 million registration forms received in the 2012 federal election cycle, 27.5 million (44.0%) were changes of address, compared to 23.8 million (38.1%) that were new registrations. (EAC 2012 NVRA Report, Table 3.) As with many of the statistics reported in the EAC's NVRA report, there are many anomalies in these statistics, so we must use them cautiously. Still, statistics such as these suggest that the vast majority of registered voters who move intrastate do in fact reregister.
[27] Following AHS, we suspect that setting a_{rb} equal to a_{ru} is justifiable, given the finding in their paper that *r* (i.e., their version of the conditional registration rate) can be safely treated as constant for all age groups. However, it seems less reasonable to assume that in-state movers register at this (relatively) low rate, especially if we believe that intrastate movers re-register at a greater rate than those who have never registered before. Not all in-state movers are registered to vote in the state they are moving from. Therefore, it is likely that the registration rate of in-state movers will be less than the registration persistence rate of intrastate movers, but still greater than a_{ru}.

In this initial analysis, we adopt the method used by AHS to estimate what we are calling the conditional registration rate.[28] We make two adaptations. First, we apply their method to data observed at the state level, so that we can produce conditional registration rates that vary across states.[29] Second, we expand the range of ages included in each cohort, from two years to four years. We do this because when we disaggregate the analysis to the state level, the estimate of the number of people registered to vote in each state by age group seemed especially prone to sample-induced variability.

Following AHS, we estimate the proportion registered by age group in each state using the Catalist 1% sample drawn in June 2014. So that we could construct each age cohort around a full presidential election cycle, we "rolled back" the sample, so that we assigned people to age cohorts according to their age in November 2012.

Implementing the AHS method to the states separately allows us to estimate conditional registration rates for each state. The conditional registration rates thus estimated range from 7.2% in Wyoming to 86.1% in Ohio. We report the rates for all states in Table 2.

[Table 2 about here]

In Figure 8, we display two scatterplots that describe the relationship between the registration proportion calculated two different ways and the conditional registration rate (a_{ru}) calculated using the AHS method. In Figure 8a, we show the registration proportion using the VRS dataset plotted against state estimates of the conditional registration rate using the Catalist

[28] AHS estimate the conditional registration rate by performing a nonlinear regression in which the unit of analysis is an age cohort *c*. They observe the proportion registered by age cohort, R_c, and estimate the proportion registered by fitting the data to the following function: $R_c = \alpha(1 - \beta^c)$, , where $\alpha = \frac{r}{1-q}$, $\beta = q$, and $q = s(1 - r)$. Using the coefficients, it is possible to back out an estimate of *r*, which is the proportion of interest to us here. (It is also possible to back out an estimate of *s*, which is their "stayer" parameter.)

[29] We recognized that the AHS model assumes that the state population is fixed, whereas our model is developed under the premise that the population is fluid. And alternative is simply to set the conditional registration rate equal to the proportion of voters in the youngest age cohort who are registered in each state.

data. There are four states in this scatterplot — Mississippi, Wisconsin, Alaska, and New Hampshire — and D.C. that have very high registration proportions but low estimated conditional registration rates. Including these four states, the correlation between the two measures is .31; excluding these four states, the correlation is .74.

[Figure 8 about here]

In Figure 8b, we measure the registration proportion using the overall registration percentage reported to the EAC in the 2012 NVRA report. As before, there is a strong correlation between the official registration proportion and the conditional registration rate, with the exception of the outliers, which in this case are Alaska, Wisconsin, New Hampshire, and D.C. Including all the states, the correlation between these two measures in Figure 8b is .31; excluding the three outliers, the correlation rises to .69.

In examining these scatterplots, one thing is reassuring and another thing gives us pause. What is reassuring is that there are fairly strong correlations between the conditional voter registration rates we get by using the Catalist data and the overall registration proportions we get when we use both official registration statistics and responses to the VRS. What gives us pause is that the conditional registration rate is greater than the registration proportion (according to official state statistics) in four states (Florida, Nevada, Oregon, and Washington). This suggests we must do more investigation of how to adapt the AHS model to state data in future iterations of this paper.

V. Results

In this section, we review the results we obtain when we run the simulations for each state. The simulations were run this way: First, a series of starting values was chosen for each state. The sources of these values are reported in Table 3. It should be noted that although the quantities estimated in our model are levels — size of the registered eligible electorate, etc. — the quantities of theoretical interest to us — estimated registration rate, etc. — are rates. Our experience running the simulations is that these rates converge rather quickly (see the discussion in the previous section for a hypothetical example), though they converge faster when seeded with values of *R* and *U* derived from actual 2012 statistics. Second, we calculated for each state the values of the estimated quantities after 100 iterations (i.e., $m=100$, or $\mathbf{C}_m$). Most importantly for our theoretical interest, we have estimates for *R*, *U*, *D*, and *M* that can be used to estimate equilibrium registration rates and deadwood rates.

[Table 3 about here]

We start by reporting estimated true registration rates for each state, that is, $R/(R+U)$. Estimated values are reported in Table 4. In Figure 9 we have plotted official registration statistics against the values in Table 4.

[Table 4 about here]

[Figure 9 about here]

Even casual perusal of the results of these simulations indicates that more work needs to be done to fine-tune the model. The most important anomaly is that in many states, the officially-reported registration rates are less than the simulated true registration rates, when the opposite should be true. The scatterplot in Figure 9 shows the relationship between the true registration rates and those calculated based on official reports in the 2012 NVRA summary. With the exception of seven states that are clear outliers (Wyoming, Hawaii, etc.), there does appear to be a correlation between estimated registration rates and officially reported registration rates. This is a start, but the fact that even the best-performing states in the simulations yield estimated true registration rates that are too high show that further work needs to be done to develop this model — a topic we return to in the final section.

Estimates of the deadwood rate

Our model also lets us examine the nature of the deadwood estimates that our model produces. To do so, we need first to specify more precisely the dynamics of the stock and flow of deadwood. We have defined deadwood as the presence of formerly eligible voters who have died or moved out-of-state on a state's voter registration rolls. Estimating the *number* on a state's voter rolls that is deadwood is a classic stock-and-flow problem. The stock of deadwood is the number of dead or moved voters who are still on the rolls. The inflow consists of new deaths and moves; outflow consists of voters who are removed due to list-maintenance operations.

In principle, list-maintenance operations that pertain to outflows are different for deaths and moves, though in practice, they may not be so distinct. Under the NVRA and HAVA, all states are required to engage in regular list maintenance — programs that are required to be uniform, nondiscriminatory, and in compliance with the Voting Rights Act. If a state receives direct notification that a voter has died or has moved out-of-state, the state may immediately remove the voter from the rolls. If the state has received no such notice, then the state may only remove a voter through a procedure that involves confirmatory mailings and documentation that the voter has not been in contact with election officials for two full federal election cycles.

It is clearly easier to remove dead voters from the rolls in a timely fashion than it is to remove out-of-state movers. Most state departments of elections receive regular reports from their state's vital records bureau about death certificates, and most also utilize the Social Security Death Index (SSDI). Thus, it is reasonable to assume that a state can remove the great majority of dead voters from the rolls in their regular maintenance activities, especially those who die within the state. For those who die outside the state, if the death is not picked up on the SSDI, then it is more likely that the voter will only be removed after not voting for a period of time.

It is harder to remove out-of-state movers from the rolls in a timely fashion. For instance, despite the fact that the rate of interstate moving is several times greater than the nationwide death rate, the number of removals nationwide due to "moved from jurisdiction" during the 2011–2012 federal election cycle was only slightly greater than the number of removals due to death. As a consequence, most movers will be removed from the voter rolls through the process that essentially relies on the registrant not voting for a period of four years. The non-voting period could even be longer, even if a state if aggressive in seeking to remove such voters, depending on the status of confirmation notices that are mailed to registrants.

These conjectures are confirmed through an examination of list maintenance statistics provided by the states to the EAC as part of NVRA-mandated biennial reporting. Death statistics taken from the CDC WONDER system reveal that over 9.7 million people aged 20 and older died in the United States between 2009 and 2012.[30] This is 4.3% of the average population during this period that was 20 years of age and older.[31] Death rates varied by a factor of two across the extremes, ranging from 2.9% in Alaska to 6.0% in West Virginia.

In contrast, the raw voter registration statistics reported to the EAC from 2009 to 2012 revealed that at least 5.9 million voters were removed from the rolls due to death.[32] We state that "at least" 5.9 million were removed because the EAC counts missing data as a zero, rather than

[30] The CDC WONDER system only allows one to query death rates in age categories that are five years in width for the adult population, which is why these statistics are expressed in terms of those 20 and older, rather than 18 and older.
[31] Across this four-year period, the average number of people in the U.S. 20 years of age and older, according to the CDC WONDER database, was 227.4 million.
[32] These statistics are taken from the EAC's NVRA reports for both the 2008–2010 cycle and the 2010–2012 cycle.

imputing non-zero missing values. With an average voter registration level during this period of 190.8 million, it would appear that at least 3.1% of registered voters were removed due to death between the presidential elections of 2008 and 2012.[33] This is at least the same order-of-magnitude as the corresponding death rate in the adult population. However, the interstate variation in the percentage of registrations removed because of death is a factor of 277, ranging from 19.3% of registrants in D.C. to 0.07% in Florida.

Figure 10 shows two sets of scatterplots describing the relationship between voter registration death removals and adult death rates in the states. The first column of figures shows the relationship between death-removals and adult deaths for all states. The second column shows the relationship only for those states that had full reports from all counties in both the 2010 and 2012 NVRA report. Under each column, the first graph shows the raw number of death removals and deaths reported by each state, while the second graph expresses these numbers as a four-year average rate.[34]

[Figure 10 about here]

The graphs in Figure 10 show a positive correlation between deaths and death removals at the state level. If we compare the two columns of graphs, we see that the correlations are greater if we confine ourselves to states with complete data reporting.

If we exclude the District of Columbia, the Pearson correlation coefficient between the death removal rate and the adult death rate is .47.[35] (This analysis is done only using data from

[33] We state "at least" 3.1% for two reasons. First, as already mentioned, we know that there are jurisdictions that probably removed some voters due to death but did not report the number to the EAC. Second, if the voter rolls are inflated, then the rate of removals due to death will under-state the actual fraction of voters removed due to death.
[34] North Dakota is excluded in all graphs because it does not have voter registration. D.C. is excluded from the "rate" graphs to preserve the scale of the graphs.
[35] If we include D.C., the correlation coefficient is .08. If we weight the states by the number of registered voters, the Pearson correlation coefficient is .40. after excluding D.C. and .30 including D.C.

the states with complete data.) Three states (Florida, Utah, and Wyoming) are significant outliers, suggesting that they either undertook virtually no removals from the voter rolls during this period or (more likely) had counties that simply failed to record the reasons they removed voters from the rolls.[36] Wisconsin and the District of Columbia (not shown on the rate graph because it would throw the scale off) are significant positive outliers. We know for certain that D.C. undertook a major audit of their voter rolls during this period, including a removal of dead registrants after years of failing to do so. We suspect that Wisconsin did something similar, though we have yet to confirm this.

In contrast, there is no correlation between the number of voter registration records removed because a voter moved out of state and the out-of-state mobility rate of a state's adults. This is illustrated in Figure 11, which plots the percentage of voters removed due to moving out of state from 2009 to 2012 against the percentage of the population that moved out of state during the same period. Not only is there no correlation between these two data series, there are several states that reported more out-of-state removals than the fraction of adults who moved from the state during the same period.

[Figure 11 about here]

How list maintenance rules impact the deadwood rate

In order to estimate the stock of deadwood in a state, we need to know something about list maintenance operations in each state. At the present, we do not have good measures of how regularly and thoroughly each state engages in list maintenance activities. Lacking that, we varied the amount of time dead and moved registrants were allowed to stay in the set of

[36] In future drafts of this paper, we will investigate this issue further.

deadwood voters before being purged in our model. These amounts of time ranged from zero ("instant purging") to 48 months. Average deadwood rates are reported in Table 5.

[Table 5 about here]

Table 5 reports the average deadwood rates across all states, after making different assumptions about how long voters remain on the rolls after dying or moving. The "instant purge" assumption is reflected in the very first cell — if registrants are removed instantly when they die or move, there is no deadwood. Table 5 also shows that if dead and moved registrants stay on the rolls for equal amounts of time, mobility contributes significantly more to deadwood than dying. For instance, if dead voters are removed immediately but movers are allowed to stay on the rolls for 48 months, deadwood will be 10.2% of the registration rolls; if the opposite is true, then only 4.4% of the registration rolls will be deadwood.

A more reasonable baseline assumption is for the average dead voters to be removed after 6 months, and for the average mover to be removed after 48 months. This assumption is consistent with all dead voters being removed annually, when the rolls are matched against death certificates, but having to wait at least four years for non-voting.[37]

Under this assumption, Table 5 tells us that the average degree of deadwood on a voter registration list with an "average" list maintenance program should be 11.2%. However, there will be some variance around this average, owing to the fact that states have different death- and out-migration rates. This variance is shown in Figure 12, which plots the estimated deadwood rate of each state.

A comparison of Figures 9 and 12 reveals that the extremes in the estimated deadwood rates, both at the upper and lower ends, tend to be the outliers when the registration rates are

[37] In other words, we assume that when the voter rolls are compared against the dead certificate list, the average dead voter will have been dead for six months.

estimated. Thus, at this point, it is impossible to know whether the range of deadwood variability we estimate here would remain after further refinements to the model.

VI. Discussion and Conclusion

The purpose of this paper is to propose a method to estimate the "true" proportion of eligible voters who are registered in each state, building that estimate from the ground up from known (or knowable) population dynamics. As the results here demonstrate, further work is needed before this model can be used to pass judgment on the actual voter registration statistics as reported by the states.

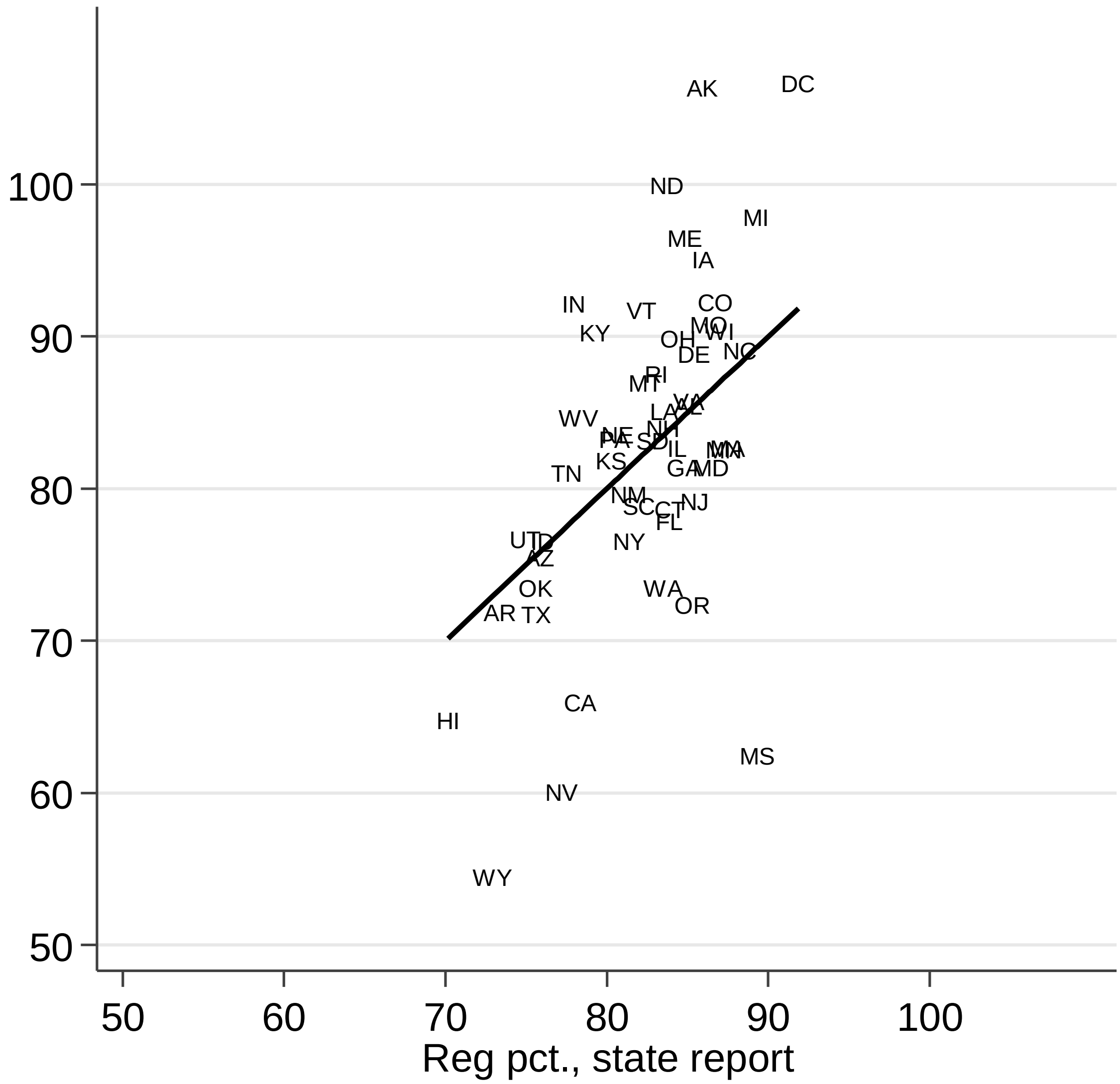
At this point, we have confidence in the overall structure of the approach, and suspect that most attention to fine tuning needs to be focused on the transition probabilities. That said, at least one major issue remains to be explored in developing the overall approach — whether to treat the conditional registration rate as fixed across time. The simulations here proceed under an assumption that transition probabilities are fixed. However, we have strong reasons to believe that the transition probabilities vary across time. Most obviously, for instance, the implementation of the Voting Rights Act over the past half-century has increased dramatically the conditional registration rate in southern states. Registration rates in other states have fallen. To the degree that the stock of currently registered voters were added to the registration rolls at different times, with different conditional registration rates, our estimates of the "true" number of registered and unregistered voters will be off.

As to fine-tuning the transition probabilities, the results shown in Figure 7 clearly suggest that the estimated true registration rates for the bulk of states are well below state-reported rates, and that there is something with the transition probabilities that bear further investigation. The two transition probabilities that do the most work in determining the registration rate are the conditional registration rate and the registration persistence rate. In this paper, we estimated the conditional registration rate using the AHS method without modification, other than moving the estimation to the state level and expanding the width of the age cohorts. It is possible that with a dynamic population, the AHS method produces estimated conditional registration rates that are too high, thus biasing upward the estimate of the true registration rate.

Alternatively the registration persistence rate may be systematically too high. Refining the method to estimate this rate will occupy future work on this project.

We conclude by noting that although our initial efforts to produce estimates of true registration rates using basic population dynamics have generated estimates that are clearly off the mark, we are confident that an approach such as this one is critical, if we are to reach a better understanding of the quality of voter registration lists in the United States. Reaching a better understanding is more than just a matter of better accounting. A non-trivial fraction of Americans believe that the voting rolls are bloated, and that this bloat gives rise to opportunities for voter fraud. Until we develop a scientific method to quantify how big voter registration rolls "should be" under ideal administrative procedures, the question of registration list quality will remain mired in a world of suspicion and recriminations.

Figure 1. Proportion of eligible population that is registered, measured by the Voting and Registration Supplement of the CPS (y-axis) and official state statistics (x-axis), 2012.



Sources: Current Population Survey, Voting and Registration Supplement, November 2012; U.S. Election Assistance Commission, NVRA Report, 2012.

Figure 2. The relationship between the eligible electorate and registered voters.

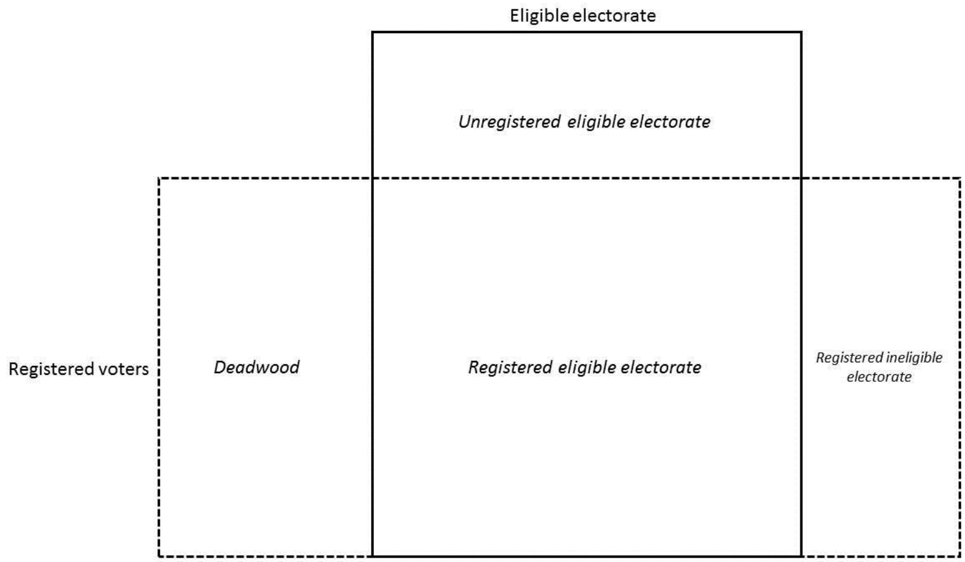
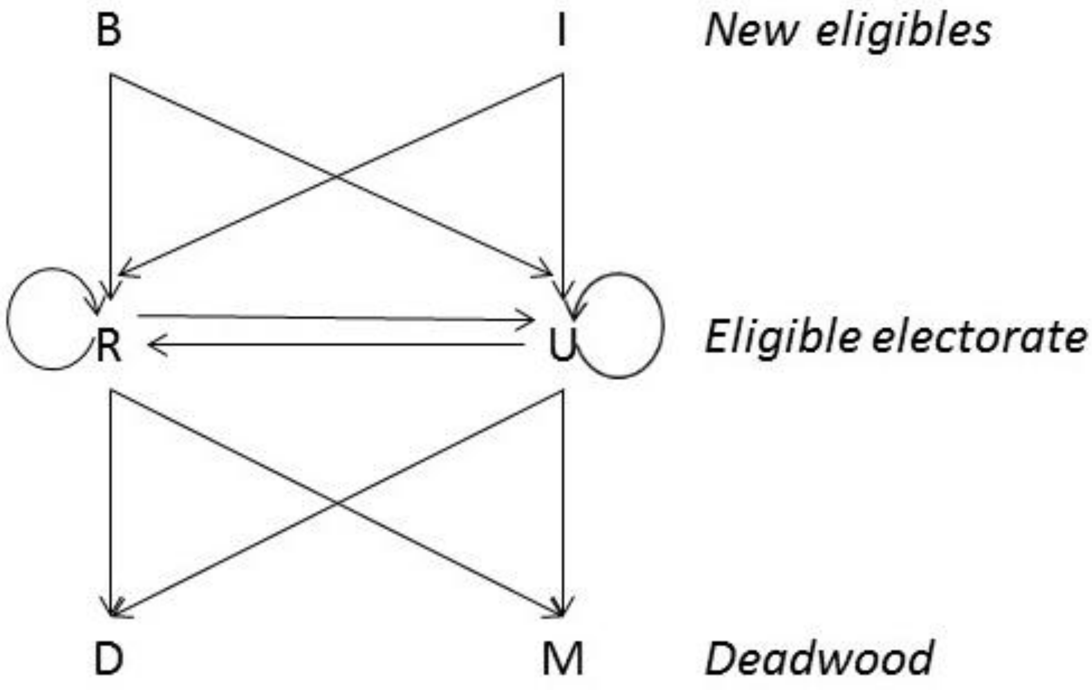


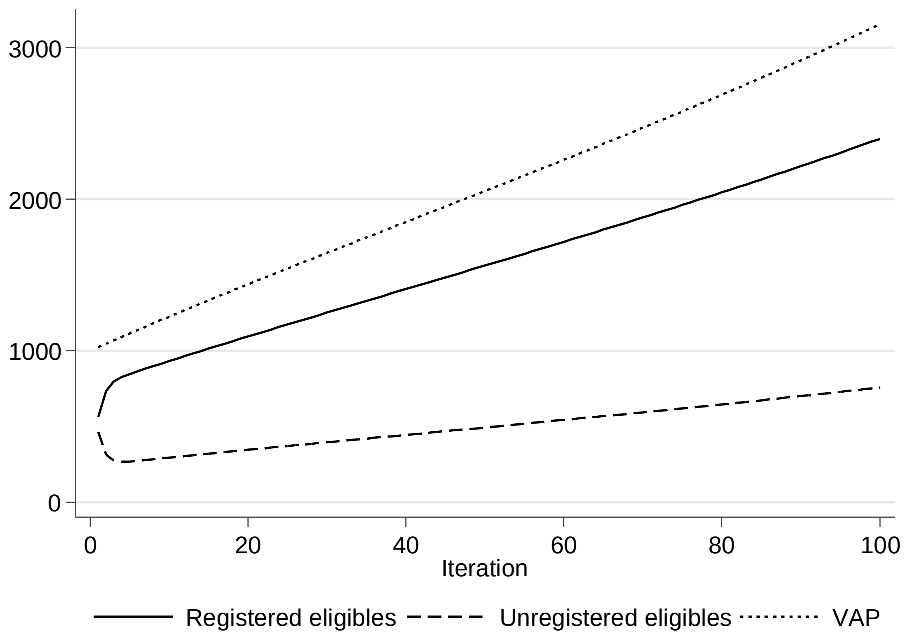
Figure 3. Dynamic representation of voter registration and deadwood.



R = registered eligible electorate
U = unregistered eligible electorate
B = newly eligible electorate due to citizens turning 18 since the last time period
I = newly eligible electorate due to moving into a state since the last time period
D = dead people who were in the eligible electorate the previous time period
M = movers who were in the eligible electorate the previous time period

Figure 4. Results from model example

a. Size of registered and unregistered population



b. Voter registration rate

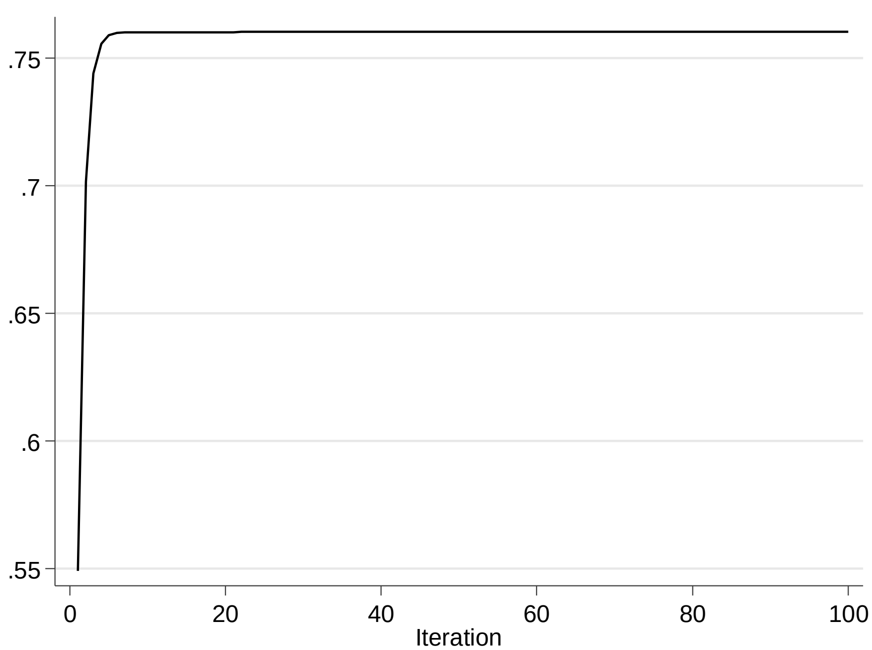
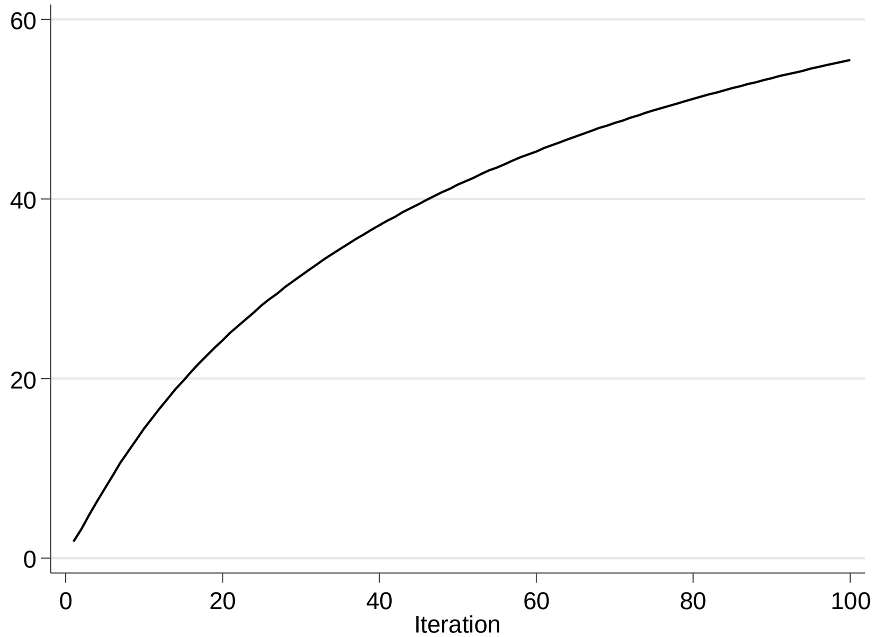


Figure 5. Deadwood percentage on voter rolls under two list maintenance regimes.

a. No list maintenance (Regime 1)



b. Deaths remain for one period, out-of-state moves remain for two periods (Regime 2)

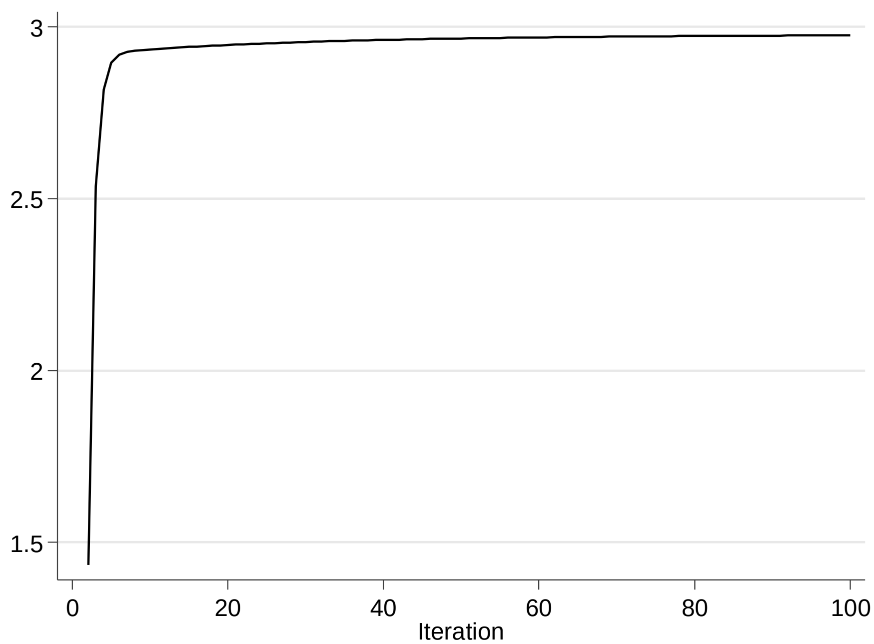
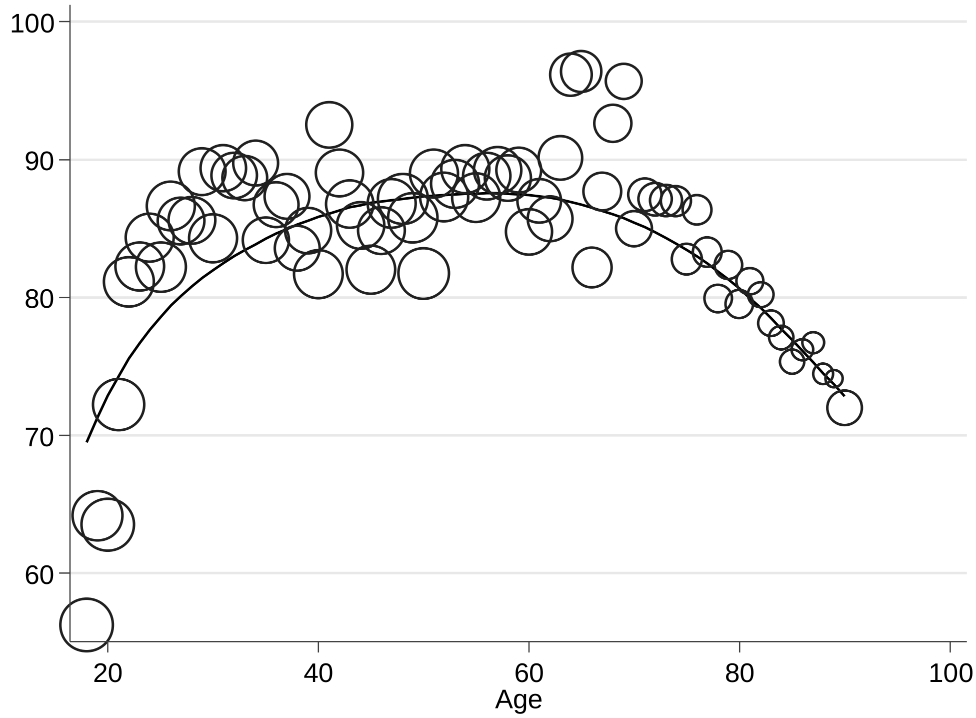
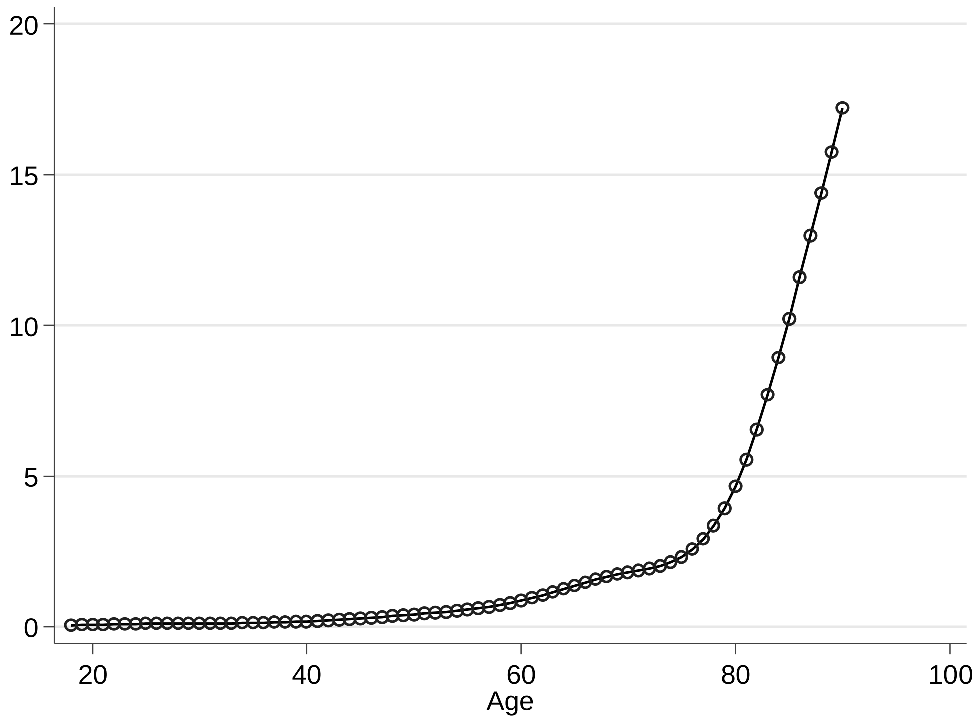


Figure 6. Registration rates by age.



Source: Catalist 1% sample file, June 2014 (registered voters); ACS micro-use data (eligible citizens).
Note: The size of the tokens is proportional to the number of eligible citizens in each age group. The line is a loess curve.

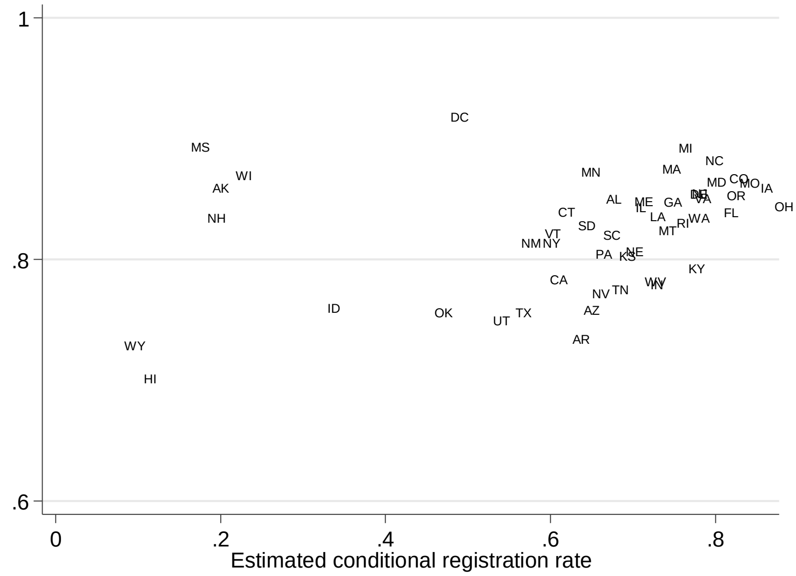
Figure 7. Death rates by age.



Source: CDC WONDER file, 2013.

Figure 8. Proportion registered in each state, measured two ways, plotted against estimates of state-level conditional registration rates.

a. Proportion registered calculated using VRS



b. Proportion registered according to official state reports

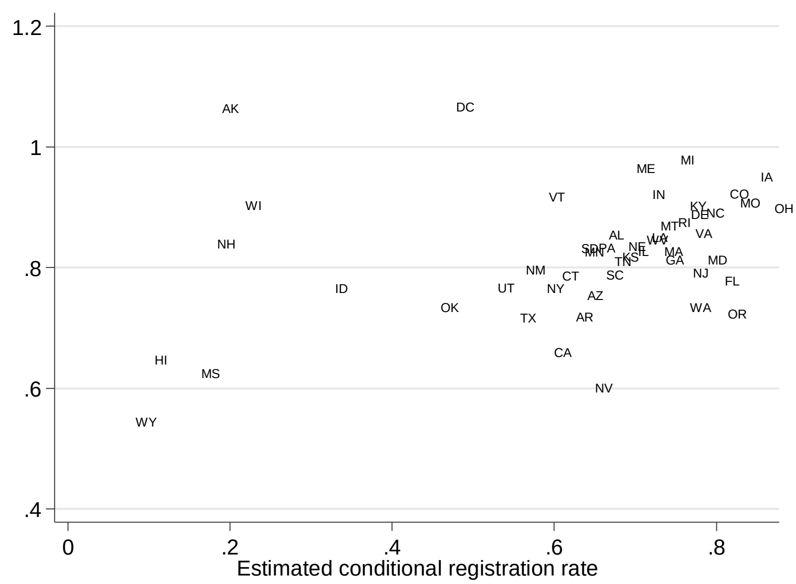


Figure 9. Estimated true registration rates plotted against officially reported registration rates in 2012.

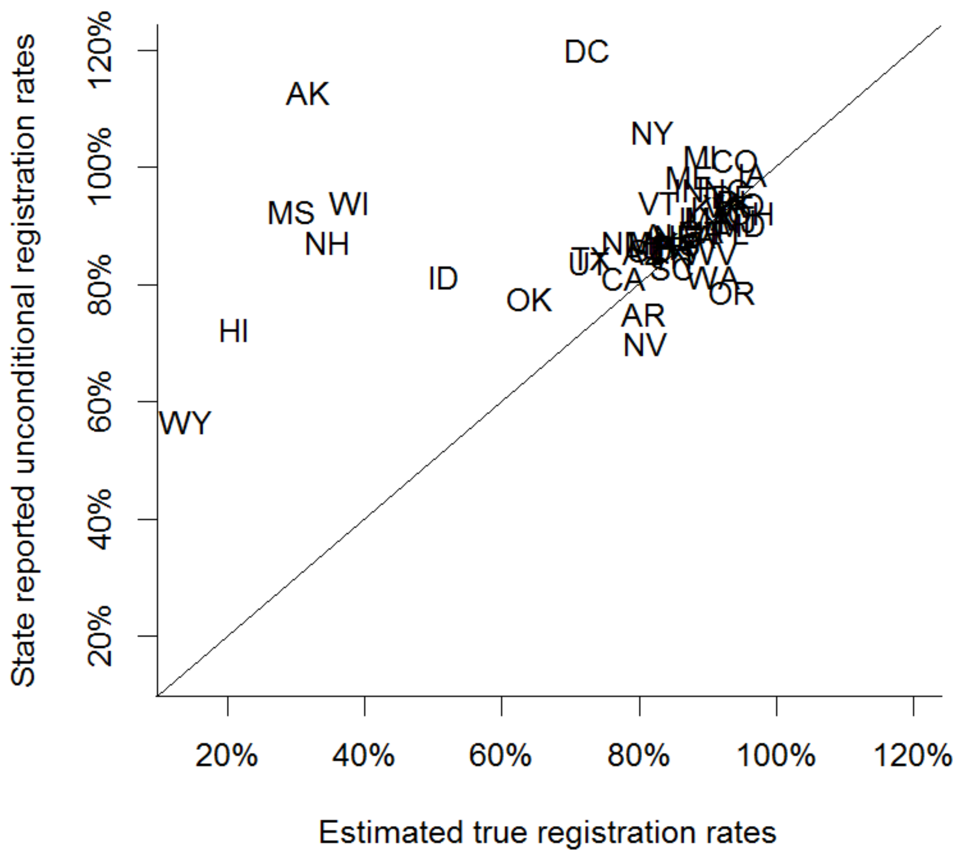
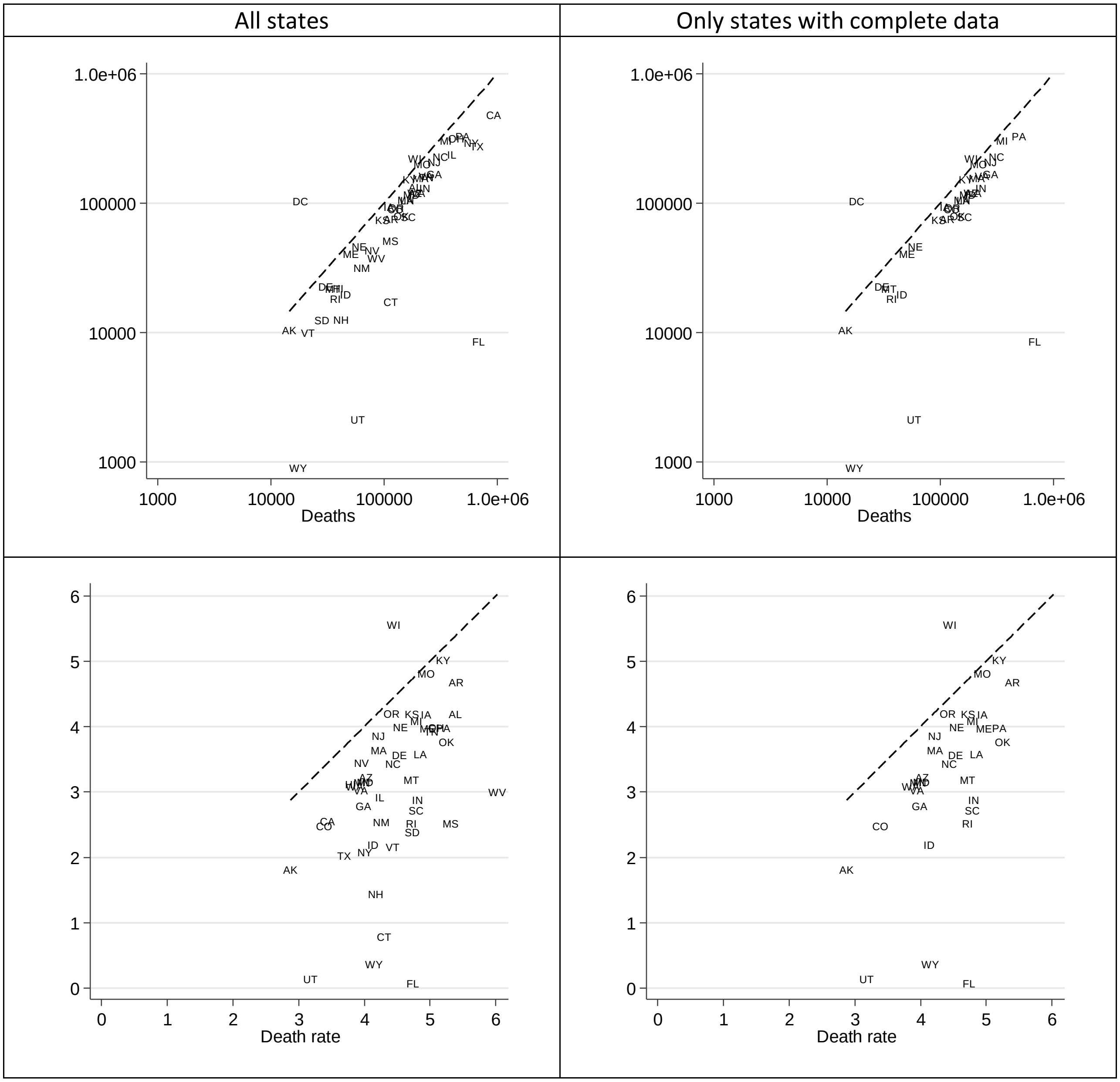


Figure 10. Relationship between death removals from voter registration lists and adult deaths at the state level, 2009–2012.



Source: EAC NVRA reports, 2008–2010 and 2010–2012 (death removals), and CDC WONDER (death rates).

Figure 11. Relationship between removals from voter registration lists due to out-of-state moves and adult out-of-state moves at the state level, 2009–2012.

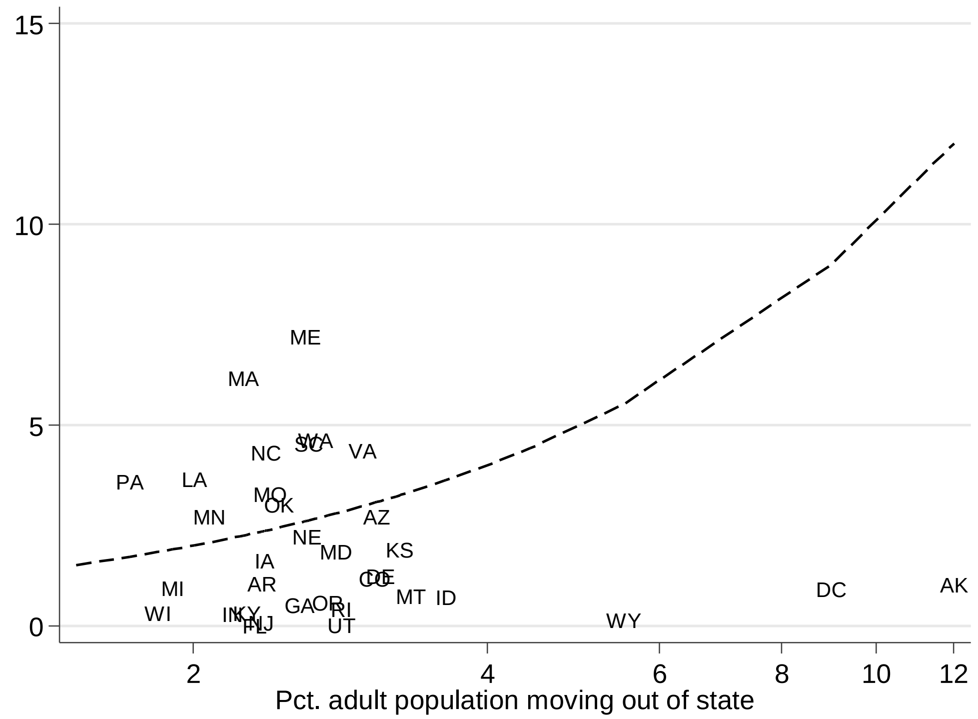


Figure 12. Distribution of deadwood percentage, with dead voters removed after 6 months and movers removed after 48 months.

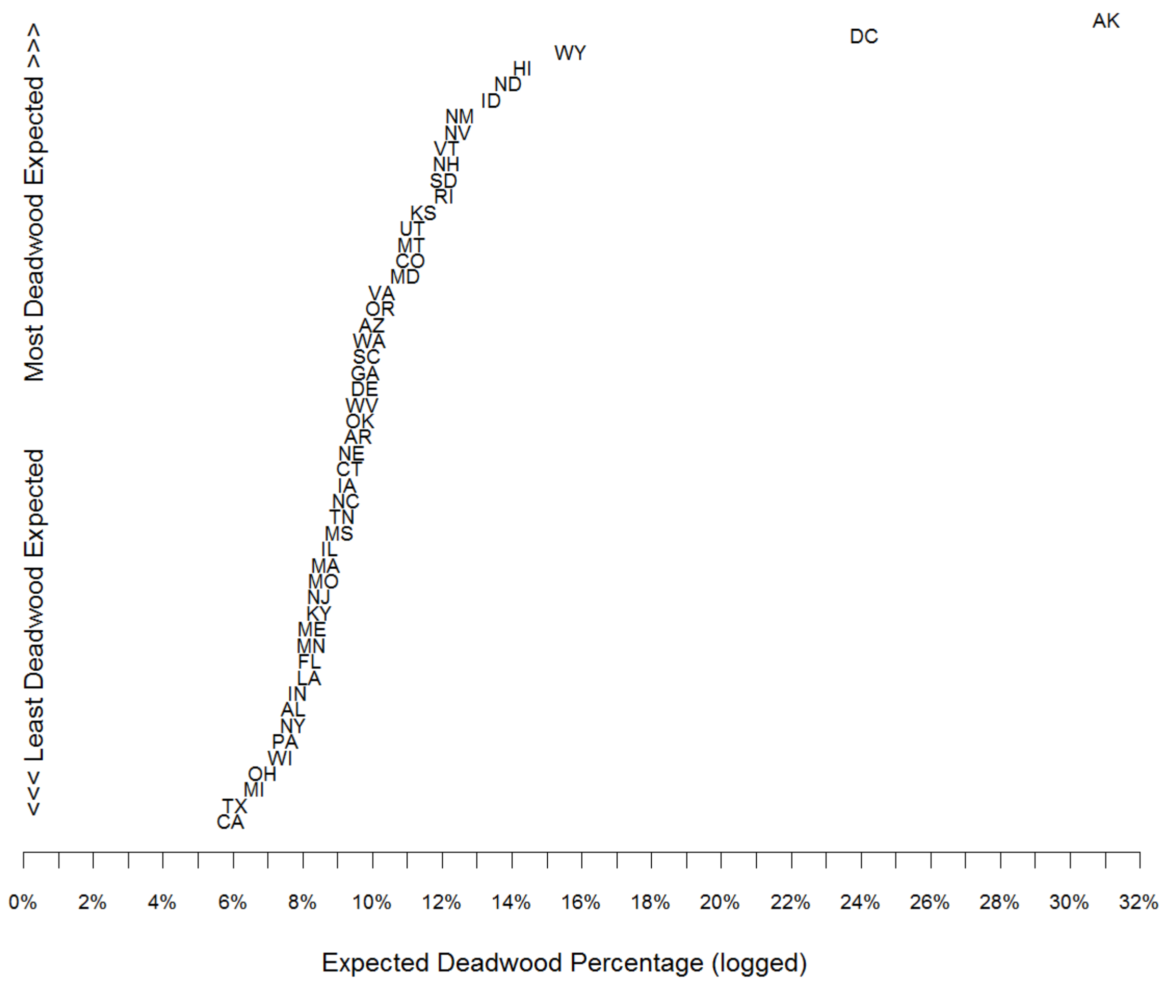


Table 1. Starting values of simulation example.

a. Population values

Registered eligibles (R)	0
Unregistered eligibles (U)	1,000
Deaths (D)	0
Moving out (M)	0
Ageing-in (B)	25
Moving in (I)	35

b. Weights (omitted weights set to zero)

Registered eligibles who stay registered (a_{rr})	0.814
Unregistered eligibles who register (a_{ru})	0.530
New adult eligibles who register (a_{rb})	0.830
New in-migrant eligibles who register (a_{rb})	0.830
Registered eligibles who are no longer registered (a_{ur})	0.167
Unregistered eligibles who remain unregistered (a_{uu})	0.451
New adult eligibles who do not register (a_{ub})	0.170
New in-migrant eligibles who do not register (a_{ui})	0.170
Registered eligibles who die (a_{dr})	0.007
Unregistered eligibles who die (a_{du})	0.007
Registered eligibles who move from the state (a_{mr})	0.007
Unregistered eligibles who move from the state (a_{mu})	0.007
Population growth due to births (a_{bb})	0.012
Population growth due to immigration (a_{ii})	0.012

Table 2. Conditional registration rates by state.

State	Conditional registration rate	State	Conditional registration rate
Alabama	0.657	Montana	0.720
Alaska	0.179	Nebraska	0.680
Arizona	0.629	Nevada	0.639
Arkansas	0.616	New Hampshire	0.173
California	0.589	New Jersey	0.760
Colorado	0.806	New Mexico	0.554
Connecticut	0.599	New York	0.580
Delaware	0.758	North Carolina	0.777
D.C.	0.468	North Dakota	0.087
Florida	0.799	Ohio	0.861
Georgia	0.726	Oklahoma	0.449
Hawaii	0.096	Oregon	0.803
Idaho	0.318	Pennsylvania	0.644
Illinois	0.693	Rhode Island	0.742
Indiana	0.710	South Carolina	0.653
Iowa	0.844	South Dakota	0.622
Kansas	0.673	Tennessee	0.664
Kentucky	0.757	Texas	0.547
Louisiana	0.709	Utah	0.519
Maine	0.691	Vermont	0.582
Maryland	0.778	Virginia	0.763
Massachusetts	0.725	Washington	0.756
Michigan	0.745	West Virginia	0.703
Minnesota	0.627	Wisconsin	0.208
Mississippi	0.153	Wyoming	0.072
Missouri	0.818		

Table 3. Starting values ($\mathbf{C}_0$) for simulations.

Variable	Description	Initial values
R	Registered eligible electorate	State-reported registered voters[a]
U	Unregistered eligible electorate	CVAP[b]-*R*
D	Dead formerly eligible voters	0
M	Moved formerly eligible voters	0
B	New 18-year-olds	Birth rate[c] ×CVAP[b]
I	Newly eligible in-migrants	In-migration rate[d] × CVAP[b]

[a]U.S. Election Assistance Commission, Election Administration and Voting Survey, 2012.
[b]Calculated from U.S. Census Bureau estimates.
[c]U.S. National Center for Health Statistics, National Vital Statistics Reports.
dU.S. Census Bureau, State-to-State Migration Flows, ACS 2013 one-year estimates, https://www.census.gov/hhes/migration/data/acs/state-to-state.html.

Table 4. Estimated true registration rates by state.

State	Estimated true reg. rate		State	Estimated true reg. rate
Alabama	83.7%		Montana	89.0%
Alaska	31.8%		Nebraska	85.4%
Arizona	80.7%		Nevada	80.9%
Arkansas	80.9%		New Hampshire	34.5%
California	77.7%		New Jersey	94.5%
Colorado	94.1%		New Mexico	78.2%
Connecticut	83.2%		New York	82.0%
Delaware	93.8%		North Carolina	92.7%
D.C.	72.4%		North Dakota	NA
Florida	93.4%		Ohio	96.5%
Georgia	89.1%		Oklahoma	64.0%
Hawaii	20.9%		Oregon	93.6%
Idaho	51.4%		Pennsylvania	85.1%
Illinois	87.7%		Rhode Island	93.5%
Indiana	87.3%		South Carolina	85.0%
Iowa	96.7%		South Dakota	81.7%
Kansas	85.8%		Tennessee	84.9%
Kentucky	90.6%		Texas	73.1%
Louisiana	88.4%		Utah	72.5%
Maine	87.3%		Vermont	82.9%
Maryland	94.9%		Virginia	92.5%
Massachusetts	90.2%		Washington	90.9%
Michigan	89.0%		West Virginia	90.5%
Minnesota	82.2%		Wisconsin	37.7%
Mississippi	29.3%		Wyoming	13.9%
Missouri	94.6%			

Table 5. Estimated percentage of deadwood on voter registration rolls, assuming different purging rules.

	Months after moving out of state							
Months after dying	0	3	6	12	18	24	36	48
0	0%	0.70%	1.40%	2.80%	4.10%	5.40%	7.80%	10.20%
3	0.90%	1.60%	2.30%	3.60%	4.90%	6.20%	8.60%	10.90%
6	1.30%	2%	2.70%	4%	5.30%	6.60%	8.90%	11.20%
12	1.70%	2.40%	3.10%	4.40%	5.70%	6.90%	9.30%	11.50%
18	0.20%	0.90%	1.60%	3%	4.30%	5.60%	8%	10.30%
24	2.50%	3.20%	3.90%	5.20%	6.50%	7.70%	10%	12.20%
36	3.30%	4%	4.70%	6%	7.20%	8.40%	10.70%	12.90%
48	4.40%	5%	5.60%	6.80%	8%	9.20%	11.50%	13.60%

Citations

Ansolabehere, Stephen and Eitan Hersh. 2013. Voter registration: The process and quality of lists. In *The measure of American elections* eds. Barry C. Burden and Charles Stewart III. New York: Cambridge University Press.

Ansolabehere, Stephen, Eitan Hersh, and Kenneth Shepsle. 2012. Movers, stayers, and registration: Why age is correlated with registration in the U.S. *Quarterly Journal of Political Science* 7(4): 333–363.

Ansolabehere, Stephen, Samantha Luks, and Brian F. Schaffner. 2015. The perils of cherry picking low frequency events in large sample surveys. *Electoral Studies* 40: 409–410.

Highton, Benjamin. 2000. Residential mobility, community mobility, and electoral participation. *Political Behavior* 22(2): 109–120.

Leslie, P.H. 1945. On the use of matrices in certain population mathematics. *Biometrika* 33(3): 183-212.

Lefkovich, L.P. 1965. The study of population growth in organisms grouped by states. *Biomeetrics* 21(1): 1–18.

Meredith, Marc and Michael Morse. 2015. The politics of the restoration of ex-felon voting rights: The case of Iowa. *Quarterly Journal of Political Science* 10(1): 41–100.

Richman, Jesse T., Gulshan A. Chattha, and David C. Earnest. 2014. Do non-citizens vote in U.S. elections? *Electoral Studies* 36: 149–157.

Shaw, Daron, Stephen Ansolabehere, and Charles Stewart III. 2015. A brief yet practical guide to reforming U.S. voter registration systems. *Election Law Journal* 14(1): 26–31.

Strate, John M., Charles J. Parrish, Charles D. Elder, and Coit Ford. 1989. Life span civic development and voting participation. *American Political Science Review* 83(2): 443–464.

Timpone, R.J. 1998. Ties that bind: Measurement, demographics, and social connectedness. *Political Behavior* 20(1): 53–77.

Tuljapurkar, Shripad and Hal Caswell. 1997. *Structured-population models in marine, terrestrial, and freshwater systems.* New York: Chapman and Hall.