The Electronic Ballot Box:
A Rational Voting Model and the Internet\textsuperscript{1}

\textbf{Jason Gainous}  
Department of Political Science  
University of Florida  
234 Anderson Hall  
Gainesville, FL 32611  
gainous@polisci.ufl.edu

\textbf{Kevin Wagner}  
Department of Political Science  
University of Florida  
234 Anderson Hall  
Gainesville, FL 32611  
kwagner@polisci.ufl.edu.

November 15, 2002

\textsuperscript{1}Each author contributed equally to this study. Thanks to Jeff Gill, Michael Martinez, Ken Webber, and Beth Rosenson for reviewing earlier drafts.
Abstract

This research creates a theoretical framework for understanding the effect of Internet voting on the electorate. Based on standard Downsian rational choice voting theory, we claim that Internet voting lowers the cost of voting for certain voting demographics based upon race, age, and income. We further contend that this electoral advantage may crystallize the growing turnout disparity between demographic groups. The theory is tested using Bayesian inferential methods with data from the Internet turnout in the 2000 Arizona Democratic Presidential Primary combined with demographic data obtained from the 2000 Census. Our findings lend support for the theory that the Internet provides an electoral bias towards white voters, younger voters, and to the more affluent.
During the 2000 Presidential Primary election, the Arizona Democratic Party brought politics and technology together by allowing party members to use remote Internet access to vote in the Democratic Presidential Primary. This was the first time a binding election for public office in the United States has ever been held on the Internet (Solop 2000). The implications of Internet voting are just beginning to be discussed (Alvarez and Nagler 2001), but largely only in empirical terms without a substantive explanatory model (Solop 2001). The purpose of this article is to propose the use of a utility model (Downs 1957) to predict the demographic impact on voter turnout with the introduction of Internet voting, along with a Bayesian analysis of the aggregate Arizona voting data to assess the value of cost-benefit analysis on turnout and voting demographics in the digital age.

1 Digital Democracy: The Early Impact

Digital Democracy is changing the nature of the political landscape. By integrating technology into the functions of government, we have altered, in fundamental ways, the power dynamics of our political system. An early warning of this impact was displayed in January 1996, when then House Speaker Thomas Foley was defeated in part through the efforts of a political action committee that was organized almost entirely on the Internet (Browning 1996). Though most notable in electoral forums, technology growth is affecting the way that government goes about its tasks in almost every aspect. From filing taxes to obtaining federal documents, the manner by which the government interacts with the people is changing rapidly.

Predicting the nature of this change and the possible impact of new technology such as the Internet on the American political system is difficult, though scholars have tried (See, e.g. Selnow 1998; Davis 1999; Browning 1996). Early works have explored the influence of the Internet on news gathering, lobbying, campaigning, and even participation (Davis 1999). Political scientists have already surmised that the implications of the Internet are substantial and have, in a short time, changed the manner in which campaigns are conducted (Davis 1999). Though a relatively recent development, a significant web presence has become critical to an effective campaign (Selnow 1998). Candidates are using the Internet to bypass
traditional campaign methods to reach voters as well as raise campaign funds (Browning 1996). In the 2000 Republican primaries, Senator John McCain repeatedly advertised his web presence and used the Internet to turn his surprise victory in New Hampshire into a fundraising juggernaut (Salant 2000). McCain raised $4 million over the Internet (A.P. 2000).

Despite the importance of these areas of focus, it is becoming more evident that the most significant impact of the Internet may be as part of the electoral machinery itself. E-voting, or remote voting over the Internet, is no longer a fiction. In the private sector, companies such as Chevron and Lucent Technologies have utilized e-voting as a means to elect union officers (Nathan 2000). Universities such as Stanford University and the University of Florida have used it for student government elections. States such as Iowa and Washington have experimented with Internet ballots. But the largest breakthrough was originated by the Arizona Democratic Party, which initiated an entirely new political discourse by allowing its members to vote over the Internet in the 2000 Democratic Presidential Primary.

Shortly after the election, early surveys suggested some possible impacts of the Internet voting. Arizona had a sharp rise in electoral participation with the rate of turnout increasing 723% percent between 1996 and 2000 (Solop 2001). Solop noted the demographic impact of the election primarily with regard to age, education and religion (Solop 2001). Though empirically interesting, the early work did not attempt to create a theoretical framework for exploring the possibly significant change in the voting population should e-voting be adopted in other jurisdictions. We create such a framework, and test it empirically. The proposition herein is that the Internet may crystallize a substantial technological gap in the voting electorate that could exacerbate cleavages already present within the electorate. More generally, the Internet may very well change the magnitude of previously existing voter turnout cleavages based on income, race, and age and thereby force a substantial change in how politicians campaign and govern. Though the scope of the primary is in itself too small to make sweeping generalities about the national electorate, it should supply indicators of the potential exacerbation of preexisting cleavages in the electorate.
E-voting presents a new and potentially significant variable in the analysis of voting behavior, but it does not change the fundamental structure of such analysis. Voting and the political system are unchanged, but the use of the Internet changes the means and operation of the system at a fundamental level by adding a new procedural lens to the equation. From a social choice perspective, E-voting does not change the “calculus of voting” as developed by Downs (1957) and later adopted with a slightly different emphasis by Riker and Ordeshook (1968) (see also, Tullock 1967; Barry 1970). The base Downsian equation is as represented as follows:

\[ R = PB - C \]  

\( R \) denotes the net reward or utility in voting. The likelihood of voting is a function of the probability that the vote will effect the outcome \( P \) multiplied by the differential benefit of the voters’ candidate \( B \) prevailing, and then subtracting the cost of voting \( C \). The theory contends that the voter will abstain from voting if \( R < 0 \). Also note that if \( R > 0 \), the voter may still abstain because there may be other competing activities that produce a higher \( R \) for that given point in time. Assuming that \( C > 0 \), then \( PB \) must be \( > C \) for the voter to vote, and since we know that \( P \) is very small, then \( B \) must be very large for a vote to occur (Gill and Gainous 2002). Obviously it would take an unrealistically large value for \( B \) to overcome this small value of \( P \). Riker and Ordeshook (1968) attempt to account for this problem by formalizing the additional Downsian satisfaction parameter \( D \). This \( D \) is added to the equation and represents the personal satisfaction/utility that a citizen receives from the act of voting regardless of the actual outcome of the election.

The revised equation is as follows:

\[ R = PB + D - C \]  

Riker and Ordeshook (1968) contend that this \( D \) value is not constant across individuals. Therefore some will vote and some will abstain, depending on whether \( D_i \) sufficiently overcomes \( C_i \) for individual \( i \). \( D \) is said to consist of various social and psychological subfactors.
such as: citizen duty, prestige, guilt relief, and a sense of continuing the political system (Gill and Gainous 2002). In short, where the cost to vote $C$ outweighs the potential utility to effect the outcome and the perceived benefit plus the various subfactors mentioned above, the voter is not likely to vote.

The equation represents the base decision-making of a voter grounded on a basic notion of utility maximization with the inclusion of the less quantifiable variable of Duty $D$, which acts largely as an error or stochastic term. As becomes readily apparent, anything that affects any of the variables in the equation can change the nature of the voting electorate. While the positive component of the model is a function of several factors, the central drag on participation is but one variable: Cost of Voting $C$. Downs (1957) refers to this component as the opportunity cost of voting based on the time and resources spent in preparing for, and participating in, the election.

More recently, this cost has been more succinctly described as the cost of registration, decision-making and turnout at the polls (Aldrich 1995). Though each of these elements are valid costs associated with participation, the cost of physically turning out at the polls seems to be the most significant. Research on the National Voter Registration Act (“Motor Voter”) has illustrated that reduced registration cost alone appears insufficient to bring voters to the polls (Martinez and Hill 1999).

The initial question presented herein is whether one could change voter participation by simply changing the nature of the most substantive elements in the cost of voting. By voting on the Internet from home and eliminating entirely the cost of turning out at the polls, the drag on participation should significantly decline. Further, it follows that those who vote by the Internet also would benefit from the information-gathering ability offered by the Internet enabling them to reduce the resources spent in preparing to vote and learning about the candidates and their issues (see, e.g. Browning 1996; Davis 1999). As a result, the Internet should significantly reduce the magnitude of the drag variable in the equation.\(^1\)

\(^1\)As noted above registration cost alone does not appear to be a significant drag on turnout, though this issue is unsettled (Martinez and Hill 1999; Wolfinger and Rosenstone 1980). Nonetheless, the Internet may also be used to reduce the cost of registration as well. In Arizona the voters in the primary were contacted through mail with Internet voting instructions (Solop 2001). In the future much of the registration could be performed with little or no cost online.
If the cost of voting $C$ is reduced, turnout should change.

Turnout, if measured in magnitude alone, is interesting and previous work has illustrated that turnout did substantially increase in Arizona during the Internet Primary. The rate of turnout in the 2000 election increased 723% from the 1996 election (Solop 2001). But such a finding is not the end of the analysis. The more significant issue is based not on magnitude of turnout, but in the possible change within the voting electorate itself. More directly, the issue is whether the Internet changes the identity of the voter at the poll or the likelihood of different groups voting. Generally, the Internet should lower in absolute terms the true cost of voting in the voting equation, but it is not uniform in its application. The ability to use the Internet is based on the availability of a computer connected to the Internet combined with sufficient knowledge of its use. One only gets the benefit of a reduced cost of voting when these threshold factors are met. Without access and knowledge, the cost of voting equation remains unchanged. Hence, the benefit and impact of the Internet is built on a divided foundation of have and have-nots (Davis 1999). The Internet will not disenfranchise anyone, but rather will disadvantage non-users as voting participants (Davis 1999).

3 Voting And The Digital Divide - Maximizing The Inequality Effect

Under the present voting system, The United States has managed to combine declining turnout with increasingly unbalanced voting electorates that over-represent the upper classes (Burnham 1987; Leighley and Nagler 1992; Rosenstone and Hansen 1993). We propose that the use of e-voting will not only be consistent with this trend, but will, with increasing impact, crystallize the distinction by changing the voting incentive and costs in the voting system along an increasingly apparent cleavage. The end result will not be a question of over-representation, but rather of significant power growth in favored groups. Where “Motor Voter” was expected to help balance the electorate through increased registration of under-represented groups (Martinez and Hill 1999), e-voting will emphasize a technological and class gap through the more significant reduction of the turnout cost in these favored groups.
It is a remarkable reversal with considerable more effectiveness. E-voting targets the key cost of turnout rather than the less substantial element of registration cost.

The impact of e-voting is predicated on an understanding of the Digital Divide. This is a vernacular reference to the disparity between those able to use informational technology such as the Internet, and those who cannot. Though the United States has more computers than any other nation, in 2000 the Department of Commerce issued a report revealing that only 41 percent of U.S. households have Internet access (DOC 2000). Ethnicity appears to be a key factor, with high rates of Internet access for Whites (46.1%) and Asian American & Pacific Islander (56.8%). There are greatly reduced rates of access for Black (23.5%) and Hispanic (23.6%) households (DOC 2000).

Similar divisions can be seen along educational lines, with high access rates for college graduates (65%) but minimal rates (11.7%) for households headed by persons with less than a high school education. Additionally, rural areas (38.9%) have lower rates than urban areas (42.3%). Age is also a factor. The U.S. Census reveals a significant distinction between homes with a resident 49 and under (63%) and those with only persons 50 and over (37%). Sex does not appear to be a significant issue, though women (51%) are slightly ahead of men (49%) (DOC 2000).

Internet access, however, is available for persons outside the home through schools and libraries. For the purpose of the cost of voting analysis, there is no significant difference between travelling to the local polling location versus the local library. Further, only (0.5%) of Americans use computers at Community Centers (DOC 2000). Hence, the cost of voting reduction is largely through remote access. Conceivably however, the analysis may change should Internet voting be made more available through remote units more widely and readily available than traditional polling stations.

If the data are consistent with the utility model, there should be an observable movement in favor of groups with technology. Based on divisions identified by the data from the United States Department of Commerce on access and use of technology, the results from the Arizona primary based on the cost benefit of the Internet should be reflected in economic, racial and age demographics. Aware of criticism, based on the digital divide, Arizona did provide Internet voting through computer polling stations (Solop 2001). As this does not provide
any substantial cost benefit related to turnout, this analysis is focused solely on the remote Internet voting data.

4 Hypotheses

To provide support for the theoretical argument concerning the cost of voting several hypotheses are tested. These hypotheses are as follows:

\( H_1 \) - there is a positive relationship between income and remote Internet turnout.

\( H_2 \) - there is a negative relationship between minority status and remote Internet turnout.

\( H_3 \) - there is a negative relationship between age and remote Internet turnout.

All of these hypotheses are directed at providing support for the assertion that allowing e-voting will magnify the advantages of certain demographic groups in the electorate.

5 Data and Methods

Our model tests the probable impact of certain demographic characteristics on Internet voting. The data used in this study are a compilation of turnout results from the 2000 Arizona Democratic Presidential Preference Primary and demographic data obtained from the Bureau of the U.S. Census. The turnout results were obtained from the Arizona Democratic Party and are available at the party’s official website (www.azdem.org/breakdown.html). These data are formatted as total turnout, remote Internet turnout, mail-in turnout, polling place Internet turnout, and polling place paper turnout by county. The data provides for 15 cases based on the number of counties in Arizona. The demographic data are also by county, which enables the compilation of each of these sources. These data are aggregate.

The small sample size makes conventional frequentist estimations problematic. This concern is addressed by the use of a Bayesian construct which involves the creation of posterior subjective probability distributions of model parameters by combining prior information
(non-sample data) with sample data (Gill 2002; Western and Jackman 1994). Bayesian methods have been advocated for research involving a small number of observations and cases involving nonstochastic data. Small data sets that produce fragile statistical inference in frequentist models are more effectively handled by the Bayesian approach through the incorporation of prior information in the estimation. This methodology provides a solution to problems associated with restricted samples and collinearity (Western and Jackman 1994). In a previous study, Western and Jackman (1994) use Bayesian inferential methods with a sample size of 20.

The use of a Bayesian model requires that we incorporate a prior distribution for each of our variables. Prior distributions are descriptions of relative probabilities that are usually based on previous research and knowledge developed in the discipline (Gill 2002). Our non-sample information used for the creation of the Bayesian prior is drawn from previous research on the Arizona primary (Solop 2001). These data include three telephone surveys: a 1200 person cross-sectional survey of Arizona adults, a 1200 person survey of registered Democrats in Arizona and a post-election panel study with 783 registered Democrats of which 318 participated in the Democratic primary (Solop 2001).

The data from the Solop study provide an important foundation for our own data, which is based on county level aggregate sample. The limited sample size would prevent us from reaching the commonly accepted threshold for statistical inference in a frequentist model (Western and Jackman 1994). But the combination of the prior information with the present sample allows for asymptotic inference. The data provide the base parameters for the prior distribution of the explanatory variables. These parameters are set forth in more detail below.

Aside from dealing with the limitations of the frequentist model, the Bayesian approach also allows us to make probability statements about the parameters of the model. The outcome is not a point estimate as in a frequentist model, but rather a probability distribution that is typically described by a mean or mode and some type of measurement of dispersion (Jackman 2000). While a brief discussion and review of the Bayesian approach is undertaken in this paper, a complete review of the applicability of Bayesian methodology is beyond the scope of this paper (for a complete review see Gill 2002; Pollard 1986; Lee 1989; Western
To create the posterior for our model, we used Gibbs sampling method of Markov Chain Monte Carlo (MCMC). The Gibbs sampler integrates the posterior identity to create probability inferences for each of the unknown parameters in the model. The model used for this study is based on a standard ordinary least squares (OLS) regression that is supplemented by the prior information provided in the Solop study. None of the Gauss-Markov assumptions are violated, and therefore, ordinary least-squares regression is the best linear unbiased estimator of the relationship. The relationship between the outcome variable percent remote Internet turnout by county and the explanatory variables is linear with a standard model as follows:

\[ Y = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k + \epsilon_i \]

This linear model is combined with the prior distribution MCMC methods. The posterior is determined by:

\[ \pi(\mu, \sigma^2) \propto L(\mu, \sigma^2|x)p(\mu\sigma^2)p(\sigma^2) = L(\mu, \sigma^2|x) \cdot p(\mu\sigma^2) \]

The model was operationalized through the use of WinBugs, a program for Bayesian analysis of complex statistical models using MCMC techniques. WinBUGS allows models to be described as Doodles (graphical representations of models) or through a text-based description. For use in WinBugs, our model was in its base text form expressed as follows:

\[
\begin{align*}
\text{for (i in 1: N)} \\
\epsilon[i] & \sim \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \Rightarrow (0, 0, \tau) \\
\lambda[i] & \leftarrow \theta_0 + \beta_1 \ast x_1[i] + \ldots + \beta_k \ast x_k[i] + \epsilon[i] \\
Y[i] & \sim \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \Rightarrow (\lambda[i], \tau)
\end{align*}
\]

The model is based on regression analysis where \( \epsilon \) is the error term with a prior normal distribution. The \( \beta \) represents the explanatory variables each with its own prior distribution based upon information obtained in the Solop surveys. The sample data for each explanatory variable are represented with an \( x \). The outcome variable is continuous and distributed
normally (refer to Figure 1) with a precision distributed \( \tau \) which is discussed in greater detail below.

Figure 1 demonstrates the normality of the outcome variable remote Internet turnout. Quantile-Comparison plots are an effective graphical means of comparing ordered data against the corresponding quantiles of a reference distribution, the normal distribution in this case (Fox 2002). The data are linear and they conform to the normal distribution within a 95% pointwise confidence envelope. This envelope is represented by the dashed-lines that band the linearity indicator.

The nature and function of MCMC techniques is complex and addressed more thoroughly in other scholarly works. For further investigation into both the modeling and MCMC techniques utilized in this paper refer to Gill (2002) and Jackman (2000).
6 Variable Description and Operationalization

6.1 Outcome Variable

The outcome variable is Remote Internet Turnout and is measured as a percentage of the turnout by county. This variable was initially coded as a raw turnout number, but for the purposes of this research was coded, along with most of the explanatory variables, as a percentage. The variable is measured with the following simple equation: Remote Internet Turnout by County/Total Population Turnout by County. The mean remote Internet turnout is 0.35 and the median is 0.34. These percentage values for each county were multiplied by 100 to convert them into whole numbers, making the interpretation more easily understandable.

The prior information for this variable was determined based on the outcome of the Solop (2001) study and parameterized by a normal distribution. Solop uses a binary logistic regression and assumes asymptotic normality. Our model uses a hyper-parameter $\tau$ based on a tightly defined gamma distribution for the precision in this variable allowing for a random effects model. The gamma distribution is appropriate because it is the conjugate of the normal distribution and our likelihood function utilizes the normal distribution. Instead of having a constant precision, the model will draw on the gamma distribution for that parameter:

$$\tau \frac{x^{\gamma} - 1 e^{-x}}{\Gamma(\gamma)} \Rightarrow (0.1, 0.1) \quad \forall x \geq 0, \gamma > 0 \quad (6.1)$$

6.2 Explanatory Variables

Income per capita was measured using the Census Bureau data by individual unit of population, or by county. The mean income per capita is $16,942 and the median is $16,569. Income was divided by 1000 to make the interpretation of the output more clear. In our data, income and education are correlated closely making the inclusion of both variables redundant ($r = .685, p < 0.01$). To make each individual case (county) relative to every other respective case, the demographic indicators of race and age were converted to percentages by county. White, African American, and Hispanic were all conformed to percentages using
the following simple formula: \( \frac{\text{Total Population by Race per County}}{\text{Total Population per County}} \). These percentages were then multiplied by 100 to make the interpretation more coherent. The mean and median percentage of whites by county was 0.70 and 0.74, African Americans by county was 0.02 and 0.01, and Hispanics by county was 0.27 and 0.25. The Native American population, which is significant in Arizona, was not included in the model to create a baseline variable.

The percent adolescent was measured and inferred based on a raw number of persons under 18 with a mean of 91,130 and a median of 33,425 converted with the same formula: \( \frac{\text{Total Population by Age per County}}{\text{Total Population per County}} \). This variable is intended to measure the proportional presence of families which serves as an indicator of young and middle aged potential voters that are more likely to use the internet to vote as the theory herein contends. Population over 65 was also measured as a ratio using the same formula as above. Senior citizens had a mean and median of 0.15 and 0.14, respectively. All of these demographic variables were converted to whole numbers by multiplying them by 100 to make them consistent with the outcome variable. These averages do not differ much from the national average, except perhaps for the lower percentage of African Americans and higher percentage of Hispanics (DOC 2000). This does not threaten the generalizability of our model because it is primarily an argument of socio-economic class structure that is associated with race. The combination of African Americans with Hispanics in Arizona approaches the combination of their national averages (DOC 2000).

The prior distributions on these variables are all based on a normal distribution with the mean centered on the prior information derived from the Solop (2001) study. Since our prior is based upon a single survey instrument instead of a compilation of research, we operationalize our uncertainty through the variance placed on the prior distributions. Larger variances reflect greater uncertainty. Precision is inversely related to variance, hence, lower precision provides for greater variance. The precision on the variables measuring income, race, and age are set at .0001 to provide a fairly flat normal so as to indicate relative uncertainty in the model. As there is only one inferential study of the impact of Internet voting, using a tight or well-defined prior would presume a state of knowledge that is not representative of the current level of scholarship. Based on the Solop study of Internet voting,
the means of our prior distributions representing the rate of participation in the Internet Primary were set as follows: Percent White (.47), Percent Hispanic (.44), and Percent over 65 (.33). Two of the variables used in the model (Percent Black and Percent Adolescent) are not measured in previous studies so we supply a flat prior with a mean at zero to indicate the lack of information for these variables. All of these percentages were multiplied by 100 to make them consistent with the explanatory variables in our model. The income per capita prior mean was set to the same value as the income indicator ($16,492) because there are not sufficient indicators of income per capita in the Solup study. This parameter was also divided by 1000 to be consistent with the income indicator in the likelihood function. The Census data used are a comprehensive sample and therefore there is no reason to expect income per capita significantly deviate from the Census value.

7 Findings

The results of our Bayesian Inference are presented in Table 1. The table describes the posterior distributions of the variables in our model.

<table>
<thead>
<tr>
<th>Demographic Variables</th>
<th>Mean Effects</th>
<th>Standard Deviation</th>
<th>95% Credible Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>0.86</td>
<td>0.51</td>
<td>-0.14</td>
</tr>
<tr>
<td>Percent Black</td>
<td>1.25</td>
<td>0.94</td>
<td>-0.62</td>
</tr>
<tr>
<td>Percent White</td>
<td>0.31</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>Percent Hispanic</td>
<td>-0.19</td>
<td>0.08</td>
<td>-0.35</td>
</tr>
<tr>
<td>Percent Over 65</td>
<td>-0.67</td>
<td>0.32</td>
<td>-1.30</td>
</tr>
<tr>
<td>Percent Adolescent</td>
<td>0.42</td>
<td>0.15</td>
<td>0.12</td>
</tr>
</tbody>
</table>


While frequentist models provide point references, Bayesian output is the description of probability distributions. The distributions can be described by their means and standard deviation. The mean is the tangent of the distribution or the most likely effect of the explanatory variable on the outcome variable. The credible interval describes the bounds
Figure 2: Posterior Distributions for Explanatory Variables

Figure 2 is a visual representation of the probability distributions described in Table 1. They were created by generating 10,000 random values given the mean and standard deviation of the posterior distributions. All of the probability densities have a normal distribution. The power of the predictive value of each variable is more easily understood by this representation as it is clear when the highest densities of the distribution fall away from zero.

From these distributions we can make the following statements. There is sufficient evidence to support these hypotheses: $H_1$- there is a positive relationship between income and remote Internet turnout, and $H_3$- there is a negative relationship between older voters and remote Internet turnout. There is partial support for $H_2$- that there is a negative re-
relationship between minority populations and remote Internet turnout. While there is some conflicting evidence in the results regarding race, these discrepancies can be accounted for by the fact that the data are at the aggregate level. This assertion is further explored in the interpretation of each of the explanatory effects explained by the model. This interpretation is followed by a diagnostic analysis of the model.

7.1 Income

The model indicates support for the hypothesized positive relationship between income and remote Internet voting. Though not large in absolute terms, the increase in income when controlling for the other demographic variables predicts an increase in remote Internet voting. While zero is bounded in the 95% credible, near 90% (-0.14 to 1.87) of the distribution falls on the positive side of zero. This indicates that the true unknown parameter while controlling for the other demographic indicators within the model is positive around approximately 90% of the time. This finding contradicts that of the Solop (2001) study, wherein his binary logistical regression indicated that income was not significant.

This result is consistent with what one would expect from one of the major indicators of growing class division as expressed in $H_1$. Figure 2 shows the posterior distribution in more stark visual terms. The graph illustrates the positive relationship between income and remote turnout based on a largely normal distribution. The mean of the distribution is positive with relatively small tails. The relationship meets the expectation that larger incomes are predictors of Internet use and ultimately Internet voting.

7.2 Race

The predictive value of race is mixed in our model. As can be seen in Table 1, there is a reliable relationship between Hispanic voters and remote Internet turnout while holding all other explanatory variables constant. The 95% credible interval indicates a negative relationship between Hispanic voters and remote Internet turnout of fairly significant magnitude (-0.35 to -0.03) while controlling for all other explanatory variables.

The expected positive relationship exists for white voters. The distribution of the white
vote indicates a positive relationship with a 95% credible interval bounded away from zero
(0.04 to 0.59). The distribution is well defined and the mean of the distribution is significantly
positive (0.31). Refer to Figure 2 for visual evidence of the relationship between white
population as well as Hispanic population with remote Internet turnout. The evidence
provides support for $H_2$

The test results of the relationship between African American population and remote
Internet turnout are not what we expected. The relationship appears to be positive with
the mean well above zero at 1.25. But this relationship is unconvincing because while the
bulk of the probability distribution is positive, a significant portion of the of the parameter
probability is distributed negatively (-0.62 to 3.14), and this interval is proportionally
larger than all of the other intervals in the model. Nonetheless, even the limited positive
relationship is counter to what we would expect. This may be based upon the demographic
makeup of Arizona. The African American population in Arizona is relatively small and the
variance between each county is extremely low. Therefore, there is not enough change in
the explanatory value, case by case, to produce a significant change in the outcome variable.
In addition, the highest of the African American populations are in urban areas that also
have high income ($r = .651, p < .01$), so the variance is absorbed by the income explanatory
variable in the opposite direction.

### 7.3 Age

Age is consistently viewed as having predictive value for Internet and computer use and
that view is supported by the most recent Census indicating that homes headed by persons
under 50 are more likely to own a computer (DOC 2000). Our model is consistent with
these indicators and shows a strong negative relationship between age and Internet voting.
For persons over 65 the negative relationship with Internet voting is apparent from the
distribution (refer to figure 2). The credible interval is negative and bounded away from
zero illustrating 95% probability that the true unknown parameter of age influence on remote
Internet turnout is between -1.30 and -0.02. This illustrates that age is a powerful predictor
of Internet use and voting. The distribution is centered far from zero at -0.67.
Figure 3 provides a visual representation of the posterior distributions for both age indicators. This figure was generated using 1,000 random values given the mean and standard deviation of each posterior respectively. The negative relationship of seniors and the positive relationship of adolescence is apparent. Almost all of the values for each are distributed away from zero in the opposite direction, demonstrating the power of age as a predictor of internet turnout.

The credible interval generated by the measure of adolescence further exhibits the strong negative relationship of age to e-voting (0.12 to 0.72). While the distribution of this variable indicates a positive relationship, the true nature of the relationship to age is inverse because the variable is measured as a percentage of youth. As the presence of adolescents increases, so does the presence of parents which are generally younger than 65. Hence, as age decreases e-voting increases. This posterior distribution has a mean and standard deviation of 0.42 and 0.15 respectively. The age indicators can also be thought of as a propensity for the use of technology. In the present context, there is a strong gap between younger families with technology and the older populations who are lagging behind in the increasingly complex digital age (Census 2000). The younger generation is likely to have less problems with technology driven advances. County level aggregate data do not allow for a more detailed view of the youth effect, though the greater penetration of technology to younger Americans is worth further study. It is unclear from these data that age will be a constant division. It may be that some of the age related divisions are not so much a matter of life cycle, but are
founded on generational differences that will decline over time as the more technically savvy generations reach retirement.

8 Diagnostic

The reliability of a posterior generated through MCMC is based upon an assumption of convergence. More succinctly, the posterior described must come after the Markov chains have found the most dense regions of the sample space for each indicator. MCMC samplers will usually get to a desired distribution, though it may take many iterations to achieve convergence. There is no perfect way to make this determination, though there are several different tests of convergence. We use the widely accepted Gelman and Rubin (1992) test. Gelman and Rubin’s convergence diagnostic uses a measure of within chain variance and between chain variance. A score of 1.2 or less is considered acceptable. In testing our model, all of the relevant posteriors achieved a Gelman and Rubin score of near 1.0.

9 Discussion

The data and findings herein are but a preliminary look at what will likely be a growing field related to the impact of the Internet on voting and behavior. As e-voting voting becomes more prevalent, more samples will become available to test rational voting models and the cleavages suggested herein. The goal of this paper was not to prove the existence of the digital divide, though its presence is clear in the data. The research herein is being used to suggest a rational utility model as a means to view the impact of the Internet on the voting electorate. Though this is certainly not the only impact, the declining cost of voting may by maximizing the voting strength of certain groups reshape voting trends and call for a new calculus in creating likely voter models. Though the data suggests that there will be no sharp change in who votes, it does indicate a magnification of the economic cleavage that has already been observed. This is a confirmation and crystallization of a significant voter

---

trend. The effect of an acceleration of this shift can be explored in multiple policy areas and in the decline in power of some demographic groups. Further, some of the change may well be new. If the trend continues to show a small over 65 presence on the net combined with growing youth usage, the power dynamic of interest groups such as the American Association of Retired Persons could change.

But ultimately, the nature of the Internet and its future relationship to voting behavior is unsettled. As noted earlier, the availability as well as use of the technology is changing, and thus, the cost of voting model is not static with relation to the electorate. In the end, with larger efforts to distribute technology and the growth of homes with Internet access there may well be some equity in the impact of the Internet across demographic barriers and groups. Yet, it is just as clear that at present there is a sharp contrast in the initial benefit of the technology. In the infancy of e-voting, this may have a significant impact on how campaigns are managed and how voters impact the government.

The Internet has become an integral part of American Society. The time has come to explore that impact in terms of voting behavior and change. This paper presents one view on how to measure that impact. As more voting moves to the Internet, additional work needs to be done to understand what will likely be one of the most significant changes in the politics and voting in the next decade.

References


