

Movers, Stayers, and Registration: Why Age is Correlated with Registration in the U.S.

Stephen Ansolabehere, Eitan Hersh, and Kenneth Shepsle*

April 25, 2012

Abstract

Age is among the strongest predictors of political participation, yet it is also among the least well understood. We offer a probability model of participation in the U.S. voter registration system - the first step in the voting process. In this model, people have a constant probability of registering to vote at any given time and a constant probability of moving. A strong relationship between age and participation arises simply as a byproduct of the rules of the registration system, namely that participation is voluntary and that it is residentially based. Specifically, the probability that someone *is* registered increases over time (and thus with age) even when the probability of becoming registered is constant. A new, national random sample of 1.8 million voter registration records is employed to test the model. The model provides a theoretical foundation for the relationship between age and participation, identifies the functional form of that relationship, and solves a puzzle about the nature of participatory bias.

*Stephen Ansolabehere is Professor of Government, Harvard University. Eitan Hersh (corresponding author) is Assistant Professor of Political Science, Yale University, and can be reached at eitan.hersh@yale.edu or (203)436-9061. Kenneth Shepsle is George D. Markham Professor of Government, Harvard University. We thank Andrew Gelman, Erik Snowberg, and participants in the LSE-NYU Conference on Political Economy, London, England, May 2010.

Age is one of the strongest predictors of participation in U. S. elections, and of particular importance is the effect of age on voter registration (Wolfinger and Rosenstone 1980, Rosenstone and Hansen 1993). Approximately 75% of citizens ages 18-30 are registered to vote, while almost 90% percent of those at least 60 years-old are in the system. We argue that the correlation between age and registration can be explained almost entirely as a product of the electoral procedures that tie registration to residential location and on the effect of those rules on a society that is fairly mobile. We present a stochastic model in which we begin by assuming two facts about people. First, they move around. In the simplest version of the model, geographic mobility is assumed constant across age cohorts. Second, people register to vote at a constant rate. Voter registration drives and nudging by friends and family all may occur but they are assumed to affect all people equally. Although the simplified model strips away many specific facts about registration and voting, the assumptions square with the main features of the U.S. voter registration system. Registration is voluntary. If someone moves they must register again in order to stay in the voting system. If someone registers and does not move, they remain in the system. We call this the Stayer-Mover model of voter registration.

The model implies an exponential relationship between age and participation, distinct from the linear and quadratic relationship assumed in most empirical studies. Even with a constant rate of registering across the life span, the model highlights that the longer a person lives and does not move, the greater is the probability that the person is in the system. The proportion registered increases with each age cohort, but the rate of increase declines over time. Since the system acts as a ratchet - once in, one remains in until moving - age becomes correlated with registration. The mobility rate of a society determines the proportion that will be registered if the population is long-lived. In a population with no mobility, that asymptote is one - eventually everyone will become registered. If the mobility rate is higher than zero, the asymptote is somewhat lower than one. For the U.S. it appears to be in the

neighborhood of 0.85. This pattern is precisely what we observe in the data.

We put the predicted relationship to the test by partnering with Catalist - a data management firm that has compiled the voter registration files for every state. Catalist's database lists every registered voter in the U.S. as well as the ages of nearly all registered voters. We combine these records with data from the U. S. Census in order to measure the registration rate of every age group. This is the actual registration data for the entire nation, and is not subject to reporting biases common in political survey research. We show that the theoretical relationship derived fits the empirical relationship between age and registration rates very well; indeed, the theoretical model leads us to a statistical specification that is much tighter than any of the commonly used specifications estimated in prior empirical research. And, it casts the existing theoretical literature in an entirely different light.

Social science survey research has repeatedly found that older people are more likely to report that they are registered to vote. Various rationales are offered: older people have more experience and knowledge of the political system; they have come to realize the importance of participation; they have denser social networks and are thus subject to more social pressure to vote, and so forth. (See, for example, Dahl (1961); Verba, Schlozman, and Brady (1995); Rosenstone and Hansen (1993); Strate, Parrish, Elder, and Ford (1989).) All of these accounts suggest that aging and maturing causes people to participate – to register to vote and to vote at higher rates. To our knowledge, none of these explanations has been derived from first principles. Nor has there been any theoretical analysis of the assumptions behind these arguments to show that they generate a relationship between age and voting. Gill and Gaiours (2002) provide an excellent survey of various models of voter turnout and registration, showing the complexity of empirical analyses that have been conducted and the general lack of theoretically-driven probability modeling of electoral participation and behavior. Too often, they point out, empirical patterns are taken as theoretical models,

rather than guided by them.¹

Our model presents an alternative explanation of the correlation between age and registration that poses a broad challenge to more common social-psychological explanations of political participation. We propose a stochastic model of a population in which people enter, exit upon moving, and can reenter. The rules and institutions of the registration system define the process through which entry, exit, and reentry may occur. People enter the registration system because at each election they have a given probability of registering, and they exit the registration system if they moved. That dynamic generates a correlation between the fraction of people registered and age. In other words, we take as given a null model of the social psychology: in our model, age does not affect the propensity to register. We further strip away the decision theoretic or game theoretic aspects of the problem that have concerned many political scientists. A simple model of population dynamics for a population in which the propensity to register is constant is sufficient to predict the correlation between age and the fraction registered at any age.

Our empirical analysis conforms with our model's predictions in two important respects. First, the actual registration data reveal that the propensity to register is in fact fairly constant across people's life cycles, a fact that runs flatly against common social-psychological interpretations of survey research results. Second, the simple model derived from the assumption that the probability of registering is constant fits observed patterns of registration and age better than any model used in the empirical literature (especially linear models and step functions).

The analysis presented here also highlights a confusion in research on political participation between the *fraction* of people participating and the *propensity* of an individual to

¹One exception is the probability of being pivotal. Rational choice models of participation conjecture that turnout is a decision based on costs and expected payoffs, and the probability of casting the deciding vote shapes the expected payoff. Gelman, King, and Boscardin (1998) and Gelman, Katz and Tuerlinckx (2002) provided explicit derivations of the probability of being pivotal and find that the probability of being decisive in an election is surprisingly high.

participate. The former is the *stock* of registrants, and the latter characterizes the *flow*. Often what is measured in regression analyses predicting registration is the stock of registrants in a group, but the interpretations offered often discuss this as a flow - as an estimate of the marginal probability that an individual in a given group will become registered. This is the case when it is argued, for example, that older people have a higher propensity to register or vote because the coefficient on age is positive in a cross-sectional model conducted on a survey sample. Such cross-sectional regressions, however, only measure differences in the fraction of people who are registered across groups. Our analysis will clearly distinguish these two concepts. The empirics show that the propensity to register does not increase with age and may even decrease, but the stock of registrants increases with the age of the cohort simply because people drop out of the system and must reregister when they move. Hence, the correlation between age and registration in the United States is explained almost entirely by population mobility and the registration rules.

Registration, rather than voting itself, may initially seem like a minor part of political participation. However, as a screen for the act of voting, registration explains much of what separates voters from non-voters. According to the Current Population Survey, in 2008 71% of citizens were registered, and 90% of registered citizens voted (File and Crissey 2010). The correlation between age and voting is at least partly explained by the correlation between age and registration. Timpone (1998) goes so far as to argue that the correlation between voting and age is driven entirely by the correlation between registration and age. In the Catalist data, among registered voters, the correlation between voting and age is just .14, and the relationship between age and voting is essentially flat after age 30. At least in the United States, then, understanding registration is the first step to understanding voting, and we show that age, or more specifically - time spent in an electoral jurisdiction - explains much of the variance in registration rates.

1 Age, Mobility, and Participation

The persistent empirical relationship between age and political participation has attracted numerous explanations. The authors of *The American Voter* favored an explanation based on party attachment. Young people are less wedded to either of the parties and thus feel less motivated and obligated to participate in politics (Campbell, et al 1960, see also Achen 2006). Glenn and Grimes (1968) suggest that older citizens have fewer alternative distractions to voting and so participate at a higher rate than younger citizens. A more common explanation is that age is related to life experience and community ties, both of which lead to an investment in the political process, expressed through the acts of registering and voting (Rosenstone and Hansen 1993, Plutzer 2002, Jankowski and Strate 1995, Abramson, Aldrich, and Rohde 1998, Jennings 1979, Niemi, Stanley, and Evans 1984, Goerres 2007).

When pushed to elaborate the causal mechanism of how community attachment or personal experience leads to participation, these explanations become quite complex and virtually impossible to test empirically. For example, Strate, Parrish, Elder, and Ford (1989) describe a causal pathway in which age affects one's income level, which in turn affect one's competence as a citizen, which in turn affect one's sense that he or she can make a difference, which in turn leads one to register or vote. This complicated pathway is estimated with linear and polynomial regression, without clear indication as to why these specifications are appropriate. Certainly such a pathway is difficult to verify.

Personal experience and community attachment face compelling challenges as explanations for the relationship between age and participation. Highton and Wolfinger (2001) show that the personal experience that citizens typically acquire with age (income, education, marriage) do not seem to be the cause of differential participation by age group. Highton (2000) poses a critique of the community attachment argument by comparing the behavior of movers who left their communities with movers who stayed within their communities. He

finds that it is the requirement to re-register, not the uprooting of social ties, that reduces participation among recent movers. Mobility may disrupt community involvement (Putnam 2000), but this apparently is not a reason why moving leads to lower election participation. Further, in a study of the registration and turnout stages of voting, Timpone (1998) finds that nearly all the effect of age on participation occurs in the registration stage. Thus, whatever process is generating a correlation between age and turnout seems to be unrelated to the actual casting of ballots. If age proxies for investment in the political process or commitment to the community, it is indeed odd that the investment and commitment would not extend to the voting part of the electoral process.

One key fact demonstrated below runs contrary to the entire social psychological literature on age and registration. We will demonstrate empirically in Figure 1 that the propensity to register (the flow of registration) does not increase with age. Moreover, we will demonstrate that the rate at which people become registered, a key parameter in the stayer-mover model, also need not bear any relationship with age in order to generate a strong relationship between age and the probability of being registered.

Apart from social/psychological explanations, another rationale for the relationship between age and registration is structural. The laws governing participation determine which demographic characteristics will correlate with registration rates. Because registration is geographically-based in the United States, when Americans move, they usually must re-register to vote. Thus structural explanations of the relationship between age and registration tend to focus on mobility. Wolfinger and Rosenstone (1980), Squire, Wolfinger and Glass (1987), Highton and Wolfinger (1998), and Highton (2000) focus on mobility as a key explanation for political participation. Wolfinger and his colleagues find that moving has similar counter-participatory affects on old voters and young voters alike. Since younger people move much more frequently than older people, this may contribute to the difference

in participation rates by age.²

However, the correlation between age and registration is not simply because younger people move more. Even if Americans of all ages moved at the same rate, mobility itself would result in a positive relationship between age and voter registration. A stochastic process of registration over time with constant rates of becoming registered and constant rate of mobility produces as a byproduct the relationship between age and participation that has been observed for decades.

This idea, that the correlation between age and participation in the election system may have nothing to do with personal changes connected with aging, deserves special emphasis. When researchers observe a cross-section of the electorate, such as in the form of a representative survey sample or a voter registration file, they see the proportion of participants at any given age group (the stock) but they must look at their data in a different way in order to see the relationship between aging and participation (the flow). Here we show that the correlation between age and participation can be explained without invoking the personal changes that are assumed to occur as individuals grow older. Those changes may affect the voting stage of election participation and they may affect other aspects of political participation, but as we will demonstrate, they do not explain the core of the relationship between age and election participation.

2 A Stayer-Mover Model

Suppose a closed society has age cohorts $t = 1, \dots, T$ and that there is migration within the country, but not into or out of the country. The index t can also correspond to the number of elections, say federal elections, that an individual has lived through. Someone with $t = 1$ is in the first election and is in the youngest adult cohort, and so forth. Suppose

²See, also, Brians (1997) and Nie, Verba, and Kim (1974) for other mobility-focused analyses.

further that all people are characterized by two behavioral parameters. The first is a stability parameter: between each election some people move and others stay put. Let s be the rate of stability (non-movement) in the population, assumed constant and the same for all people in the basic model but allowed to vary in extensions. The second is a mobilization (or registration rate) parameter: between two elections there is a common rate at which people enroll to vote. Let r_t be that for cohort t . This is the flow parameter of registration.

Importantly, this is a model of the registration rate in the entire society or nation, such as the entire United States. The assumption of a closed society is not necessary but greatly simplifies the analysis. It is possible to allow for immigration or naturalization of citizens, but the key insights follow from the model presented here.³ It should also be noted that the analysis characterizes the national registration rates rather than the registration rates of individual areas such as specific states. Modeling local area registration is a worthy extension.⁴ In addition, the registration system is assumed to be current and once a person registers they remain in the system. The model does not allow for purging or obsolescence of the rolls. These variations can also be included, and might be of particular interest to research about registration reform. For the analysis presented here they are of small practical consequence, as discussed in the subsection on Purging below.

A simple accounting identity expresses how the stock of registration changes over time. Let R_t be the proportion of cohort t that is registered to vote. R_t is the stock of people who are registered. Although we assume constant r , it is useful to begin with a general accounting identity that characterizes the proportion of age cohort t who are registered, the stock of registrants of age t :

$$R_t = R_{t-1}s_t + (1 - R_{t-1})s_t r_t + (1 - s_t)r_t.$$

³See Popkin and McDonald (2001) for an analysis of the citizen population.

⁴One obvious first step is to allow the registration systems to vary across states to accommodate EDR and other laws.

In words, the proportion of cohort t registered at any time equals the proportion who were registered in the previous time and did not move, the proportion who were not registered in the previous time and did not move but decided to register this time, and the proportion who moved and registered. Equivalently, the proportion of cohort t registered equals the proportion of those who were not registered and become registered plus those who were registered and remain in the system (either because they did not move or because they moved and reregistered). That is, $R_t = R_{t-1}(1 - (1 - s_t)(1 - r_t)) + (1 - R_{t-1})r_t$. As an analytical matter, it is convenient to express this relationship as $R_t = r_t + R_{t-1}q_t$, where $q_t = (1 - r_t)s_t$.

We derive R_t as a function of r , s , and t . Here we present three key results that emerge upon considering variations on this accounting model in which (1) r and s are constant for all people, (2) r is constant but s varies over the course of an individual's life, and (3) r differs across types of people. Our argument is that r is approximately constant over the lifecycle for all people. That is, as people age their propensity to participate is not affected by aging or factors correlated with age, such as knowledge, social network density, education, or experience with the political system. This assumption is consistent with data on registration (see Figure 1 below). In assuming r is constant for individuals (even if we permit heterogeneity across cohorts), we are effectively stacking the deck against finding an "age effect," and are certainly not building it into the analysis by assumption. As we will show in the remainder of this section, even when there is no age effect on the probability that someone registers (r), there will still be a relationship between age t and the likelihood that someone *is* registered (R_t). We show in the next section that the empirical patterns of registration fit best the model in which r is constant and s varies over the life cycle, but the data are fit exceptionally well by the simplest model in which r and s are both constant.

2.1 Basic Model: Constant r and Constant s

The simplest version of the process assumes the probability that someone registers r and the probability that someone stays at their residence s are independent of time. That is, $r_t = r$ and $s_t = s$. Under these assumptions the fraction of people of a given age cohort t who are registered is increasing exponentially in t . Hence, even if there is no causal effect of age t on the probability of registering r there will still be a correlation between age t and being registered R .

RESULT 1: If r and s are constant over an individual's life and the same for all individuals,

$$R_t = r \sum_{j=0}^{t-1} q^j = r \frac{1 - q^t}{1 - q}, \quad (1)$$

where $q = s(1 - r)$. All proofs in the appendix.

This proposition captures the core insight of the paper. The propensity to register (r) does not vary with age, but the proportion of people registered to vote (R_t) is increasing in age (t). The reason is that the registration system acts as a ratchet. Once someone registers he or she remains registered until the individual moves. Because there is a constant propensity to register the system will accumulate registrants over time, and the proportion not registered will be the product of the likelihood of not registering in all previous elections. What slows the rate of accumulation of registrants is the mobility rate $1 - s$. As people move they are removed from the lists, and must start anew the registration process. The distinction, then, between r and R_t is that of a flow (r) versus a stock (R_t). The flow parameter in this model is constant by assumption and we have shown that, even still, the stock—the percent registered—is increasing in age.

Analysis of Equation (2) and variations on it will deepen understanding of the relationship between age and participation. One important initial implication is that a very strong

connection between age and registration may emerge and not be a causal relationship. Aging does not cause registration in the usual sense of a cause, as the parameter r does not vary with age. Nonetheless a tight relationship between age (t) and proportion registered (R_t) emerges simply from the interaction of two independent stochastic processes.

We will examine below the consistency of this model with observed patterns of registration in the United States. Before doing so, it is important to develop a variation of the model in which s varies with age.

2.2 Variable- s Model

Empirically, the parameter s varies with age. Census estimates of geographic mobility suggest that people under 30 are highly mobile. Approximately 50 percent of those between 18-35 years old remain in the same residence for two years. Those over 35 become more sedentary, and approximately 83 percent do not move in a two-year interval. There are further variations within these groups. The 24-29 year olds have somewhat higher mobility than the 18-23 year olds. And, those over 70 are the least mobile, with 87 percent having resided in the same residence for at least two years. Given the large variation in mobility over the life-cycle, we are particularly interested in the implications of a model in which s varies with age, but r does not.

A simple expansion of the model accommodates two generations. Denote generations as a for younger and b for older. Younger people, in this model, have mobility parameter s_a , and they may participate in the first T_a elections, indexed $t_a = 1, \dots, T_a$. The older generation has mobility parameter s_b . If they move less often, then $s_b > s_a$. The older generation has already experienced the first T_a elections. The older generation faces elections $T_a + 1, T_a + 2, \dots, T_b$. For sake of mathematical representation, we will index the generation b elections as $t_b = 1, \dots, T_b$.

RESULT 2: If generation A has parameter s_a and generation B (older) has parameter s_b , then

$$R_{t_a}^a = r \frac{1 - q_a^{t_a}}{1 - q_a}$$

$$R_{t_b}^b = r \left(\frac{1 - q_b^{t_b}}{1 - q_b} + q_b^{t_b} \frac{1 - q_a^{T_a}}{1 - q_a} \right)$$

where $q_a = s_a(1 - r)$ and $q_b = s_b(1 - r)$, generation a consists of T_a periods, and generation b consists of T_b periods. If $s_b > s_a$, then $q_b > q_a$ and the function R_t^b will be flatter, but follow a trajectory higher than R_t^a .

These formulations may be extended to cover more than two generations. One may generalize the spline function above to cover many different generations if s is highly variable, but that propagates parameters. As a practical matter it will be easier to estimate the variant of the model in equation (1). Namely, $R_t = R_{t-1}s_t + (1 - R_{t-1})s_t r + (1 - s_t)r$, which reduces to $R_t = r + (1 - r)s_t R_{t-1}$.⁵ We will estimate this linear model below, in which the dependent variable is R_t and the independent variable is $R_{t-1}s_t$. The intercept equals r , and the slope equals $1 - r$, so a handy test of the appropriateness of the model will be whether the intercept equals one minus the slope.

2.3 Variable- r Model

Variation in r across types of people accommodate possible social psychological effects such as heterogeneity in the degree of engagement (Campbell 1960) and learning.

2.3.1 Heterogeneity in enrollment arising from "types." The parameter r may vary across types of people. Campbell's (1960) distinguishes high engagement people (r_{high}) and low

⁵One may also express this recursion in terms of r and s_t . $R_t = r(1 + \sum_{j=1}^t \prod_{i=1}^j s_i(1 - r)^{i-1})$, where $s_1 = 0$.

engagement people ($r_{low} < r_{high}$). Suppose α percent of voters have r_{high} (in the extreme $r_{high} = 1$) and $1 - \alpha$ have r_{low} . Note that r is constant for each type; it does not vary with age or previous record of registration. Under these assumptions, application of Result 1 to the two types of people yields:

$$\begin{aligned} R_t &= \alpha r_{high} \sum_{j=0}^{t-1} q_1^j + (1 - \alpha) r_{low} \sum_{j=0}^{t-1} q_0^j \\ &= \alpha r_{high} \frac{1 - q_1^t}{1 - q_1} + (1 - \alpha) r_{low} \frac{1 - q_0^t}{1 - q_0} \end{aligned} \quad (2)$$

where $q_0 = s(1 - r_{low})$ and $q_1 = s(1 - r_{high})$.

One may readily expand (4) to include more types of people, with the result that the fraction of cohort t who are registered to vote is a weighted average of the relevant fractions for each type of person. The model with heterogeneity implies a much flatter relationship between t and R_t than Results 1 and 2.

2.3.2 Learning arising from prior registration. Arguments that people learn how to register by being registered or that registration is habit-forming imply that the fact that one is in the system changes the probability of registering in future years. Assume the registration rate of those not registered in the prior period is r_0 and the registration rate of those registered in the prior period is r_1 . In period 1, no one has been registered before so $R_1 = r_0$. In period t , anyone who has registered and moved has a higher probability of reregistering than someone who is presently not registered and has not moved. Hence, $R_t = s[R_{t-1} + (1 - R_{t-1})r_0] + (1 - s)[R_{t-1}r_1 + (1 - R_{t-1})r_0]$. Expressing the recursion fully,

$$R_t = r_0 \sum_{j=0}^{t-1} p^j = r_0 \frac{1 - p^t}{1 - p}, \quad (3)$$

where $p = s(1 - r_1) + (r_1 - r_0)$. Result 1 (Equation 2) is the case where $r_1 = r_0$.

As a practical matter it is exceptionally difficult to distinguish the model with learning effects from the more parsimonious model with constant r . The fact that r is constant or perhaps even declining with t (see Figure 1) suggests that learning effects are likely quite small.

2.4 A Complexity: Purging

An extension of the model would incorporate purging of the lists. Purging occurs for two reasons - obsolete registration, say due to moves or deaths, and non-participation. Obsolete registrations are captured in the model already, though they do present a complication in data analysis, as discussed in the empirical portion of the paper. Purging due to non-participation, on the other hand, applies to people who registered, did not move, but did not vote, and thus can create “leakage” from the registration lists.

Specifically, the National Voter Registration Act allows states to remove people from the active voter files only when the individuals have not voted in two successive federal elections. Those persons may then be placed on an inactive voter list and eventually purged after at least two elections. That mechanism is a constraint on purging; many states allow for an even longer grace period before removing an inactive registration.

Purging can be introduced by creating a second process (voting) and a purge rule that is a function of that second process. One may assume, for example, that there is a constant probability of voting, v , and failure to vote in n successive elections leads to a purge. This additional “leakage” process would flatten the predicted relationship between age and voting, as newer registrants could not be purged but people who have been in the system n elections have a probability of being purged of $(1 - v)^n$.

We do not incorporate this phenomenon here as it complicates the basic insight and is an order of magnitude smaller than mobility. Analysis of the Catalist data suggest that approximately 5 percent of registered voters who did not move failed to vote in two successive

elections.⁶ Every 4 years, almost 40 percent of the U. S. population moves at least once. Purging due to obsolete registrations is clearly the main source of churning in the lists, not failure to vote. Consequently, we will focus the remainder of this analysis on the implications of mobility.

3 Empirical Predictions of the Model

The relationship between age and registration defined in Equation (2) provides a foundation for empirical analysis. Already one important insight emerges. A positive correlation between age and registration may emerge simply because of population mobility. The model allows much more precise analysis of that relationship. Many empirical studies estimate mean registration conditional on age in a linear or quadratic form. The rationale behind such a relationship is rarely spelled out. We have derived from first principles an exponential relationship. Specifically, $R_t = \gamma(1 - q^t)$, where $\gamma = r/(1 - q)$. In the next section we estimate this relationship using data on registration in 2008. Here we derive some basic predictions.

3.1 Comparisons

Consideration of Equation (2) reveals that the percent registered in any country with a registration system like the U. S. system will exhibit a positive correlation between age (t) and percent registered (R_t). Any increase in the propensity to register (r) will increase the percent registered in each cohort and an increase in mobility will flatten the correlation between R_t and t .

A particularly interesting case to consider with this model is a system in which $s = 1$. One interpretation of such a system is that there is no mobility in the society. A second

⁶We examined a three year panel of data from Florida, Illinois, and Connecticut. This result is consistent with a simple calculation. Approximately 85 percent of registered Americans vote in Presidential elections and 70 percent vote in Congressional elections. If non-voting is independent over time, then we expect 4.5 percent of people ($.3 \times .15$) to fail to vote in two federal elections.

interpretation is that this is a system in which registration is mobile, it moves with the person. Such a system is actually used in many countries as it is an important feature of what are called “universal registration” systems. The individual must still opt in but once in, the person never drops out. Under that rule, we have

$$R_t = 1 - (1 - r)^t$$

The percent registered in the society and at any age cohort but the first is much higher than in the US-style system in which people must re-register. It should be noted, however, that universal registration systems such as in many European systems will still lose people due to mobility depending on how their registration system operates. Any opt-in sort of rule, then, will create a correlation between age and the percent registered.

3.2 Statistical Characteristics

Normative analysis of the registration systems focuses on the representativeness of the registered electorate compared with the population. Deviation of the average age of the registrant pool from the average age of the adult population captures the unrepresentativeness of the electorate. Under the simple model, the average age of the registered electorate will be higher than the age of the typical adult. The question is by how much.

Here we derive the median and mean cohort t (which is age divided by 2) for registrants and the population using a simple model of the population (zero growth). It is worth noting that zero growth means that we are holding the population sizes of the cohorts constant, which sharpens the theoretical prediction. To characterize fully the statistical or aggregate behavioral patterns predicted by the model requires an additional assumption about cohort sizes and population growth. Zero growth implies that each cohort is of the same size, $P_t = n$. There are T cohorts, so the total population is $P = T \times P_t = Tn$. In the appendix,

we present results for constant growth. Further modeling would carry insights about the effects of baby-boomers and other population shocks. Such analyses are for later work.

The expression for R_t defines the percent of a given cohort who are registered. The number of people in a given cohort who are registered is $R_t P_t$. We characterize the age distributions of registrants and the total population here and, later, compare them with the actual empirical distributions observed in the U. S. population.

The total number registered in cohort t is, from equation (2), $nR_t = \frac{nr(1-q^t)}{(1-q)}$. The total number of people in the entire population who are registered is $R = \sum_{t=1}^T nR_t = \frac{nr}{1-q}(T - \sum_{t=1}^T q^t) = \frac{nr}{(1-q)}(T - \frac{q(1-q^T)}{(1-q)})$. Hence, the fraction of all registrants who are cohort (age) t is

$$f_t^0 = \frac{nR_t}{R} = \frac{1 - q^t}{T - \frac{q(1-q^T)}{1-q}}$$

The proportion of all people who are of a given cohort (age) is $h_t^0 = \frac{P_t}{P} = \frac{1}{T}$.

In a stable population, the average and median cohorts of the population are $T/2$. Under the assumption of zero growth, the average age of the registrants is defined as $\mu_0 = \frac{1}{T} \sum_{t=1}^T tR_t$. The median age of registrants is defined as the value m_0 such that $\sum_{t=1}^{m_0} nR_t = .5 \sum_{t=1}^T nR_t$. Hence the mean (μ_0) and median (m_0) values of m among registrants in a stable population are defined as follows.

RESULT 3. The average and median ages of registrants under zero population growth are:

$$\mu_0 = \sum_{t=1}^T t f_t^0 \approx \frac{T}{2} + \frac{1-q}{q} \tag{4}$$

$$m_0 \approx \frac{T}{2} + \frac{q^2}{2(1-q)}. \tag{5}$$

Recall that the median cohort in the population is $T/2$, which is the first term on the

right hand side of the approximations of the mean and median. If q is about equal to $1/2$, the mean age of registrants is predicted to be about 1 cohort (or 2 years) higher than that of the population, and the median age of registrants is predicted to be about $1/4$ of a cohort (or a half year) older than the median age of the adult population.

The magnitude of the difference between average or median age of the population and average or median age of registrants is a particular concern for those interested in representation. If the registration process yields large disparities between the medians of registrants and the population, then the electorate may differ significantly from the rest of the adult public, and representatives will respond to their electorates, rather than to the adult population. Importantly, under either constant and zero growth, we expect that the average and median ages of registrants will be only slightly higher than the population. Linear regression analyses of survey data about participation, such as the Wolfinger and Rosenstone (1980) and Verba et al. (1995) classic studies, find only slight deviations between the the average ages of the participant pool and the general population, just as predicted.

4 Empirics

The statistical predictions emanating from the model square with some of the basic facts about registration in the United States. Registration is predicted to be correlated with age, and the mean and median ages of registrants are expected to be only slightly higher than the adult population. How well the data describe the overall pattern of registration across age groups is a more challenging question. The model points toward a more exact analysis of the relationship between age and registration. An investigation of actual registration records will not only permit us to test the model's predicted relationship between age and participation, but will also allow us to test key model assumptions, such as the constant propensity to register across age groups.

Registration statistics are derived from a 1% national random sample of registration records from Catalist, a Washington, DC based firm that vends data to the Democratic Party and left-of-center interest groups. Catalist acquires voter files from state and county election offices in all fifty states. The firm cleans the lists, removes deadwood and duplicates, and matches records against a variety of other data sources such as the National Change of Address (NCOA) database, the Social Security Death Index, and lists provided by commercial vendors. These costly services improve the efficiency of voter targeting for the political campaigns that Catalist serves. For a detailed discussion of the Catalist data-cleansing process, the quality of records originating from the official records in every state, and concerns about partisan bias in the Catalist data, consult Ansolabehere and Hersh (2010).

The 1% sample of all registration records was drawn in the Spring of 2010. The initial sample includes 1,898,427 individual records. Of these, about 70,000 do not have birth years available either through official records or commercial matched records. Another 5,000 records are from the very small group of people who turned eighteen after the 2008 election and registered since then. Because the age cohorts in this analysis are conceived of as 2-year cohorts, we begin the first cohort with the population that was 18 during the 2008 general election. Finally, we remove “deadwood” from the sample. Catalist creates a model that predicts how likely it is that each record is obsolete. A combination of factors, including vote history, age, inactive status, address changes, Social Security death notices, and others, are used to predict deadwood status. Roughly four percent of the sample is predicted to be “deadwood” or “probably deadwood.” Designation as deadwood and age are not positively correlated because while older people are more likely to be deceased (and thus appear as deadwood), younger people are more likely to move residences (and thus appear as deadwood). After removing registration records that appear to be deadwood, the final sample of registrants with age information includes 1,744,560 records.

The baseline citizen population is derived from the 2008 American Community Survey,

administered by the U.S. Census Bureau. The weighted three million person sample enables us to estimate precisely the population distribution of citizens by age. Other official data sources are available for alternative estimates of the citizen population, such as the full population counts from the Census or the Current Population Survey (CPS); however the Census full population counts are not available for detailed subsets (e.g. counts of citizen non-movers by age group) and the CPS samples are smaller than the ACS.

Aside from citizen population counts and registrant population counts by age cohort, the only other statistic necessary to estimate for this analysis is s_t , the rate of non-moving for citizens by age. The ACS asks individuals if they lived in the same residence one year ago. Because we want to rate of non-movement for a two-year election cohort, we square this rate for each group. This strategy probably makes an unrealistic assumption of statistical independence, but the analysis is robust to other ways of estimating s , such as using statistics for the full population (rather than the citizen population) or by using estimates from the CPS. As a robustness check, though, the estimated rate of non-movement derived by squaring the one-year non-mobility rate in the ACS is consistent with the rate of non-mobility derived from citizens who report not moving in a two-year period in the 2008 Cooperative Congressional Election Study (CCES). Squaring the ACS one-year non-mobility rate is not perfect, but it is preferable to other estimation options that would sacrifice precision by using estimates from smaller samples, or that would increase bias by using a non-mobility rate from statistics on the full population rather than the citizen population.

Table 1 details the statistics from the ACS and Catalist that are used in this analysis. The entire empirical analysis (except Figure 4, see below), including all regressions, can be replicated using this complete set of data listed in Table 1. It is worth emphasizing that, until now, the simple set of variables displayed in Table 1 have been very difficult to analyze. Only a few years ago, when registration records were not yet all digitized and computational power and data storage mechanisms were not as well advanced, scholars

Table 1: Summary Statistics from the 2008 American Community Survey (ACS) and Catalist 1% Sample

Age Cohort	American Community Survey			Catalist		s_t $Col(3)^2$	R_t $Col(5)/Col(2)$
	Citizen Sample (1)	Est. Citizen Pop. (2)	Non-mover for 1 yr. (3)	Registrant Sample (4)	Est. Reg. in Pop. (5)		
18-19	87,256	8,342,610	0.71	50,298	5,029,800	0.50	0.60
20-21	83,270	7,961,492	0.67	53,836	5,383,600	0.45	0.68
22-23	76,138	7,279,567	0.63	59,630	5,963,000	0.40	0.82
24-25	74,919	7,163,078	0.65	58,343	5,834,300	0.42	0.81
26-27	75,345	7,203,816	0.69	60,318	6,031,800	0.47	0.84
28-29	71,768	6,861,780	0.73	58,479	5,847,900	0.53	0.85
30-31	68,265	6,526,878	0.77	55,951	5,595,100	0.59	0.86
32-33	65,465	6,259,104	0.79	54,311	5,431,100	0.63	0.87
34-35	67,879	6,489,920	0.82	55,078	5,507,800	0.67	0.85
36-37	74,425	7,115,848	0.84	60,857	6,085,700	0.70	0.86
38-39	76,230	7,288,431	0.85	63,576	6,357,600	0.73	0.87
40-41	76,520	7,316,113	0.87	61,306	6,130,600	0.75	0.84
42-43	79,400	7,591,426	0.88	63,743	6,374,300	0.77	0.84
44-45	86,956	8,313,938	0.89	69,797	6,979,700	0.79	0.84
46-47	85,828	8,206,106	0.90	71,133	7,113,300	0.80	0.87
48-49	86,646	8,284,294	0.91	71,494	7,149,400	0.82	0.86
50-51	86,619	8,281,695	0.91	71,085	7,108,500	0.83	0.86
52-53	81,002	7,744,685	0.92	68,800	6,880,000	0.84	0.89
54-55	77,933	7,451,190	0.92	65,765	6,576,500	0.85	0.88
56-57	71,667	6,852,093	0.93	62,271	6,227,100	0.86	0.91
58-59	68,917	6,589,165	0.93	58,617	5,861,700	0.87	0.89
60-61	70,212	6,713,007	0.93	59,919	5,991,900	0.87	0.89
62-63	53,118	5,078,618	0.93	46,133	4,613,300	0.87	0.91
64-65	52,879	5,055,757	0.94	45,389	4,538,900	0.88	0.90
66-67	44,598	4,264,000	0.94	39,196	3,919,600	0.89	0.92
68-69	39,950	3,819,604	0.95	34,215	3,421,500	0.89	0.90
70-71	37,144	3,551,374	0.95	31,610	3,161,000	0.90	0.89
72-73	33,792	3,230,866	0.95	29,175	2,917,500	0.90	0.90
74-75	31,441	3,006,064	0.95	26,039	2,603,900	0.90	0.87
76-77	29,905	2,859,186	0.95	24,420	2,442,000	0.89	0.85
78-79	27,921	2,669,500	0.94	23,020	2,302,000	0.88	0.86
80+	111,921	10,700,797	0.92	90,756	9,075,600	0.84	0.85

Note: To ease comparison between the ACS and Catalist statistics, it is helpful to scale them to the size of the full population. Because the Catalist sample is a 1% random sample for registration records, we multiply the cohort sizes by 100. For the ACS, we take the Census estimate of an adult citizen population of 206,072,000 during the 2008 election (File and Crissey 2010), and weight up the ACS figures so that age cohorts sum to that population estimate. The precision of this weighting is not critical to the main findings, as the age groups are all multiplied by constants.

needed to rely on small-N public opinion surveys to study election participation. Survey self-reports of election participation are problematic because of sampling bias, misreporting bias, and small numbers of observations. For example, in recent years, 85-90% of National Election Studies respondents report being registered. When the reported non-registrant pool in a survey consists of only 100-200 people, as in the National Election Studies, very little insight can be gained.

The national Catalist data listed in Table 1 is also superior to a method that relies on one or more state voter files. As the theoretical model is at the national level, it is important here to have the national set of registered voters, which only the political data vendors like Catalist produce. Furthermore, because Catalist tracks movers and cleans records of deadwood, estimates of the registered population are cleaner than if we relied only on raw voter files. Thus, the relationship we can now observe using Catalist's registration data and the Census Bureau's estimates is a major step forward in the study of election participation.

As a final point about the data, it is worth noting that while Catalist records (as well as raw voter file records) list each voter's registration date, it is not advisable to use registration date as an indication of when someone entered the registration system. Knowing when a voter entered the system could provide us with leverage on the flow of participation over the lifespan. However there are two problems with relying on the date of registration. First, a voter's registration date is often changed by election officials after the voter moves and re-registers or even after the voter changes something about his or her record (like a shift in party affiliation). Second, once voters leave the election rolls, we no longer have a record of their registration date and so we cannot capture the flow of participation among this important subset of individuals. For these reasons, we must focus here only on the stock - the proportion of voters in any given age group currently registered to vote. Of course, this is the same quantity observed in the vast majority of vote studies that rely on cross-sectional survey samples.

4.1 Propensity to Register Does Not Increase with Age

The key assumption of the Stayer-Mover model is that the *propensity* to register (r) does not increase with age even while the *proportion* registered (R) does increase with age. This assumption challenges the canonical notion that aging itself leads to a higher propensity to participate. By virtue of stronger ties to political parties, more information about politics, or increasing levels of social pressure, the propensity to participate has often been assumed to increase with age (e.g. Strate, et al 1989). Because the Stayer-Mover model takes issue with this long-held assumption, our first empirical test is to estimate r with data.

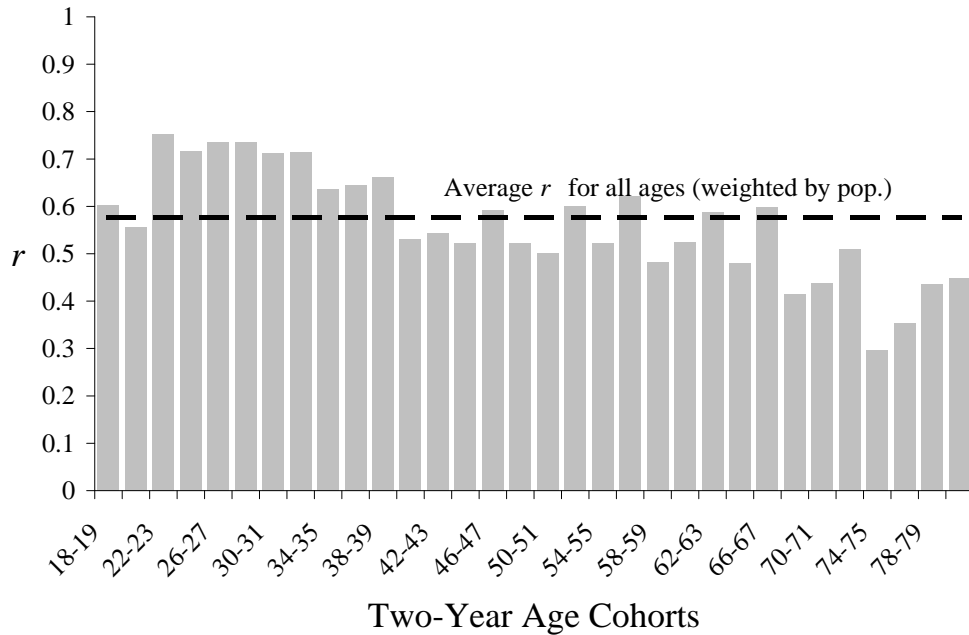
We calculate r using a generalized form of the model. Allowing r and s to vary with age yields a recursive formula: $R_t = R_{t-1}s_t + (1 - R_{t-1})s_t r_t + (1 - s_t)r_t$. Solving for r_t gives

$$r_t = \frac{(R_t - R_{t-1}s_t)}{(1 - R_{t-1}s_t)}. \quad (6)$$

Notice that r_t is a function of the proportion registered in the current age cohort, the proportion registered in the previous cohort, and the mobility rate by cohort. These parameters can easily be estimated with the data in Table 1. Using Census data on mobility and population age group size and Catalist data on registration, we estimate r_t for each cohort and display the relationship in Figure 1. The relationship depicted in the Figure can be best interpreted as the combined rate of new registrations and re-registrations for each age group.

Contrary to expectations from previous models of aging, the propensity to register does not increase with age. On the contrary, the rate of registration is declining with age. We conjecture that the negative pattern is a consequence of heterogeneity of types: in the younger cohort, there is a larger class of non-registrants who will register automatically. As citizens age, the pool of still-unregistered voters consists mainly of people who require more prodding to sign up. Whatever the reason, the observed propensity to register runs contrary to much of what has been argued in the past about the reasons age and registration are

Figure 1: New or Re-Registrations by Age Cohort (r_t)



correlated.

Other estimates of r_t are possible by studying the re-registration behavior of movers, where movers are identified in the Catalist database through the National Change of Address (NCOA) registry. In that estimation, the general pattern is the same as in Figure 1 (i.e. declining with age), though estimation using the NCOA flag is problematic because many people, particularly younger people, do not fill out NCOA forms. Figure 1 provides the cleanest estimates of r_t . No matter how we estimate r , the key model assumption that the propensity to register does not increase with age is confirmed. Though slightly declining with age, age groups exhibit a rate of registration near the average of 0.58, as indicated in Figure 1.

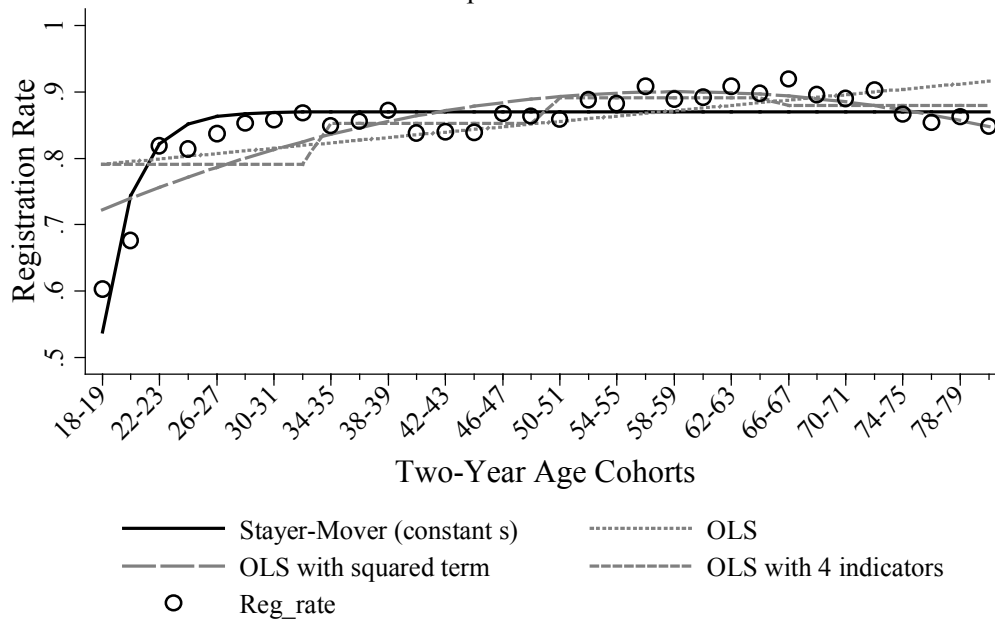
4.2 The Model versus Common Statistical Specifications

Having established the key model assumption, we now observe the proportion registered by age group (R_t). The general prediction from our analysis is that the relationship between

registration rates and age follow an exponential form. Researchers commonly estimate linear and quadratic forms. That the relationship between age and registration is non-linear in the parameters implies that no linear function, including no transformed polynomial function, is appropriate to fit to the data. This suggests that most previous models of the relationship between age and political participation are misspecified.

We fit registration data to the model through nonlinear least squares regression. This will allow us to derive average estimates of r and s . Equation (2) can be transformed into a simple exponential regression model of the form $R_t = \alpha(1 - \beta^t)$, where $\alpha = \frac{r}{1-q}$ and $\beta = q$. (Recall that $q = s(1 - r)$.)

Figure 2:
Actual and Expected Registration Rates by Age Cohort
Comparison of Models



Note: Each observation represents the registered population in the cohort, estimated from voter registration records nationally ($N=1,744,560$), divided by the citizen population, estimated from Census data. Predicted values from four alternative models are compared. The OLS model is of the form $R_t = \alpha + \beta t$; the OLS model with squared term is of the form $R_t = \alpha + \beta_1 t + \beta_2 t^2$; and the four-indicator model is of the form $R_t = \alpha + \beta_1 I_{34-49} + \beta_2 I_{50-65} + \beta_3 I_{66+}$, where I represents a dummy variable for ages noted in subscript. The exponential model is of the form $R_t = \frac{r}{1-q}(1 - q^t)$, as described in the text.

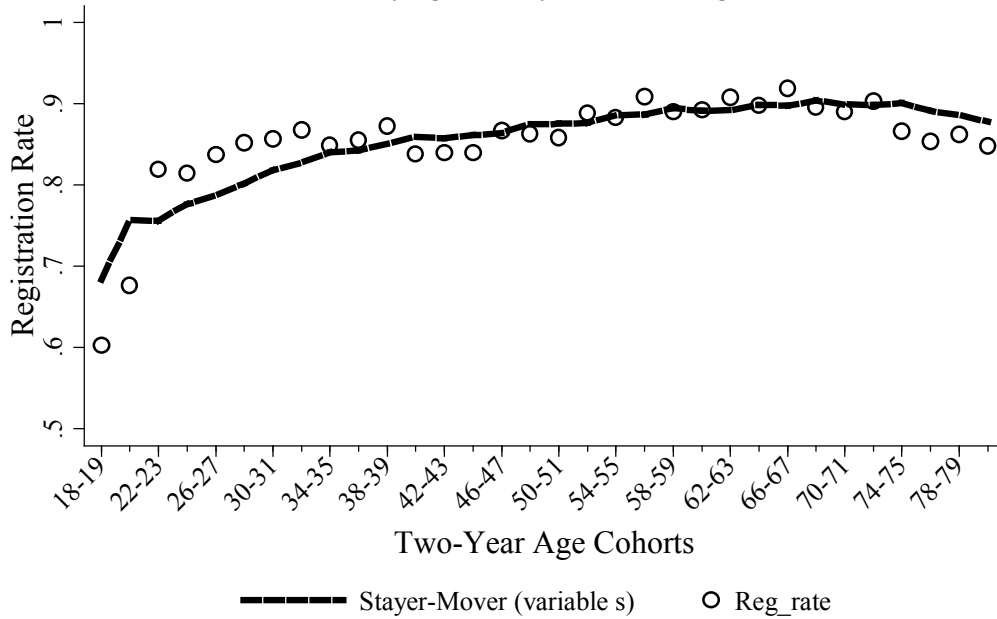
Figure 2 displays the predicted values from the exponential regression model. The fit is not perfect but clearly outperforms all of the standard specifications, such as the linear specification, the linear specification with a squared term, and a specification in which age groups are broken into four dichotomous variables and fit with OLS. The reason why the exponential fit is superior is plain to see: most of the increase in registration by age is concentrated in the first few age groups. After the third age group, the relationship looks flat. It is the first jump between ages 18 year olds and 22 year olds that the exponential model captures and that the other models cannot accommodate.

The coefficient estimates from the non-linear least squares model are $\hat{\alpha} = 0.87$ (0.01) and $\hat{\beta} = 0.38$ (0.03), where parentheses contain standard errors. From these coefficients, we derive $\hat{r} = \hat{\alpha}(1 - \hat{\beta}) = 0.54$, and $\hat{s} = \frac{\hat{\beta}}{1 - \hat{r}} = 0.83$. The theoretical expectation is that r will equal the rate of registration among the first adult age cohort, 18-19 year olds, i.e. $R_1 = r$. The implied rate of 54% is quite close to the actual rate of registration for this cohort in the 2008 general election, 60%. It is also quite consistent with the average estimate of r_t shown in Figure 1 of 58%. The non-mobility rate estimate of 83% is ten percentage points higher than the national non-mobility rate of 74%, estimated by squaring the percentage of ACS respondents who reported that they had not moved in the last year.

4.3 Estimation Allowing s to Vary

The difference between the Census mobility estimate and the model estimate may be an artifact of how mobility is measured. In the model, it is assumed that all movers must re-register. In the United States, the vast majority of moves take place within a county jurisdiction, and re-registration is not necessarily required. We suspect that a deeper issue with the specification is that the basic model is too simple. It is well known, for instance, that population mobility rates (i.e., $1 - s$) change with age. To make things more complicated, the mobility rate does not decrease linearly with age. Mobility actually increases for the first

Figure 3:
 Actual and Expected Registration Rates by Age Cohort
 Varying Mobility Rates with Age

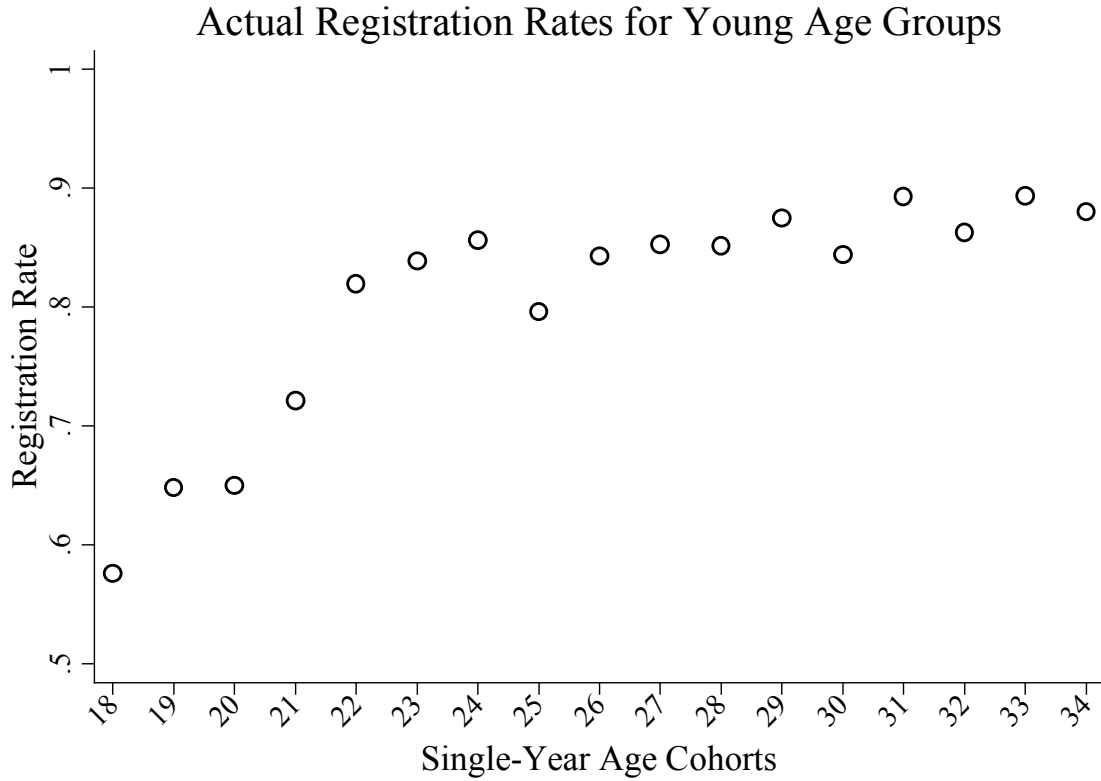


Note: Each observation represents the registered population in the cohort, estimated from voter registration records nationally (N=1,744,560), divided by the citizen population, estimated from Census data. Predicted values are generated from an OLS regression of R_t on $R_{t-1}s$ where $\alpha = r$ and $\beta = 1 - r$.

few age cohorts and then begins to decrease.

To accommodate variation in the non-mobility parameter, s , we return to the recursive formula $R_t = r + (1 - r)R_{t-1}s_t$. We can observe R_t and R_{t-1} from the Catalist and Census data and s_t from Census data. The recursion, then, takes a linear form $y = \alpha + \beta x$, where α will equal r and β will equal $1 - r$ and where $y = R_t$ and $x = R_{t-1}s_t$. For this estimation to be valid, α should be equal to $1 - \beta$. Figure 3 shows predicted values from this model. From the model, we derive $\hat{\alpha}$ as .68 (.02) and $\hat{\beta}$ as 0.27 (0.03), with standard errors in parentheses, which comes very close to satisfying the assumption that $\alpha = 1 - \beta$. As is evident from Figure 3, compared to the formula with a constant s , this recursion formula does not fully capture the sharp increase in participation amid the youngest cohorts, but overall, the fit is

Figure 4:



Note: Each observation represents the registered population in the cohort, estimated from voter registration records nationally (N=1,744,560), divided by the citizen population, estimated from Census data.

very close to the data.

As one final view of the data, we confront the concern that our estimation utilizing two-year age cohorts is highly dependent on one or two data points representing the young cohorts. Figure 4 shows the registration rate of single-year age cohorts for young voters, using the same Catalist database and Census citizenship statistics as employed in the preceding analysis. In effect, this figure “zooms in” on the cohort analysis in Figures 2 and 3, and focuses on the age group where there is real movement in the rate of registration. We see that there is a steady and sharp increase in the rate of registration for the first groups and then a leveling-off.

5 Discussion

The correlation between age and participation is one of the most enduring regularities in political science. It is also one of the least well understood. We have derived a simple model in which aging does not affect the marginal propensity to register to vote, and is thus not causal in the usual sense of that term. Nonetheless, an exponential relationship between age and registration rates emerges through the process of mobility. To a first order of approximation, this functional form fits the observed relationship between age and registration well. Indeed, it fits better than a function in which registration linearly increases with age, as is commonly used in the empirical literature. The analysis here puts the relationship between age and registration on a new theoretical footing, and calls into question the large literature that asserts an actual causal connection between aging and registration, based on correlations between age and registration.

The analysis pushes forward the empirical, as well as the theoretical, understanding of the relationship between age and registration. A few points deserve emphasis. First, empirical researchers often conflate the probability that an individual is registered, R_t , a stock, and the probability that an individual is mobilized or becomes registered in a given election, r , a flow. For example, Rosenstone and Hansen (1993) predict that the probability of registering and voting is a function of campaign contact (which would affect r) as well as age and length of residence (which correlate with R_t). The model presented here clarifies that there are two distinct notions of the registration rate (R_t and r) which should be distinguished in empirical research.

Second, the accounting we have offered leads to a specific framework and functional form within which to study variations. The exponential relationship between age and registration – that expressed in equation (2) – defines how age, mobilization rates (r), and social mobility (s) shape the registration rate. This emerges from an accounting identity that must hold.

Almost all empirical research in this domain uses a linear or quadratic relationship between age and registration, which has no basis in theory. The parameters of this model (r and s) may be further parameterized as functions of other factors and variables in order to explore how variables that shape r and s influence registration, but the framework proposed here is the generic accounting of the relationship between the stock of registered persons and the propensity to register, mobility, and time.

We see five avenues for further extension of this framework. First, we do not explain where the marginal propensity to register, r , comes from. It is roughly constant with age (or possibly declining), and it is probably a product of the political and social system. Second, we have set aside the question of voting, the second stage to electoral participation. The proportion of the entire population that votes will follow the relationship derived here, as registration is a condition for voting in nearly all states. It remains an open question whether any additional correlation between age and voting can be shown to result from population mobility. Other forms of political participation apart from voting may be correlated with age (though Rosenstone and Hansen (1993) do not find this to be the case (see their Figure 5-1)). Third, more refinements of the model can incorporate *purging*. Some of the residual from our estimation, especially among older cohorts, is surely attributable to the failure of election authorities to purge from their rolls voters no longer in residence.

Fourth, there is the prospect that this model, developed with the current U.S. rules in mind, can be extended to other registration systems or consideration of other systems. Even “universal” election systems like those found in Canada and Europe often have trouble keeping the mobile population in the system. Research by Elections Canada finds that there is a correlation between age and the percent registered, which, they argue, accounts for some of the observed correlation between age and turnout (Black 2003). A 2005 study of the voter registration system in Britain finds similar relationship between age and registration as we find here, attributable in part to issues of mobility (Electoral Commission 2005). A

cross-national study would be a worthy extension of this research.

Finally, the relationship derived here allows us a framework within which to examine the effects of other demographic characteristics, such as family structure, home ownership, or race. Our approach delineates two ways in which particular demographic factors will matter – mobility rates and marginal propensities to register. Even if the marginal propensity to register is the same for every individual, the argument here reveals exactly how differential mobility will affect the degree to which people are excluded from the voting system. According to the American Community Survey

- 28% of renters move each year, compared with 5% of home owners
- 21% of households with a single parent with children under 18 move each year, compared with 10% of married households.
- 19% of unemployed persons move each year, compared with 12% of employed persons
- 13% of those with annual income less than \$30,000 move each year, compared with 7% of those with income over \$100,000.
- 8% of those in the Northeast move each year, compared with 14% in the West and 13% in the South.

Renters, single parents, unemployed persons, and Westerners and Southerners routinely show up in correlational analyses as people less likely to be in the registration system. Given their relatively high mobility rates, these people are much less likely to be registered simply because they moved. Our argument suggests that these are not necessarily out of the system “by choice,” but because they moved and the system places an additional burden on them of having to re-register. And as we have shown with younger persons, for that reason alone they are much less likely to register and, thus, to vote.

6 Appendix

6.1 Constant Growth Model

As a companion to the discussion in the text on statistical predictions emanating from an assumption of zero population growth (section 3.2), consider an analysis using the assumption of constant population growth. Constant growth means that the percentage change in the population from one cohort to the next is the same, say g . Assume the first cohort is size n and the rate of growth is g . Hence, the size of any cohort t is $P_t = n(1 + g)^{t-1}$ and the total population is $P = \frac{n}{g}[(1 + g)^T - 1]$.

The second $n(1 + g)$, the third $n(1 + g)^2$, and the j th is $n(1 + g)^{j-1}$. For the constant growth model it will be convenient to reverse the order of summation so that $t = 1$ is the youngest cohort and $t = T$ is the oldest cohort, and any cohort above T dies. The total size of such a population is $P = \sum_{t=1}^T n(1 + g)^{T-t} = \sum_{t=0}^{T-1} n(1 + g)^t$. It will be convenient to write $w = \frac{1}{1+g}$. Hence, $P_t = n\frac{w^t}{w^T}$ and $P = \sum_{t=1}^T n\frac{w^t}{w^T} = \frac{nw(1-w^T)}{w^T(1-w)}$.

$$f_t = \frac{nR_t}{R} = \frac{(1 - w^t)w^t}{J},$$

where $J = \frac{(1+g)^{T+1} - (1+g)}{g} - \frac{(1+g)^{T+1} - w^{T+1} - (1+g-w)}{(1+g-w)}$. The proportion of all people who are of a given cohort (age) is

$$h_t = \frac{P_t}{P} = \frac{w^t(1-w)}{(1-w^T)w}.$$

The average and median age of the population and registrant pool are characterized as follows.

RESULT 4B. The average and median ages of registrants under constant population

growth are:

$$\mu_g = \sum_{t=1}^T t f_t \approx \frac{1}{K} \left[\frac{w}{1-w} \left(\frac{(1-w^T)}{1-w} - T w^T \right) - \frac{qw}{(1-qw)^2} \right], \quad (7)$$

$$m_g \approx \frac{\ln(.5(a+w)^T)}{\ln(w)}. \quad (8)$$

where $K = \left[\frac{w}{1-w} ((1-w^T) - \frac{qw}{(1-qw)}) \right]$ and $a = \frac{(1+g)(1-q)}{(1+g)(1-q)+gq}$.

Under constant population growth, the statistical distribution of the registrant population is biased toward older people. Consider the differences in average ages. The average age (cohort) in the population is $E[t_p] = \sum_{t=1}^T t g_t = \frac{1}{w} - \frac{T w^T}{1-w^T}$. Further algebraic manipulation of μ_g reveals that $\mu_g = \frac{E[t_p] - \left[\frac{qw}{(1-qw)^2} \right] \left[\frac{1-w}{w(1-w^T)} \right]}{1 - \left[\frac{qw}{(1-qw)} \right] \left[\frac{1-w}{w(1-w^T)} \right]} > E[t_p]$.

The median age of registrants under the model also exceeds the median age of the adult population. The median age in the population under constant growth is $m_p \approx \frac{\ln(.5(1+w)^T)}{\ln(w)}$. Because $a < 1$, $m_R < m_p$.

6.2 Proofs of Propositions 1 and 2

Proof of Result 1: From (1), $R_t = qR_{t-1} + r$ for the t -th cohort and $R_1 = r$ for the first cohort. Repeated substitution in the recursion formula yields

$$\begin{aligned} R_t &= r + r(1-r)s + r(1-r)^2 s^2 + \dots + r(1-r)^{t-1} s^{t-1} \\ &= r[q^0 + q^1 + q^2 + \dots + q^{t-1}]. \end{aligned}$$

This partial series yields Equation (2).

Proof of Result 2: For the first generation the percent registered at age (or time) t_a is the same as equation (2). Now, however, the mobility parameter is s_a and $q_a = s_a(1-r)$. That

is,

$$R_{t_a}^a = r \sum_{j=0}^{t_a-1} q_a^j = r \frac{1 - q_a^{t_a}}{1 - q_a}. \quad (9)$$

The second generation differs in several respects. The year $T_a + 1$ is the first year of generation b. The registration rate of those in year $T_a + 1$ follows a similar formulation as before, but with a different start point and with a different mobility rate. Already $R_{T_a}^a$ of those in year $T_a + 1$ are registered. Between elections T_a and $T_a + 1$ the probability that an individual does not move is s_b . Proceeding as before, the registration rate of an individual of generation b in election $T_a + 1$ (i.e., $t_b = 1$) is: $R_1^b = R_{T_a}^a s_b + (1 - R_{T_a}^a) s_b r + (1 - s_b) r$. In the next election, $R_2^b = R_1^b s_b + (1 - R_1^b) r s_b + (1 - s_b) r$, and so forth. In the t_b th election, $R_{t_b}^b = q_b R_{t_b-1}^b + r = r + r q_b + r q_b^2 \dots + r q_b^{t_b-1} + q_b^{t_b} R_{T_a}^a = r \frac{1 - q_b^{t_b}}{1 - q_b} + r q_b^{t_b} \frac{1 - q_a^{T_a}}{1 - q_a}$. The result follows on application of Result 1 to each of the cohorts.

Gender, Race, Age and Voting: A Research Note

Stephen Ansolabehere* Eitan Hersh†

December 8, 2011

Abstract

In this brief analysis, we use a new dataset of two million voter registration records to demonstrate that gender, race, and age do not correlate with political participation in ways that previous research has shown. Among Blacks and Latinos, women participate at vastly higher rates than men; many Blacks participate at higher rates than Whites; and the relationship between age and participation is both not linear and varies by race and gender. Survey research is unable to capture the true relationship between demographics and participation, on account of survey bias and, more importantly, the non-linearity of effects. As a result, theories of participation, like the dominant resources-based models, have been built on faulty premises and tested with inadequate data. Our evidence calls for a renewed effort to understand election participation by utilizing large datasets, by being attentive to linearity assumptions, and by returning to theory.

*Professor, Department of Government, Harvard University. 1737 Cambridge St., Cambridge, MA 02138.

†Assistant Professor, Department of Political Science, Yale University. 77 Prospect St., New Haven, CT 06520.

Using a sample of two million registration records, we demonstrate that the ways that race, gender, and age correlate with political participation in the United States are misunderstood. Political scientists have long used survey samples of 1,000 to 15,000 respondents to study political participation. Because of the predictors of self-reported behavior that routinely arise in such surveys, scholars have inferred that a certain set of social and political processes must drive participation. Specifically, arguments that people’s “resources” drive participation dominate the field. These arguments predict that people who are disadvantaged or discriminated against in society are less likely to get involved (e.g. Rosenstone and Hansen 1993, Verba, Schlozman and Brady 1995).

We show that some of the most basic facts that led scholars to formulate the resource theories are wrong. Survey research finds that Whites participate in politics more than Blacks and men participate in elections at similar rates as women. According to Verba, Burns, and Schlozman (1997), there is a “well known equivalence between men and women in electoral turnout.” Based on how voting models are often estimated, participation is also thought to increase linearly or quadratically with age. Each of these findings and assumptions will be shown to be incorrect in ways that call for a rethinking of the resource-based theories themselves. The theoretical contribution of this essay is to challenge resource-oriented participation models with new evidence that demographic groups considered to be lacking in participatory resources are registering and voting at the highest rates.

Data

In 2010, we partnered with Catalist, a data vendor to Democratic politicians and liberal interest groups that collects voter registration records, cleans and updates them, and provides a database of all registered voters in the U.S. to its political clients.¹ For the present study, we use Catalist’s national sample of 1% of all active registration records. In most jurisdictions,

¹For more information, consult Ansolabehere and Hersh (2010).

age and gender are available on the voter file. When gender is unavailable, it can be predicted by registrants' first names. Voters in eight Southern states report their race in the public record. For voters in other states, Catalist makes a prediction of race based on voters' names and Census block-group contextual data. In a study in which we tested the accuracy of Catalist's racial predictions, we found that 91% of the time Catalist's prediction of race matches a voter's self-reported race. In this analysis, we use Catalist's racial predictions for the entire sample; however, if we were to restrict the sample to voters who report their race on the voter file, or to voters whose races are predicted with the highest degree of confidence, the results do not meaningfully change.

Catalist's records provide estimates of registrants and voters. For this study, we focus on turnout in the 2008 general election as recorded on voter files by election officials.² However, the effects identified are not particular to 2008. We also utilize citizen population estimates as a baseline for participation. To create such estimates, we start with the Census Bureau's full population estimates by gender, race, and single year of age for 2009.³ We then use the American Community Survey (ACS) to generate the percentage of each age-gender-race cohort that are citizens, and we weight the population statistics accordingly.⁴ The ACS has a sample size of three million residents, and is sufficiently large to provide accurate citizen estimates for all age-race-gender cohorts.

Results and Discussion

Table 1 shows the rate of registration among citizens, the rate of voting among registrants, and the rate of voting among citizens, by gender and race. In all racial groups, women are registered to vote at *higher* rates than men. The gap differs by race: it is 2-3 points for

²In a study of election administration quality, Ansolabehere and Hersh (2010) find turnout statistics on voter files to be highly accurate.

³"NC-EST2008-alldata: Monthly Population Estimates by Age, Sex, Race and Hispanic Origin for the United States," November 2009, United States Census Bureau.

⁴American Community Survey, "2008 1-year PUMS File," United States Census Bureau.

Whites and Asians, but 14-16 percentage points for Hispanics and Blacks. Given that a person is registered, there is a separate effect of gender on voting, with women in all racial groups voting at higher rates. Again, these differences are small for Whites and Asians and large for Blacks and Hispanics. The two gender effects together (the effect on registration and the effect on turnout) contribute to the overall participation gap, as witnessed in the third section of the table. The largest gender gap is among Blacks: Black women are 17 percentage points more likely to vote than Black men. Apart from identifying substantially larger gender effects than previously found, the evidence in Table 1 also challenges the assumption that Whites participate more than Blacks. In fact, Black women register at a higher rate than any other group, and overall Black women vote at a higher rate than White men.

Figure 1 adds age into the analysis. For each race-gender cohort, the rate of registration among citizens (column 1), the rate of voting among registrants (column 2), and the rate of voting among citizens (column 3) are shown. The figures estimate voters in 3-year age cohorts, beginning with 18-21 year olds. The gender gap is somewhat higher among young people, especially for Blacks, but it is persistent across most age groups. The relationship between age and participation is very non-linear, and it varies substantially by racial group.

The evidence presented holds important methodological and substantive lessons. Methodologically, most social science surveys lack the statistical power to measure the non-linear relationships detected here. This paper estimates participation rates for twenty-one age cohorts, four racial groups, and two genders, or 168 cells. The results reveal a strong three-variable interaction. The 2008 National Election Study, which contained an over-sample of Black respondents, included only 238 Black men. Before even considering issues of vote misreporting and sample selection, the NES would not have the power to detect even sizable interactions of race and gender. Similarly, Burns, Schlozman, and Verba (1997) report regression coefficients that in fact show a large interaction between race and gender, but they discount the results because of the standard error. Larger samples like the Current

Population Survey (CPS) can be more helpful, though these surveys still face the problems of misreporting and sample selection (Silver, Anderson and Abramson 1986)), and even the CPS cannot measure with much accuracy the behavior of relatively small groups, such as Asians.

Without the standard modeling assumption that demographics like age, race, and gender affect election participation identically across groups, and having enough data to examine sub-groups, we need not turn to regression in this analysis. Even if our analysis observed just five age cohorts rather than twenty-one, a regression table interacting race, gender, and age would necessitate 40 coefficient estimates, and that would be before incorporating any other control variables that one might consider important. The methodological take-away is not that scholars must estimate fully-interacted models or else their results are unreliable and since most studies do not use data sources with two million observations, all hope is lost. Rather, the methodological lesson is that theory must guide the modeling process. When empirical tests of theoretical models are susceptible to non-linearities and group-size confounders, as they are in regard to the resources model of participation, then models should be estimated to capture the nuances of the relationship.

In substantive terms, the data show a surprising interaction between gender and race that runs contrary to most work on this subject. Most studies of gender and participation offer theories of why men participate in politics more than women, not the other way around. Some writers suggest that men have a greater taste for politics than women (e.g. Burns, Schlozman and Verba (1997, 2001)). This blanket argument cannot explain why among Blacks and Hispanic women participate in elections at vastly higher rates than men. There is also some debate over the basic facts. Some scholars claim that women participate less, or that the participation rates by gender are equivalent; other scholars have found evidence in surveys that women participate slightly more than men (e.g. Lien 1998, Seltzer, Newman and Leighton 1997), and a few scholars have noticed that Black women report higher turnout

than Black men (e.g. Seltzer, Newman and Leighton 1997, Hornor 1999, Lien 1998). Among Latinos, the previous evidence has been mixed about differential participation in voting between men and women (see Montoya, Hardy-Fanta, and Garcia 2000).⁵ Looking at a national sample of voting data, we see that Black and Hispanic women vote at much higher rates than Black and Hispanic men, and the difference is much greater than for Whites. Moreover, young Black women participate at even higher rates than young White women. It is difficult to square these patterns with resource-based theories, and even more difficult to reconcile them with arguments that men are more attracted to politics than women.

There are several obvious potential explanations for the large gaps in participation between minority women and men. One suggestion is felony disenfranchisement laws, which disproportionately affect minority men. While these laws deserve further attention, a state-by-state analysis of the Catalist data reveals that in states with the most liberal felony disenfranchisement laws, the gender gap is even greater than in states with harsher laws.⁶ Socio-economic status and marriage rates hint at other hypotheses. Black and Hispanic men are less likely to be well-educated, wealthy, and married, and this in turn may lead to lower participation rates. Our own preliminary analyses discredit this view, as we find that at all socio-economic levels there is a substantial gender participation gap among minorities, at all socio-economic levels there is only a modest gender gap among Whites, and the gap exists among married as well as unmarried minorities.⁷ A third potential explanation is the religious and community involvement of minority women. Since the late 1800s Black churches have been mostly female (Wiggins 2005, see also Lincoln and Mamiya 1990, Greenberg 2001). While church attendance and community participation may lead to political participation,

⁵Jones-Correa (1998) offers an interesting account of participation by gender among immigrants.

⁶Miles (2004) has currently the most sophisticated analysis of the effect of these laws on turnout, and he finds no effect.

⁷Because the data come from public records, we measure socio-economic status with block-group level Census statistics and marriage based on household information in the voter file (e.g. many married voters can be identified by age, gender, and last name of individuals in the household).

such an explanation is complicated by endogeneity. We do not know whether community participation causes political participation or whether some other factor affects both, and we suspect the latter (see Atkinson and Fowler 2011).

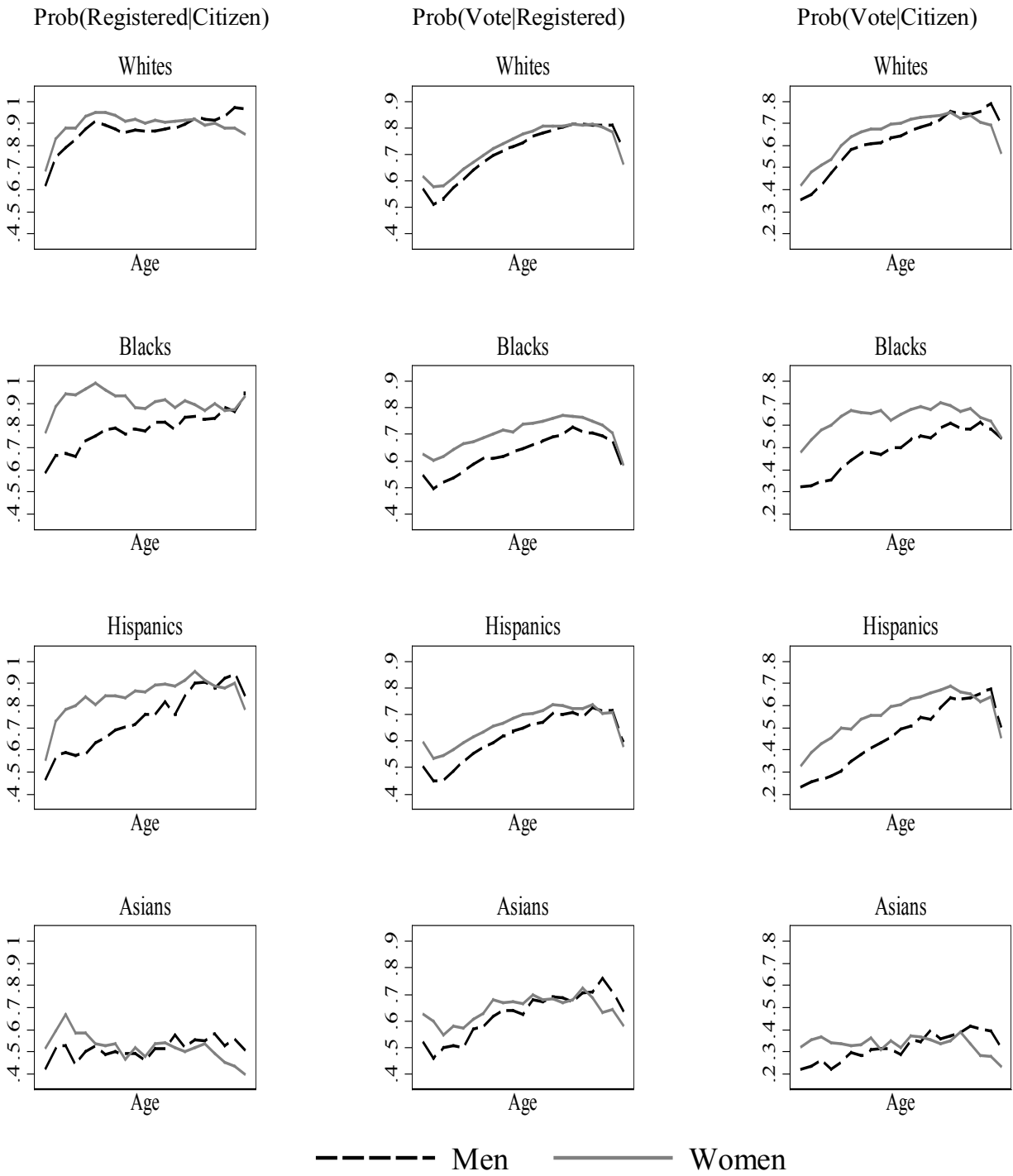
We are at an historical moment in political science. We can study political participation not by administering surveys but by observing the full population based on official records and consumer profiles. Upon doing so, the patterns of participation complicate the conclusions drawn from surveys in ways that challenge long-standing theories. Most importantly, a resources-based model of participation is out of synch with reality. African-American women in their 30s are not the demographic group thought to be at the apex of politically-relevant resources, as defined in seminal works of participation. But, they are registered at higher rates than any other demographic, vote at rates nearly 25 percentage points higher than Black men, and vote more than White men of similar age. This finding is just an initial step in a move to rethink the nature of political participation based not on what people say they do but on what they actually do.

Table 1: Mean Voting and Registration Rates by Race and Gender

Registration among Citizens		
	Male	Female
White	86.2%	89.1
Black	75.1	90.9
Hispanic	67.4	81.4
Asian	50.7	52.6
Voting Among Registrants		
	Male	Female
White	70.9	72.7
Black	61.0	69.0
Hispanic	59.1	64.3
Asian	61.0	64.1
Voting Among Citizens		
	Male	Female
White	61.1	64.7
Black	45.9	62.7
Hispanic	39.8	52.3
Asian	30.9	33.8
Obs. (Registrants in Sample)		
	Male	Female
White	614,771	681,196
Black	87,409	123,441
Hispanic	65,493	81,370
Asian	16,843	19,151

Note: Voting, registration, race and gender statistics are from a 1% sample of active voter registration records, drawn in 2010 from Catalist. Racial identification is predicted from a voter's first and last names and Census block group characteristics when not available on the public registration file. 'Voting Among Registrants' statistics are estimated entirely from Catalist's data. For citizen estimates, we start with the Census full population estimates by race, gender, and age. For each race-gender-age combination, we use the American Community Survey to estimate the number of individuals who are citizens. By taking each comparable race-gender-age group in the Catalist 1% sample of registration records, multiplying by 100 and dividing by the Census counts, we arrive at estimates of citizen registration rates and citizen vote rates.

Figure 1: Voting and Registration by Race, Gender, and Age



References

- Ansolabehere, Stephen and Eitan Hersh. July 14, 2010. "The Quality of Voter Registration Records: A State-by-State Analysis." Caltech/MIT Voting Technology Project Report.
- Atkinson, Matthew D. and Anthony Fowler. March 31-April 3, 2011. The Effect of Social Capital on Voter Turnout: Evidence from Saint's Day Fiestas in Mexico. Annual Meeting of the Midwest Political Science Association Chicago, IL: Unpublished Conference Paper.
- Burns, Nancy, Kay Lehman Schlozman and Sidney Verba. 2001. *The Private Roots of Public Action*. Cambridge: Harvard University Press.
- Greenberg, Anna. 2001. "Race, Religiosity, and the Women's Vote." *Feminist Studies* 22(3):59–82.
- Hornor, Louise L., ed. 1999. *Black Americans: A Statistical Sourcebook*. Palo Alto: Information Publications.
- Jones-Correa, Michael. 1998. "Different Paths: Gender, Immigration, and Political Participation." *International Migration Review* 32(2):326–349.
- Lien, Pei-Te. 1998. "Does the Gender Gap in Political Attitudes and Behaviors Vary Across Racial Groups?" *Political Research Quarterly* 51(4):869–894.
- Lincoln, C. Eric and Lawrence Mamiya. 1990. *The Black Church in the African American Experience*. Durham: Duke University Press.
- Miles, Thomas J. 2004. "Felon Disenfranchisement and Voter Turnout." *Journal of Legal Studies* 33(1):85–129.
- Montoya, Lisa J., Carol Hardy-Fanta and Sonia Garcia. 2000. "Latina Politics: Gender, Participation, and Leadership." *PS: Political Science and Politics* 33(3):555–561.

- Rosenstone, Steven J. and John Mark Hansen. 1993. *Mobilization, Participation, and Democracy in America*. New York: Macmillan.
- Seltzer, Richard A., Jody Newman and Melissa Leighton. 1997. *Sex as a Political Variable: Women as Candidates and Voters in U.S. Elections*. Boulder, CO: Lynne Rienner Publishers.
- Silver, Brian D., Barbara A. Anderson and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80(2):613–624.
- Verba, Sidney, Kay L. Schlozman and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge: Harvard University Press.
- Verba, Sidney, Nancy Burns and Kay Lehman Schlozman. 1997. "Knowing and Caring About Politics." *Journal of Politics* 59(4):1051–1072.
- Wiggins, Daphne C. 2005. *Righteous Content*. New York: New York University Press.