

The Electronic Ballot Box:
A Rational Voting Model and the Internet¹

Jason Gainous

Department of Political Science
University of Florida
234 Anderson Hall
Gainesville, FL 32611
gainous@polisci.ufl.edu
352-374-4595

Kevin Wagner

Department of Political Science
University of Florida
234 Anderson Hall
Gainesville, FL 32611
kwagner@polisci.ufl.edu
561-252-1794

May 17, 2004

¹Each author contributed equally to this study.

The Electronic Ballot Box: A Rational Voting Model and the Internet

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 merged 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.

Keywords: Internet voting, electoral participation, rational choice, Bayesian inference

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 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). Though we cannot attribute all of the increase to the role of the Internet in the election, it is difficult to dismiss the impact of the new voting procedure given this dramatic rise. 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 generalizations about the national electorate, it does supply indicators of the potential exacerbation of preexisting cleavages in the electorate.

2 Cost of Voting and The Internet

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 $R < 0$. Obviously it would take an

unrealistically large value for B to overcome the often small value attributed to 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. It 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 Gaiours 2002). The revised equation is as follows: $R = PB + D - C$. In short, where the cost to vote C outweighs the potential utility to affect the outcome and the perceived benefit plus the various subfactors mentioned above, the voter is not likely to vote.

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 (Browning 1996; Davis 1999). As a result, the Internet should significantly reduce the magnitude of the drag variable in the equation.¹ 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 (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 any differentiation in 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).

¹As 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.

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 Data and Methods

4.1 Data

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 whole numbers representing turnout by county across several categories including total turnout, remote Internet turnout, mail-in turnout, polling place Internet turnout, and polling place paper turnout. The data provides for 15 cases based on the number of counties in Arizona. The demographic data are also by county and coded as whole numbers, which enables the compilation of each of these sources. These data are aggregate. This small sample size is addressed by using a Bayesian approach that will be further explained in the methods section.

4.1.1 Outcome Variable

The outcome variable is Remote Internet Turnout and is measured as a percentage of the total turnout by county. Internet voting constituted 35 percent of the total turnout. As previously mentioned, this variable was initially coded as a whole number by the Arizona Democratic Party, 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 Turnout by County*.

4.1.2 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 of all counties is \$16,942. Income was divided by 1000 to make the interpretation of the output more intuitive. 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: *Total Population by Race per County/Total Population per County*. These percentages were then multiplied by 100 to make the interpretation more intuitive. The mean percentage of whites by county was 0.70, African Americans by county was 0.02, and Hispanics by county was 0.27. 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: *(Total Population by Age per County/Total Population per County) * 100*. 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. 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).

4.2 Methods

4.2.1 The Bayesian Way

The small sample size makes conventional Null Hypotheses Significance Tests (*NHST*) 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 knowledge) with sample data (Gill 2002; Western and Jackman 1994). Using a small N sample is no longer controversial. 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 (Robert 2001). In a previous study, Western and Jackman (1994) use Bayesian inferential methods with a sample size of 20.

As noted above, the Bayesian process of data analysis allows researchers to incorporate previous knowledge into a statistical model. This is operationalized through prior distributions which are descriptions of relative probabilities that are usually based on previous research and knowledge developed in the discipline (Gill 2002). In order to create our Bayesian model we included a prior distribution for each of our variables based on the findings of the Solop (2001) study. These prior distributions are described in greater detail below.

Aside from dealing with the limitations of conventional *NHST* models, 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 *NHST* model, but rather a probability dis-

tribution that is typically described by a mean or mode and some type of measurement of dispersion (Jackman 2000). In contrast to the *NHST* method of deciding strength of conclusions based on the magnitude of p-values, Bayesian inference presents evidence by simply summarizing the posterior distribution, and therefore there is no artificial decision based on the assumption of a true null hypothesis. Posterior summary is typically done with quantiles and probability statements such as the probability that the parameter of interest is less than/greater than some interesting constant, or the probability that this parameter occupies some region.

The posterior is created by joining the prior and likelihood with Bayes' Law:

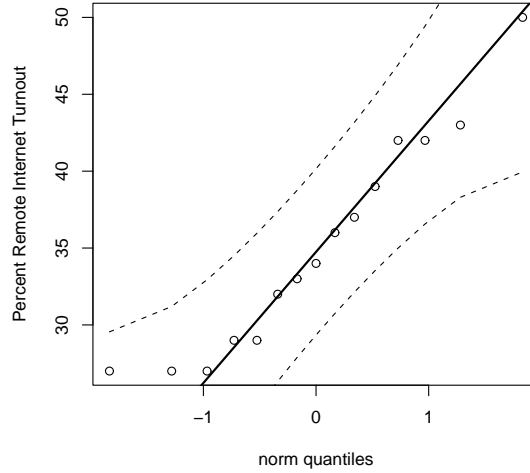
$$\pi(\boldsymbol{\beta}|\mathbf{X}) = \frac{p(\boldsymbol{\beta})L(\boldsymbol{\beta}|\mathbf{X})}{\int_{\Theta} p(\boldsymbol{\beta})L(\boldsymbol{\beta}|\mathbf{X})d\boldsymbol{\beta}} \quad (4.1)$$

to produce the posterior distribution, where $\int_{\Theta} p(\boldsymbol{\beta})L(\boldsymbol{\beta}|\mathbf{X})d\boldsymbol{\beta} = p(\mathbf{X})$. Here the $\pi()$ notation is used to distinguish the posterior distribution for $\boldsymbol{\beta}$ from the prior. The posterior distribution can be summarized by various $\boldsymbol{\beta}$ given the observed data, \mathbf{X} , and other information that is summarized in the prior distribution. 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; Robert 2001; Pollard 1986; Lee 1989; Western and Jackman 1994).

4.2.2 Modeling Remote Internet Turnout

Though Bayesian inference permits a wide range of modeling specifications, the design 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 lin-

Figure 1: QUANTILE-COMPARISON PLOT OF REMOTE INTERNET TURNOUT



ear unbiased estimator of the relationship. The association between the outcome variable percent remote Internet turnout by county and the explanatory variables is linear and is modeled as follows before integrating prior information:

$$Y = \alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \epsilon_{ik}$$

Further, the outcome variable is distributed normally. This is demonstrated through the use of a quantile-comparison plot. 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.

4.2.3 Inclusion of Prior Information

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 prior distributions on the explanatory 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. The variance is operationalized by what is called the *prior precision*. The 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 create a largely diffuse prior, 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.²

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 from the statistical results reported in the referenced study as follows: Percent White (0.47), Percent Hispanic (0.44), and Percent over 65 (0.33). These values represent the percentage of Internet voting derived from the the Solup survey. Solup factored income into several categories, and we selected the mean internet turnout expected by his highest category ($\geq \$75K$), setting the mean to

²Using flat priors can create pathologies in the resulting posterior distribution. For a detailed explanation see Hobert and Casella 1996.

(0.69). Two of the variables used in the model (Percent Black and Percent Adolescent) are not measured in previous studies so we supply a diffuse normal prior with a mean at zero to indicate the lack of information for these variables.

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.

4.2.4 Obtaining the Posterior Through Simulation

To create the distribution of the posterior for the parameters of interest, 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 underlying premise of this technique is that if an iterative chain of consecutive values can be setup carefully and run long enough, then *empirical* estimations of quantities of interest can be obtained from chain values. In order to estimate multi-dimensional probability structures such as our posterior distributions, we began a Markov Chain in the sample space and let it run until it settled into the correct distribution. From this distribution we were able to gather statistics including the means, variances, and quantiles from the simulated posterior.

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.³ The specifications for a Bayesian model are typically given in a stacked notation that sets forth the distribution assumptions of the priors and the likelihood:

$$\begin{aligned}
Y[i] &\sim \mathcal{N}(\lambda[i], \sigma^2), \\
\lambda[i] &= \beta_0 + \beta_1 x_1[i] + \dots + \beta_k x_k[i] + \epsilon[i] \\
\epsilon[i] &\sim \mathcal{N}(0.0, \tau) \\
\beta[i] &\sim \mathcal{N}(\beta_P, 0.0001)
\end{aligned} \tag{4.2}$$

for $i = 1 : n$.

Though the notation of the equation may be unfamiliar, the Bayesian approach is direct and open with all of the assumptions delineated in the equation itself. As indicated above, the outcome variable (Y) is assumed to be distributed normally around the systematic component with fixed variance. The systematic effect (λ) is based on a linear specification but includes a random effects term. The hyper-parameter τ is 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.

All of the coefficient estimates (β_P) are given diffuse normal priors but are centered around information obtained from the Solop surveys. The actual values used are explained in greater detail above. The explanatory variables are denoted x_i and are columns from the explanatory variable matrix \mathbf{X} .

³The specification of our model for use in WinBugs is set forth in the appendix

5 Findings

The results of our Bayesian Inference are presented in *Table 1*. The table describes the posterior distributions of the variables in our model. 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 of the distributional effects of the explanatory variables that serve as predictors of e-voting based on our sample and non-sample (prior) information. For the purpose of this paper, we have used a 95% credible interval.

Table 1: Explanatory Posterior Distributions of Internet Voting

Demographic Variables	Mean Effects	Std. Dev.	95% Credible Intervals	
Income	0.90	0.23	0.44	1.36
Percent Black	105.6	39.97	25.88	183.40
Percent White	29.65	6.22	17.40	41.83
Percent Hispanic	-18.35	3.69	-25.61	-11.04
Percent Over 65	-66.34	14.79	-95.37	-36.98
Percent Adolescent	39.91	8.10	24.45	56.23
Constant	0.92	2.21	3.57	3.90

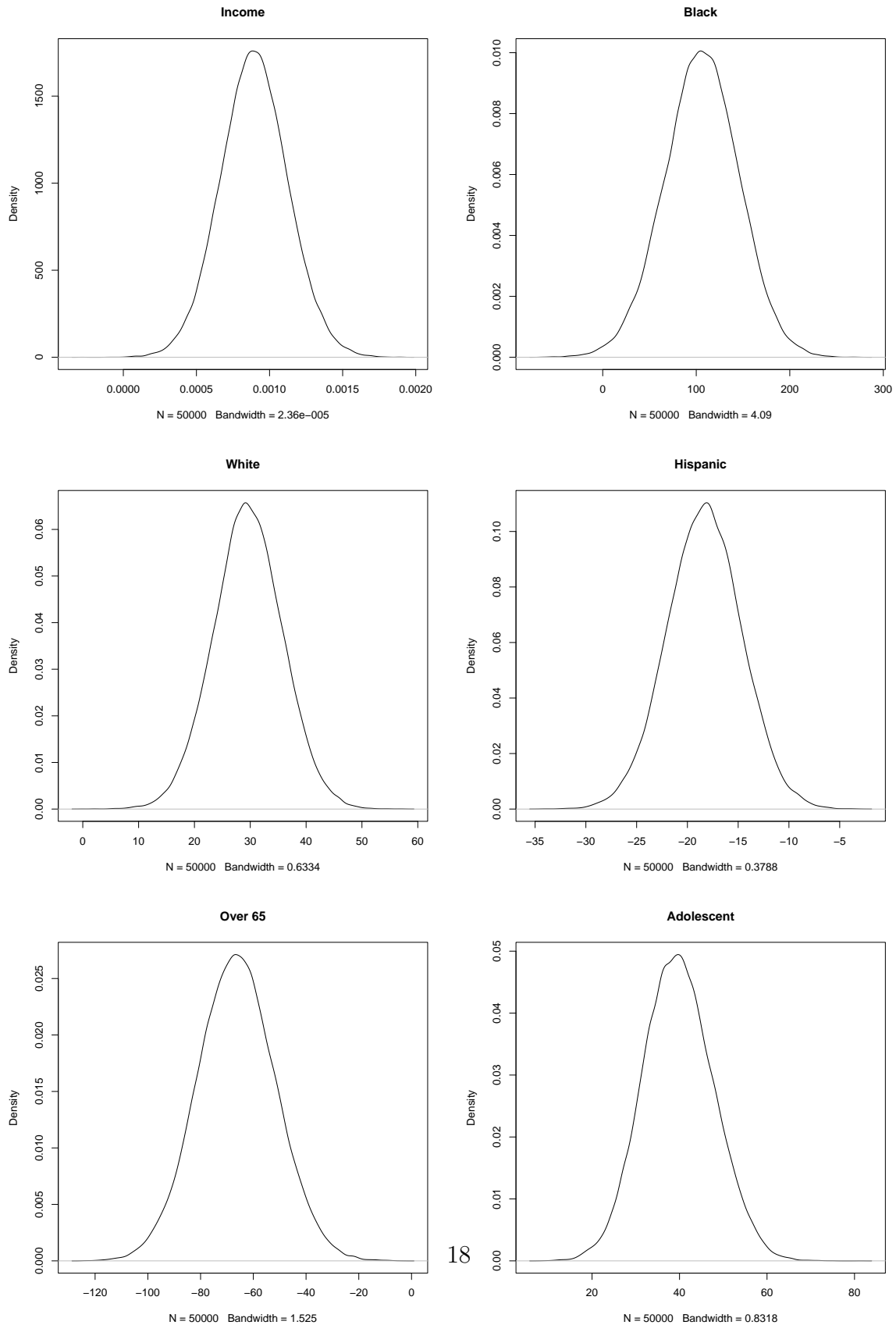
Posterior Standard Error of τ 0.9023

Source: www.azdem.org/breakdown.html, U.S. Census 2000, and Solop 2001 survey.

Note: The income parameters are multiplied by 1000 to make the values easier to interpret.

Figure 2 is a visual representation of the probability distributions described in *Table 1*. These distributions represent the Markov Chain Monte Carlo chain values for each of

Figure 2: Posterior Distributions of the Explanatory Variables



the explanatory variables. 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.

The posterior distributions offer sufficient evidence to provide support to the theory that Internet voting lowers the costs of voting for certain demographic groups. There is a positive relationship between income and remote Internet turnout, a negative relationship between age and remote Internet turnout, and a negative relationship between Hispanic population and remote Internet turnout. While there is some conflicting evidence in the results regarding race because of the positive relationship between black population and remote Internet turnout, 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.

5.1 *Income*

Though not large in absolute terms, as income increases remote Internet voting increases when controlling for the other demographic variables. Zero is not bounded in the 95% credible interval. This indicates that the true unknown parameter while controlling for the other demographic indicators within the model is positive around approximately 95% 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. *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 at the aggregate level are

predictors of Internet use and ultimately Internet voting.

5.2 *Race*

The model is consistent with the expected relationship between race and remote Internet turnout. As can be seen in Table 1, there is a negative relationship between Hispanic voters and remote Internet turnout while holding all other explanatory variables constant. The 95% credible interval indicates that this negative relationship between Hispanic voters and remote Internet turnout is reliable (-25.61 to -11.04). The expected relationship also exists for white voters. The distribution indicates a positive relationship with a 95% credible interval bounded away from zero (17.4 to 41.83). The distribution is well defined and the mean of the distribution is reliably positive. Refer to *Figure 2* for visual evidence of the relationship between white population as well as Hispanic population with remote Internet turnout.

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 105.6. 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 ($r = 0.651, p < 0.01$), so the variance is absorbed by the income explanatory variable in the opposite direction of what would be expected from low income families. It is simply a product of aggregate categories.

5.3 *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 -95.37 and -36.98. This illustrates that age is a powerful predictor of Internet use and voting. The distribution is centered far from zero at -66.34.

The credible interval generated by the measure of adolescence further exhibits the strong negative relationship of age to e-voting (24.45 and 56.23). 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 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 39.91 and 8.10 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.

5.4 Diagnostics

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 regions of highest density in 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 used the widely accepted Gelman and Rubin (1992) convergence diagnostic. This procedure measures the within chain variance and between chain variance for multiple chains with diffuse starting points. 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. We also confirmed convergence using the Geweke (1993) test, as well as by performing a visual inspection of the chains through the use of trace plots.⁴

6 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 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

⁴For a more detailed description of the Markov chain convergence tests refer to Cowles and Carlin (1996), Johnson (1996), Robert (1995), and Gelman and Rubin (1992).

suggest a rational utility model as a means to view the impact of e-voting on the 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, 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 of voting in the next decade.

Acknowledgements: Thanks to Jeff Gill, Michael Martinez, Ryan Bakker, R. Keith Gaddie, Terri Fine, Ken Webber, and Beth Rosenon for reviewing earlier drafts.

7 Appendix

For use in WinBugs, our model was in its base text form expressed as follows:

```
for (i in 1 : N) {
  epsilon[i]~dnorm(0.0, tau)

  lambda[i] <-      beta0 +
                    beta1*X1[i]  + beta2*X2[i]  + beta3*X3[i]  +
                    beta4*X4[i]  + beta5*X6[i]  + beta6*X7[i]  +
                    epsilon[i]

  Y[i] ~ dnorm(lambda[i], tau)}
```

References

- Aldrich, J. 1995. *Why Parties?* Chicago: University of Chicago Press.
- Aldrich, J. 1976. Some problems in testing two rational models of participation. *Am. J. Polit. Sci.* 20(4), 713-733.
- Alvarez, R.M., Nagler, J. 2001. The Likely Consequences of Internet Voting for Political Representation. *Loyola of Los Angeles Law Review* 34:3.
- Barry, B.M. 1970. *Sociologists Economists and Democracy*. London: Collier-Macmillan.
- Browning, G. 1996. *Electronic Democracy: Using The Internet To Influence American Politics*. Wilton: Pemberton Press.
- Burnham, W.D. 1987. The Turnout Problem. In: Reichley, A.J. (Ed.), *Elections American Style*. Washington DC: Brookings Institution.
- Conventions Boost Online Fundraising. 2000. Associated Press. 18 August.
- Cowles, M.K., Bradley, P.C. 1996. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association* 91(434): 883-904.
- Davis, R. 1999. *The Web of Politics: The Internet's Impact on American Political Systems*. Oxford: Oxford University Press.

- Downs, R.A. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.
- Fox, J. 1957. *An R and S-plus Companion to Applied Regression*. Thousand Oaks: Sage.
- Gelman, A., Rubin, D.L. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Sciences* 7(4): 457-472.
- Geweke, J. 1993. Bayesian Treatment of the Independent Student-t Linear Model. *Journal of the Applied Econometrics* 8, S19-S40.
- Gill, J. 2002. *Bayesian Methods for the Social and Behavioral Sciences*. Boca Raton: Chapman & Hall/CRC.
- Gill, J., Gajdos, J. 2002. Why Does Voting Get so Complicated? A Review of Theories for Analyzing Democratic Participation. *Statistical Science* 17(4): 1-22.
- Hobert, J.P., Casella, G. 1996. The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association* 91(436): 1461-1473.
- Jackman, S. 2000. Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo. *American Journal of Political Science* 44(2): 375-404.
- Johnson, V.E. 1996. Studying Convergence of Markov Chain Monte Carlo Algorithms Using Coupled Sample Paths. *Journal of the American Statistical Association* 91(433): 154-166.
- Lee, P. M. 1989. *Bayesian Statistics: An Introduction*. New York: Oxford University Press.
- Leighley, J.E., Nagler, J. 1992. Socioeconomic Bias in Turnout 1964-1988: The Voters Remain the Same. *American Political Science Review* 86(3): 725-736.
- Martinez, M., Hill, D. 1999. Did Motor Voter Work. *American Politics Quarterly* 27(3): 296-315.
- Nathan, S. 2000. More Investors Click to Cast Proxy Votes. *USA Today*, 27 March, 13B.
- Pollard, W. E. 1986. *Bayesian Statistics for Evaluation Research: An Introduction*. Beverly Hills: Sage.
- Press, S. J. 1989. *Bayesian Statistics: Principles, Models, and Applications*. New York: Wiley.
- Riker, W. H., Ordeshook, P.C. 1968. A Theory of the Calculus of Voting. *American Political Science Review* 62(1): 25-42.
- Robert, C.P. 2001. *The Bayesian Choice: A Decision Theoretic Motivation*. Second Edition. New York: Springer-Verlag.
- Robert, C.P. 1995. Convergence Control Methods for Markov Chain Monte Carlo Algorithms. *Statistical Science* 10(3): 231-253.
- Rosenstone, S., Hansen, J.M. 1993. *Mobilization, Participation, and Democracy in America*. New York: Macmillan.
- Salant, J.D. 2000. Bush campaign last in Internet fund-raising. *Associated Press*, 5 February.
- Selnow, G.W. 1998. *Electronic Whistle-Stops: The Impact of The Internet on American Politics*. Westport, CT.: Praeger Publishers.
- SeniorNet.org. <http://www.seniornet.org/php>

- Solop, F.I. 2001. Digital Democracy Comes of Age: Internet Voting and The 2000 Arizona Democratic Primary Election. *Political Science and Politics* 34(2): 289-293.
- Solop, F.I. 2000. Public Support for Internet Voting: Are We Falling into a 'Racial Ravine'. Paper presented at The American Association of Public Opinion Research, Portland, Oregon.
- Tullock, G. 1967. *Towards a Mathematics of Politics*. Ann Arbor: University of Michigan Press.
- United States Department of Commerce. 2000. *Falling Through The Net: Toward Digital Inclusion*.
- U.S. Dept. of Commerce. Bureau of Census. *Statistical abstract of the United States 2000*. Washington D.C.: Government Printing Office, various.
- Western, B., Jackman, S. 1994. Bayesian Inference for Comparative Research. *The American Political Science Review* 88(2): 412-423.
- Wolfinger, R.E., Rosenstone, S.J. 1980. *Who votes?* New Haven: Yale University Press.