An Examination of Likely Voter Screens Using Panel Data

Steven S. Smith & Patrick D. Tucker

Presidential elections may occur on a four year cycle, but presidential election polling has no offseason. For example, with respect to the 2012 presidential election reputable organizations began releasing two-candidate general election polls as early as mid-March 2009[1]. The forthcoming 2016 presidential election inspired organizations to begin even earlier; the same organization conducted its first study in late January 2013[2]. While pollsters conduct surveys at a seemingly constant rate, their methods are quite dynamic. Specifically, conductors' decisions on whom to include in their reported samples vary across time and institution. Perhaps the most significant difference of opinion exists concerning when to consider whether likely or registered voters are the appropriate individuals to estimate the those citizens actually turning out for the general election contest. The purpose of this paper is to examine how well likely voter screens predict actual turnout behavior. Whereas other studies focus on cross-sectional samples of citizens to determine the strength of voter screens, we employ panel data to determine if the information provided by common screens varies during the presidential election campaign. To our knowledge panel data have yet to be used when evaluating the accuracy of the likely voter screen. By examining individuals over an extended period of time, we add to the understanding of polling samples by identifying who the likely voter screens correctly and incorrectly identify as potential voters.

Registered Voters

For all but one American state, registration is a prerequisite to vote. Hence, its predictive power of election day behavior is clear: those who have already taken the time to register showed some interest in voting in an election (at a certain point). Yet, registration status for many citizens may be an artifact of retired political involvement, rather than an indicator of current political engagement. Nonetheless, multiple polling organizations rely heavily on the registration status of survey respondents

1 Public Policy Polling, "2012 Matches Close" March 18, 2009.
2 Public Policy Polling, "Clinton, Rubio lead primary contests." February 7, 2013.

when constructing their samples.[3] One of the main reasons for this reliance is that historically, registered voters have a high probability of showing up to the polls (Crespi 1988). While a registered individual is clearly more likely to vote in any given election than a non-registered individual, the predictive power of registration is not constant across all electoral cycles. In fact, evidence exists that during presidential elections when media devote and the general population pay a great deal of attention to the campaign, registered voter screens almost completely identify those people most likely to participate (Mitofsky 1980). Even in midterm or state elections, where a fraction of the voting age and registered population turn out, the registered voter screen greatly decreases the error in predicting which citizens turn out (Zukin, Crespi 1988).

One of the greatest difficulties in using the registered voter screen is its validity. Registration, like voting, is a social norm. As such, the survey respondent may feel pressure to report that she maintains the required status, even when she does not. Due to differing state laws, respondents may have moved from one state to the another; incorrectly believing their previous status carries over to their new residence. Respondents may not be aware of any laws necessary to register to vote, believing such activity is a citizen granted right; as a result, they could incorrectly report. Whatever the reason, Traugott (1985) finds that over-reporting is common and can be as large as twelve percentage points. Combined with over-reporting in the aggregate, underreporting may occur at the individual level. Since the 1970s, many barriers for registration have dissipated, making the process much easier, and as a result, much less memorable (Highton 2004). Abelson, Loftus, and Greenwald (1992) suggests that in addition to not remembering a political behavior, respondents may be "telescoping" into the future their intention to act. That is, if they plan to register or vote, they will respond with the affirmative, even if they have not do so at the time of survey fielding. Subsequently, it may be the case that individuals

3 For example: http://www.pewresearch.org/2012/08/29/ask-the-expert-determining-who-is-a-likely-voter/

simply put little value in their registration and do not remember. A common approach to reduce the amount of survey respondent error is to eliminate the “don't know” option from the self-reported registration question or mark all such responses as “no.” Crespi, notes however, that very little evidence has found such a strategy to prove effective. Another option is to include in the question prompt that individuals who are not registered do, in fact, exist. Such an addition may reduce the social pressure placed upon the respondent to conform to the societal norm. Crespi, laments, however, that this tactic is rarely employed.

Unlike many social science problems plagued by survey error, the variable of interest here is of the public record. Hence, the ability exists to verify registration, putting its measurement somewhat less at the mercy of individual reporting. For this reason, many pollsters will solely rely upon registered voter lists with computerized lists of phone numbers and partisan identification labels (Crespi 1988, 70). Still, such methods are not without pitfalls. Crespi notes that through his interviews with the nations main polling organizations, many states perform poor purging of their non-registered voters, while others have inconsistent maintenance of phone numbers. Other lists are inaccurate or out of date. As registration laws have liberalized, the deadline for registration to vote in the upcoming election has been delayed. Close or heavily scrutinized campaigns that intensify as the election cycle moves along could attract more citizens to register later. If this group of late-registers is in anyway statistically different from the group registered at the beginning of the campaign, early polls could contain a significant degree of bias with respect to the decision to turn out as well as the election's predicted outcome.

Likely Voters?

If all citizens voted, the necessity to construct an accurate likely voter poll would not exist: pollsters would need only properly compile a statistically powerful random sample of the population to produce reliability. Setting the threshold of civic participation lower, if all registered citizens voted, few problems would exist for creating strong voter screens. The only problem that would need addressing would be the of standardization of registered voting records. Unfortunately for the purpose of this study, voting is neither mandatory nor a consensus. For this reason, pollsters must determine which members of the population are most likely to vote based on some additional measurements.

While it is the goal of most surveys to obtain a sample that is most representative of the national population, the sample reported by election polls has a much different goal: it needs to capture a sample that best estimates the voting population. Were the population of non-voters a random sample of the country, the likely voter screens would be unnecessary (Gerber and Green 2010). Yet, as numerous studies show those who do not vote have different electoral preferences than those who do show up to the polls (Highton and Wolfinger 2001, Shaffer 1982, Citrin et al. 2003, etc.). For this reason, accurate polling demands difficult decisions on whom to include in a reported sample.

To be sure, changes in the sample of reported polls affect the potential outcomes generated by high-profile national organizations. Erikson, Panagopoulos, and Wlezien (2004) find that a sudden shift in Gallup's polling methodology erroneously provided a double digit swing in favor of George W. Bush. Such a shift occurred independent of any significant change in preferences at the aggregate level.

One of the most common ways to screen for voters is to use a battery of questions about one's interest in the current election campaign and past history of voting. Murray, Riley, and Scime (2009) provide an excellent overview of some of political science's contributions in determining

Gallup Likely Voters: “The Gold Standard”

No name is more associated with presidential polling than Gallup. For over sixty years the organization's likely voter screen has been used in the days preceding presidential elections to determine who can be identified as having a reasonable probability for showing up to the polls. Within popular media, their reputation is solid; while lamenting the state of likely voter screens as a whole, chief NBC news analyst Chuck Todd concedes that they are "the gold standard of polling."[4] Even with political science research, Gallup Polls reported under the likely voter screen have been found to be relatively accurate. During the 2000 election they were found to be one of the most accurate predictors of vote, while their final 2004 poll released in conjunction with *The New York Times* and *CNN* was found to have the least amount of bias when compared to other major national firms (Traugott 2001, 2005). Most recently, however, the polls have come under fire. First, popular and left leaning media sources point to numerous instances when it appears that the likely voter method unfairly favors Republican candidates (Memmott 2004). The main reason for this perceived Republican bias is that the thresholds established by the polling firm (discussed below) are too strict in their classification of likely voters. Thus, the less transient (and relatedly, youthful) Americans have difficulty in attaining likely voter classification. Since age corresponds with Republican voting behavior, it is possible that this method could act as a cheerleader for Republican candidates and electorates in closely contested elections, increasing that party's turnout. Secondly, Gallup's likely voter screens tend to place emphasis on traditional forms of voting; that is, by polling place. As a result, those who vote by mail, which is quickly becoming more common among the electorate, have a higher probability of exclusion from the polling sample (Berinsky 2005). Giving some anecdotal credence to these complaints is the evaluation of all major polls by *The New York Times* (save those by the paper itself). In their analysis they found Gallup to actually have provided the most biased and incorrect results for the popular vote in the general election: 7.5 percentage points in favor of Mitt Romney.[5]

Although this snapshot of the Gallup poll's failings is in our rear view mirror most recently, one cannot deny the salience of the firms polls. For this reason, it is a good reference point for beginning an

---

4 http://www.huffingtonpost.com/2008/07/29/gallups-likely-voters-pol_n_115623.html "Gallup's Likely Voters Poll: A Snapshot or a Hypothesis?" *The Huffington Post* 7/29/08.

5 538 blog, Nate Silver.

analysis of likely voter screens. According to their website's description of the screening process, the goal of the firm is to “winnow down national adult or registered voter surveys to a subset of respondents who are most representative of the likely voter electorate.”[6] Over the more than sixty years of its existence, the Gallup survey researchers have developed and validated a series of seven questions that gauge the interest of each respondent in the forthcoming national election. In addition to thought provided, the questions cover past voting behavior and basic intention to vote (See Appendix for the exact wording of Gallup Question).

Answering each question with a response that Gallup determines to be demonstrative of a likely voter will award each individual one point. Providing a response that indicates a lower propensity to vote will provide the surveyed person with a zero for that question. The threshold for including one in the Gallup likely voter pool is six point or higher. To clarify, this means that a person cannot exhibit two out of 7 poor voting behaviors. Gallup justifies this threshold by noting that many individuals over-report their intention to vote. Adding a high threshold simply cuts down on the large number of over-reporting by non-voters. In an effort to keep up with the changing voting laws, Gallup will also ask individuals when they plan to vote. If that date has already occurred, the respondent will receive a score of seven, regardless of all other answers.

To optimize their likely voter sample, Gallup adjusts scores in some very basic ways. First, those who are not registered or say they do not plan to vote are immediately awarded a score of zero, independent of other reported behaviors. This choice means they will automatically not be included in the reported polling numbers and that all Gallup likely voter polls consist only of those who report being a registered voter. Perhaps in response to a common criticism, Gallup attempts to account for new voters by adjusting the overall score for teenagers and young adults who would not have been

6 All of this information will come from the same website.

eligible to vote during the previous presidential election cycle. For these citizens, a score of four or higher will result in inclusion to the likely voter sample.

Methods and Data

The majority of the data for our analysis is drawn from the American Panel Survey (TAPS). TAPS is a monthly online survey of about 2000 people. Panelists were recruited as a national probability sample with an addressed-based sampling frame in the fall of 2011 by Knowledge Networks for the Weidenbaum Center at Washington University. Individuals without internet access were provided a laptop and internet service at the expense of the Weidenbaum Center. In a typical month, over 1500 of the panelists complete the online survey. More technical information about the survey is available at taps.wustl.edu. Survey data for this paper come from the months of October 2011 to October 2012. The maximum number of panelists by month was in June with about 1700, while the minimum number occurred in December 2011 (1213).

To construct our own likely voter screen, we mimicked the efforts of Gallup. A seven screen battery was designed in order to gauge the likelihood of our panelists' voting proclivities. For the most part, this series of questions is identical to that used by the Gallup polling firm during the 2012 presidential campaign. Two key differences exist. First, since our data collection began in the fall of 2011 and Gallup does not report on likely voters until the waning months of the presidential election campaign, our questions are modeled upon the set supplied for the 2010 Congressional elections. For that year's national campaign, following the question about intention to vote, each respondent who identified with the affirmative was asked to provide their certainty of their behavior (“Absolutely Certain,” “Fairly Certain,” and “Not Certain” were the responses). Only those providing an “Absolutely Certain” response were assigned a “1” for the purposes of the aggregate score A similar, but not identical, question was substituted in 2012. This addition included more of a free-response format, asking those who reported they would vote to classify their certainty of voting on a scale from one to ten, with one being the least certain. Only those respondents who claimed higher than a seven on the ten-point scale were given a positive score for their response.

The second difference between our set of likely voter questions and those of Gallup stems also from our basis of the 2010 wave as a point of reference. The question about most recent voting in that national election asked about participation in the most recent election. Following Gallup's lead, we asked if the panelist voted in the fall of 2010. For the 2012 screen, however, Gallup adjusted their questionnaire to inquire about voting behavior in the most recent *presidential election,* 2008's contest, rather than the Congressional elections of 2010.

These differences could be problematic if we hope to measure the strength of the Gallup likely voter screen. Although the first question difference measures the same variable, certainty of response, the switch from a three category to a ten-point response set could bias our outcomes. With only three responses, panelists may feel more pressure to choose the absolutely certain option even if they do not really feel that way. On the surface, greater problems exist with the second difference. Here, we are measuring a factually different variable from that of our reference. As numerous studies show, midterm-election is a strong indicator of regular voting habits, but there are large portions of the population that limit their participation to general election contests. While it may be the case that including this measurement provides a very strong prediction among those who vote on a regular basis, it could also be the case that its inclusion will force too strict a standard on our replicated likely voter score, meaning we could be under-predicting turnout. Nevertheless, we choose to include the 2010 measurement because it is the closest variable we have to the question asked in 2012.

Voter registration is key to our understanding of determining who is a likely voter. As discussed above, two options exist for determining survey respondents' registration status: self-report and verification through Secretaries of State record systems. The former's benefits are found in its feasibility and cost: there is little added difficulty in asking our panelists to report their registration status. Unfortunately, we did not ask a voter registration in each month that included the likely-voter battery. As a result such an option is not available. Instead, we rely upon verified voter registration information provided by the data firm Catalist. Catalist cross-referenced mailing address information of our panelists with all fifty states' voter registration records. With this information, the firm was able to match each individual in our sample to a corresponding name in their given state of residence. Not only is Catalist able to provide the registration status of an individual, it is also able to provide the date of registration. Hence, Catalist provided registration dates have two key advantages over the self-report method. First, they do not suffer from over-reporting of status due to societal norms or respondent survey errors. Second, they provide an accurate estimate that allows us to chart increases of registration over the course of the campaign. Thus, our measurements of registration are not fixed based upon the timing of when our panel was asked the likely voter battery.

With likely voter status and registration, the third key element of this study is the actual activity of voting. Once again, as discussed above, voting measures are subject to large amounts of mis-reporting when dependent upon survey methods. Over- or under-reporting is a common source of survey error. TAPS is not immune from such activity. For example, of those responding to the November 2012 panel, over 90 percent reported voting. Meanwhile, the United States Elections Project reports that only 58.2 percent of voting eligible public participated in the presidential contest.[7] With respect to this problem, Catalist efforts may prove useful, as well. In addition to scouring voter registration records, they are also able to provide a certain level of confidence that each individual in our panel turned out to vote. Those with whom there is a two-thirds or greater probability were kept in the sample, while all those with lesser levels of confidence were dropped. Catalist reports both election

7 http://elections.gmu.edu/Turnout_2012G.html

participation and method of voting (such as by polling place, absentee, or by mail). To compare the verified vote information to our self-reported vote totals, we find that of those who were in the panel at any time during the period of November 2011 to November 2012, only about three-fourths turned out to vote. When compared with those who reported turning out, of those who reported that they voted, verification could only be provided for those individuals eighty-four percent of the panel. Corroborating the theory that a large segment of the population overreport participation. It should also be noted that our estimates may still be over-reporting turnout due to the two-thirds confidence threshold we have established. The possibility exists that those people for whom Catalist lacks confidence in their respective match chose not to vote.

Likely voting questions were asked of the panelists in seven months (December 2011, January, March, April, June, September, and October 2012). Panel structures often suffer from varying levels of respondent attrition; not every member who begins the process chooses to participate in every wave. TAPS is no different. Additionally, our survey introduced 329 new panelists in June 2012. Just under 700 panelists (689) participated in each wave and met or exceeded the Catalist threshold for certainty of election day behavior.

Survey data were also gathered to understand the demographics of our panelists. Upon entering the panel, each interviewee is required to answer a profile survey. Variables extracted from this wave administered in both November 2011 and June 2012 include race, sex, household income, symbolic ideology, age, and years of education. The exact wording and response categories for these variables can be found in the Appendix.

Multivariate Models

To better understand the two elements of the polling puzzle discussed here, we have developed two sets of statistical models. First, it is of interest to determine the efficiency of the likely voter battery across time. For various months, we predict turnout among our sample using each of the Gallup indicator questions. If each questions provides information about the likelihood to vote, each question should produce a significant, positive effect on the average individual's decision. In addition to investigating effects of the Gallup questions, we also include a control for the panelist's registration status for the first of the month of the given wave. Included also in the model is a battery of control variables that previous research suggests affect the likelihood of voting: sex, years of education, age, race, and income. While not often considered an influence on the choice to turn out, we also include a control for symbolic ideology. It is commonly held that the likely voter sample is distinct from the registered voter screen in that Republican voters are disproportionately included (Erickson, Panagopoulos, and Wlezien 2004). For this reason, we include such a control in all of this paper's models. The outcome variable of this model, clearly, is the decision to turn out. As discussed above, this variable is drawn from the Catalist verified vote database. Since the vote variable is dichotomous (1=verified voted, 0=otherwise), we chose to run a logit model that can be summarized as shown below:

P(y=1)=b0+**b2*Gallup**+**b2*Controls**+b3*Registered

We are also interested in possible biases that the Gallup likely voter battery may produce in the sample of prospective voters. Traditionally, biases in polling screens are thought of as unidirectional. That is, most discussion of Gallup's demographical faults are drawn with respect to over-identifying certain groups as non-voters, even when large portions of such populations do turn out to the polls. As discussed above, Gallup acknowledges such an under-prediction in its past and has attempted to correct for this bias by instituting its age adjustments. To analyze bias in the direction of under-prediction, or the disproportionate "screening out" of the likely voter pool, we composed a model subsetted of those who were predicted to not vote when we applied the Gallup likely voter model to our panel data (those panelists who received lower than 6 on the Gallup likely voter battery). The outcome variable in this model is misclassification. That is, this variable is coded as "1" if the respondent voted and "0" if they did not vote. Hence, voting indicates the Gallup model incorrectly predicted our panelist as a non-voter. In order to best investigate biases in the screen, we restrict our analysis to the wave closest to the election (October 2012). The right hand side variables here include the same various demographical variables that were included as control in the model above. It can be summarized as shown in the equation below:

P(Voted=1)= b0+ **b1*Demographics+ e**

In addition to estimating misclassification in the traditional direction, our data also provide us the ability to investigate misclassification in the opposite direction. That is, we are interested if there are certain individuals that Gallup "over-predicts" as voters, or "screens in" to the likely voter sample based on their responses to the given questions. To estimate such biases, we subset our data based on the application of the Gallup score, including only those individuals who were categorized as likely voters (those scoring a 6 or above on the Gallup method). Again, the model only examines classifications of the October survey and the same set of controls will be used as the explanatory variables. As before, the outcome variable is again the misclassification of the panelist. Yet, in this instance the variable is coded a "1" if the respondent did not vote in November, as verified by Catalist, and "0" if they did vote. The model can be summarized as:

P(misclassification=1)=b0+b1*Demographics+e

Results

Improvement as Election Day Nears?

To determine if the gallup battery of questions improves as election day nears, we first disaggregated the questions to determine the individual accuracy of each. Table 1 displays the results for three selected months of those panelists who participated in every wave of the American Panel Survey. Weighted percentages provided in the cells represent the proportion of panelists who responded to each question with a corresponding "1" response on the Gallup score and were found to have voted by Catalist and the percentage of those individuals who responded with a corresponding “0” answer and were found to not vote. For example, the 73.99% in the Voted in 2010 cell for the month of March 2012 demonstrates that nearly three-fourths of the panel claim to have voted in 2010 and voted in 2012 or claim to have not voted in 2010 and abstained from voting in the presidential election of 2012. The first noticeable finding from this table is that each Gallup likely voter indicator is a reasonably strong predictor of November behavior. Nearly all of the questions have predict whether the panelist will show up to the polls with about seventy-five percent accuracy or better. Such a finding shows that Gallup chooses indicators that are highly correlated with the act of voting.

Yet, while the relatively high accuracy across all questions appears to provide validity to the measurement, further investigation of the disaggregated questions provides some skepticism to the Gallup model. Intuition and conventional wisdom suggest that likely voter identity becomes clearer to the analyst as November nears. That is, the questions provided should be better predictors of vote in October than they are in March. Among our consistent panelists, however, we find that not all of the questions improve in their accuracy as time passes. In fact, two questions, previous behavior in one's district or precinct and voting behavior in the previous election, are *worse* predictors in October, when Gallup employs the likely voter screen, than in March. Proponents of the Gallup model could point to significant improvements in prediction that we find with respect to the knowing the polling location question, voting often, certainty of voting, and intention to vote questions over the seven month period. To be sure, nearly all of these questions, with the exception of voting often, should see improvement as the election approaches. For example, intention and certainty of voting will become clearer to voters and non-voters as they obtain information about the candidates and the campaign (Hillygus and Jackman 2003). While we do not deny this possibility, we cannot ignore that while these all of these indicators' improve from the spring surveys, some still perform worse in the fall than they do in the month preceding the election.

Finally, and perhaps most importantly, with respect to Table 1, our application of the Gallup screen to the panel finds that the Gallup composite score monotonically improves across the three sampled months. Thus, as the organization argues, their definition of "likely" is better suited for the fall than for previous months. Such back-patting for Gallup is short lived, however; within each month provided in table 1, the registered voter screen sample still outperforms the Gallup likely voter screen in turnout accuracy. In fact, some of the questions, such as intention and certainty of voting, individually outpace composite score across all waves.

One could argue that the analysis in Table 1 is a poor application of the likely voter screen since those who remain within the panel across all waves are hardly representative of the regular Gallup sample. We concede this point. In order to better simulate the Gallup method, we also investigated the trends in disaggregated questions with respect to pooled samples. That is, Table 2 replicates the processes of Table 1 for all those individuals who participated in any of the same three waves. Once again, all questions appear to have a reasonably high level of correlation with actual turnout; no question provides less than seventy-percent accuracy of voting behavior. Using these different samples of respondents, we find that the accuracy of all individual questions increases over the course of 2012. Such improvements appear to hold monotonically across all months. Improvement associated with the composite score is almost identical to that found among our panel. Another similarity to the panel sample can be found in relation to the registered voter-likely voter screen comparison: as with the previous group, the registered voter screen provides a more accurately predicted sample of voters than the likely voter screen. Additionally, both samples of individuals find that in its current form, many of the questions included in the Gallup likely voter battery actually take away from the predictive power of the voter screen. When the screen is employed in October, the likely voter score is a worse predictor than not only the registered voter screen, but also the intention to vote and certainty of voting questions. Hence, we must question the wisdom of how pollsters create the Gallup likely voter screen and their decision on when to use it.

Finally, to further illustrate the choice of whether and when to switch from a registered and likely voter screen, Figure 1 adds to our skepticism. Across all waves in which the Gallup questions were asked, the registered voter screen proves to be a better predictor of voting. To be clear, the figure shows that Gallup's wisdom in poll reporting is correct early on in the campaign: in the spring and summer, registered voter screens are more accurate than their likely one. Yet, when the campaign enters its final months, and presumably polling accuracy is more highly valued, the polling firm switches to an inferior method of prediction. It appears that Gallup's polling could do no worse in its representative ability if it simply stuck to its registered voter screen. Although it is essentially a certainty that not all registrants will actually vote, such a simple screen is a better method than the current usage of likely voting questions.

If we are to better understand the value that each question of the Gallup likely voter model provides, we must continue with our disaggregated approach. The results of our first set of models can be found in Table 3. Once again, we sample the months of March, June, and October. The first column seems to confirm most of the polling world's instincts. Very little evidence of a statistically reliable positive effect on the likelihood of voting exists. Only with respect to the vote intention question does a significant effect arise. Further bolstering Gallup's practices, a stronger, statistically reliable effect exists for the panelist's registration status.

As the pooled samples move towards November, the magnitude of the indicators' effects increases for these variables. We find that estimated influence increases by more than fifty percent for registered voters over the seven months. Vote intention's significant effect more than doubles once the panel reaches October, further confirming Gallup's claims. Time indicates that individual questions gain reliability and influence as we move closer to election day. For example, October shows that the voting often question and certainty of voting question become statistically significant. While these results all support what Gallup argues, the results of these models also find no evidence that half of the likely voter indicators have any significant effect on turn out. Thinking about the election, knowing one's polling location, previous election behavior, and precinct voting all prove to be indistinguishable from a null effect on voting. Perhaps even more disturbing, although statistically insignificant, some of these effects are shown to be negative. Thus, these results suggest that it may best serve the purposes of polling firms to re-evaluate the seven question model in that more than half of the battery provide little information about voting when controlling for other factors.

Screening Bias?

Following the example of Best and Krueger (2012) we can further explore the predictive power of the Gallup likely voter screen to the actual voting population with respect to certain groups. As stated, they found that Gallup under-predicted the proportion of the voting electorate that was young, female, and non-white. To begin this analysis, we subsetted the population of our October panel using these three demographic characteristics. Out of our own interests, we also did so for those living in households earning less than $40,000 per year. Once again, we applied the Gallup likely voter battery to each of these groups. Those among the group that achieved a score of 6 or better are identified as likely voters. Across all five groups, this application predicts that turnout will be between forty and sixty percent. Yet, when we examine the actual turnout rates using the Catalist verified vote, we find that the Gallup likely voter model under-predicts behavior by about ten to fifteen percentage points for all categories. Confidence intervals at the 95% level demonstrate that such differences are statistically reliable. Hence, these initial cursory steps would seem to confirm the criticism that Gallup's likely voter model has a screening bias against certain groups, particularly those most associated with high levels of voting for the Democratic party.

While such graphical analysis provides clarity of interpretation, it is somewhat simplistic; drawing inferences without adequate controls should be met with great caution. Hence, it is best to look for misclassification among certain groups with a multivariate analysis. The first set of misclassifications of interest are those in the traditional direction: citizens are predicted not to vote by a likely voter screen, but they end up voting. In order to most efficiently observe this phenomenon, we subsetted our October panel based on the Gallup model: included are only those panelists scoring a 6 or above with respect to the battery. Since all are predicted to vote, the outcome variable is an incorrect prediction: coded "1" if the individual does not vote and "0" if the individual does vote. The main explanatory variables include demographics that were gathered from each person upon entry into the panel.

The four categories included in the analysis above are employed as explanatory variables in the model. While simplified to a dichotomous variable in the graphical analysis, age is operationalized continuously. Household income is measured using fifteen categories; within this model it is treated as a continuous variable. Race is included as a categorical variable with a baseline category of "White, Non-Hispanic." Finally, sex is included as a dummy variable with values of "1" signifying "Female." Two other explanatory variables not employed in the Krueger and Best study may prove to have an effect on misclassifications. The less educated may be less likely to provide the necessary responses to the likely voting battery, independent of their actual intention to vote; thus, such a trait could lead to biases of being screened out of the system. Here, we test for such an effect by including a continuous years of education variable. Likewise, political blog posts and pundits often complain that likely voter models systematically favor conservatives over liberals. At the same time, conservative talking heads proclaim that sample selection processes underestimate the number of conservative voters. As a result, we choose to include a symbolic ideology variable. Here, the panelists' symbolic ideology is measured on a seven point scale with "1" being "very liberal" and "7" representing those identifying as "very conservative."

Since the dependent variable is a dichotomous model, we ran a binomial logit model. The first column of Table 4 displays the results. Age provides a statistically significant effect on the likelihood of a misclassification. That is, as a panelist gets older, she is less likely to be misclassified compared to a younger panelist. Although the strength of interpretation is somewhat muted, this finding follows in the tradition of conventional wisdom; young people are more likely to be misclassified than older people using a likely voter screen. This model would also seem to confirm our initial suspicions concerning biases and race. Two ethnic categories possess strong, reliable effects on the probability of a misclassification. If the panelist is black or Hispanic, they are less likely to be incorrectly predicted to vote than their white counterparts. More clearly, white individuals are more likely to be incorrectly classified, ceteris paribus, than ethnic minorities. Since white individuals are over-included in the likely voter screen, it is the case that non-white minorities are under-included. Finally, we find support for our hypothesis that lower income individuals are more likely to be misclassified as non-voters than the wealthy. The model suggests that income is negatively related to a misclassification. As income increases, the probability of being incorrectly labeled a likely voter decreases. With respect to our other hypotheses, we find no evidence that a screening bias exists for sex, education, or symbolic ideology in this direction.

One of the strengths of our panel data is that we can examine not only those who are incorrectly predicted to vote, but we can also examine those who are incorrectly predicted not to vote. To identify such individuals, we identified those who scored less than a 6 on the Gallup battery and subsetted accordingly. Whereas the previous dependent variable was coded a "1" for not voting, here a misclassification occurs when the subject votes; in effect, the dependent variable is "1" for voting and "0" otherwise. This second model is shown in the second column of Table 2. All other hypotheses and explanatory variables remain in the model.

Once again, we find some (although less pronounced) effect for race in this direction. Compared to white panelists, Hispanic identity reduces the likelihood of being incorrectly predicted not to vote compared to white panelists. Still, compared to the many other ethnic groups, evidence does not suggest that whites are disproportionately included. Although sex had no statistical effect on misclassification in the previous direction, here we find that being female has a statistically significant effect on the likelihood of being incorrectly predicted to vote. Such a result confirms the report of Krueger and Best that compared to men, women are disproportionately excluded from the polling sample when Gallup uses its likely voter screen. In a similar vein, this direction finds a positive effect for years of education. As the number of education years increases, the likelihood of the individual being incorrectly predicted to not vote decreases. This result relates to our suspicions in that among those predicted to vote, higher education seems to lead to higher incidences of error. Such a finding runs counter to our expectations. Finally, we find a significant negative effect with respect to household income. This figure suggests that as income increases, the likelihood of being incorrectly predicted to abstain decreases. More directly, lower income individuals are more likely to be excluded from the likely voter pool than those with higher incomes. In other areas we find little evidence of effects on the likelihood of misclassification in the exclusionary direction.

Conclusion and Discussion

Our research in this project was motivated by two common beliefs about likely voter screens: their effectiveness changes as the campaign progresses and they produce samples that are statistically different from the actual voting population. The first motivation was found to be true, but with a serious caveat: while likely voter screens in October *are* much more accurate in predicting voting behavior than in March, they are *less* accurate than a simple registered voter screen. A registered voter screen, found to be most accurate in the early days of the campaign, when applied to those participating in The American Panel Survey was more accurate than the likely voter screen during all months for which data were available. To emphasize, our exploration concludes that in the campaigns final months, Gallup willingly switches to a less accurate model of voter prediction.

Following this line of analysis, we find that many of the questions in the Gallup likely voter battery provide little to no predictive information about actual turnout. In fact, some of the questions actually reduce the predictive power of the voter screen; that is, as the campaign approaches November, individual questions are less reliable than they were seven months previous. Multivariate multidimensional regression confirms this finding. At the same time, certain questions provide a great deal of information about likelihood of voting. For example, registration status, intent to vote, certainty of voting, and voting history are all excellent predictors of actual turnout.

Finally, we take up the charge that Gallup systematically biases against traditionally marginalized groups with respect to voting. Comparison of means and multivariate analysis all confirm that the likely voter screen over-predicts the proportion of men, whites, and higher income individuals who will vote, while the under-predicts the turnout of women, ethnic minorities, the young, and lower income households.

From these results we propose that polling firms such as Gallup thoroughly re-evaluate their current methods of likely voter selection. Although not explored here, we do not doubt that the underestimation of the groups shown is directly related to the firm's infamous over-prediction in the direction of Mitt Romney in the fall of 2012. To begin, such changes should focus on the reduction of the likely voter threshold, specifically reducing the number of questions in the screening process. We show that many of the questions provide little information. Furthermore, individual questions may bias against groups that are particularly mobile, such as the young and lower income. If presidential polling is to maintain its command of the American public's attention, it must improve its most basic methods. Hopefully, this paper serves as origin for such change.