

ECONOMIC AND DEMOGRAPHIC DETERMINANTS OF THE ELECTORAL  
VOTE FOR PRESIDENT\*

Robert McNown

Department of Economics, International Affairs Program,  
and Institute of Behavioral Science  
University of Colorado, Boulder

August 2008

Robert McNown  
Department of Economics  
256 UCB  
University of Colorado  
Boulder, Colorado 80309  
Email: mcnownr@colorado.edu

---

\* The author is indebted to Stefan Awad for his careful and thorough preparation of the data and assistance with the estimation, and to Kali Gajewsky for help in preparing the 2008 election forecasts.

ECONOMIC AND DEMOGRAPHIC DETERMINANTS OF THE ELECTORAL  
VOTE FOR PRESIDENT

**Abstract**

This paper presents a model of the determinants of the electoral vote for President, estimated with panel data across 51 states plus the District of Columbia over nine election years from 1972 to 2004. The model expands upon others that are driven by national economic variables to test the importance of state-specific demographic characteristics. Among these demographic factors, the poverty rate and the proportion African American do not significantly affect the division of the two-party vote within the states, while educational attainment, the elderly proportion of the population, and the percentage living in non-metropolitan areas have significant effects on the electoral vote. Inclusion of demographic variables lowers considerably the prediction errors of the last two elections, compared to forecasts of a model that excludes these factors. Lags in the model structure allow real time election forecasts to be produced before the end of July of a Presidential election year.

## **I. Introduction**

Economic models of Presidential voting predicted a strong win for George W. Bush in 2004. The median forecast from the seven models summarized in Campbell (2004) called for President Bush to capture 53.8 percent of the two-party popular vote (See Table 1). Ray Fair's (2004) model in October 2004 predicted that President Bush would receive 57.7 percent of this vote compared with the 51.6 percent that he actually received.

The forecast errors for the 2004 election generated some skepticism regarding economic models as predictors of Presidential election outcomes. Although some commentators cited a shift away from economic voting towards values-based decisions, the 2004 election results are not consistent with such an interpretation. The common understanding of values voting implies an increased concern over traditional family values, which would tend to favor Republican over Democratic candidates. If values voting had been a prominent force in the 2004 election, President Bush would have won by an even greater margin than had been predicted by the economic models.

Hillygus and Shields (2005) analyzed data from a post –election internet survey for explanations of the 2004 election outcome. They found that neither gay marriage nor abortion issues affected individuals' votes in the Presidential election, except in Southern states where the pro-Bush vote was dominant in any case. Rather the traditional factors of party affiliation and the economy had the greatest influence on individual votes.

Analysis of voter turnout by Burden (2004) and Abramowitz (2004) also failed to support the claims that values voting figured importantly in the 2004 election. Burden

(2004) analyzed exit poll data from the 2004 election to examine the claim that Bush was re-elected by a surge in the numbers of socially conservative voters. Voter turnout by groups that represent social conservatives did not change appreciably between 2000 and 2004. Regular church-goers comprised 27% of the voters in 2000 and 26% in 2004. A regression analysis of the 2004 vote across states showed no significant effect on Bush's vote share from the presence of a gay marriage initiative on the ballot.

Likewise, Abramowitz (2004) found that gay marriage initiatives had no significant effect either on voter turnout or on Bush's share of the vote in 2004. Both Burden and Abramowitz found that Bush's gains in vote share between 2000 and 2004 were spread evenly across the states, a finding that is consistent with the positive effects of a first term of Republican occupancy of the White House and a moderately strong economy.

The poor prediction performance of the economic models in 2004 as shown in Table 1 is the motivation behind this paper, taking a broader perspective on economic determinants of voting decisions. The hypothesis underlying this analysis is that the relation between national economic performance and Presidential voting is conditioned on socio-economic characteristics of the electorate. Specific demographic groups identified as economically disadvantaged in US society hold long term allegiances to the Democratic party, due to the perception that social programs favored by Democrats will work to their advantage. Variations in the demographic composition of the population influence the partisan leanings of the electorate, which, together with macroeconomic performance, affect Presidential election outcomes. One objective of this study is to examine the contribution of these socio-economic characteristics in an economic model

of Presidential voting. An unusual finding is that states with higher proportions of metropolitan residents tend to favor Republican candidates, contrary to some conventional views about urban voters (see, for example, Sperling and Helburn 2005; Cook 2005). In addition, the view that African American voters and the poor tend to vote for Democratic candidates was not confirmed in this study.

Presidential elections are decided by the aggregation of 51 state outcomes through the Electoral College. Consequently, these elections are most appropriately modeled at the level of the individual states in a pooled data analysis. A few studies have employed state-specific data in a pooled model involving causal variables (Strumpf and Phillippe 1999; Campbell 1992; Holbrook 1991; Peltzman 1990), but none have found these models to be useful for predicting the electoral vote outcome.

Jones (2008) surveys the predictive performance of several statistical models designed to predict the electoral vote in 2004. None of these are in the explanatory model tradition. For example, the regression based models used to forecast the 2004 electoral vote only employ polling data taken within states. Jones concludes that the average errors in forecasting the electoral vote tend to be quite high and variable. The average absolute error in electoral vote forecasts across seven methods was five percent, with a range from 0 to 11 percent, these two extremes coming from two alternative forecasts produced by the same modeler. Furthermore, these errors are from forecasts produced during the week before the election (or the night of the election in the case of the exit polls). These electoral vote forecasting models, therefore, lack the theoretical foundations of causal models, suffer from high forecast error rates, and do not provide predictions with useful lead times, scoring poorly on three important characteristics of successful forecasting

models (Lewis-Beck and Tien 2008). One goal of this study is to develop a theoretically based model that can produce accurate electoral vote forecasts with a lead time of at least three months.

The models presented here provide estimates of the effects of both national economic and state level socioeconomic variables on presidential voting over the nine elections since 1972. The motivation behind the models is explanatory or causal, in the sense that polling data and Presidential approval ratings are not included. Such voter opinion variables are not causal determinants of election outcomes, but rather are reflections of underlying socio-economic variables that motivate voter behavior.

Although the model underpinnings are explanatory, forecast error analysis is employed for model validation and to examine the contributions made by economic and demographic variables to the explanation of election outcomes. Peltzman (1990) and Strumpf and Phillippe (1999) contend that economic variables add little to the predictive content of a model containing only incumbency and state fixed effects. An additional goal of the forecast analysis is to re-examine this claim, assessing the contribution of both economic and demographic variables to the predictive power of the model. The complete model of this study shows a 40 percent reduction in the average root mean squared error relative to that of the baseline model that includes only a terms-in-office variable and state fixed effects. This indicates that the economic and demographic factors can add substantially to the predictive performance relative to the baseline model. In an out-of-sample analysis of the 2004 election, all but three state outcomes are correctly forecasted by the complete model of this study, compared with errors for fourteen state elections

from the models that exclude demographic factors. The resulting electoral vote forecast gave President Bush an additional 24 electoral votes above the 286 he actually received.

## **II. A Model of Voting Behavior**

The model of voting behavior employed here is grounded in principal-agent theory. In Presidential elections voters hold the incumbent party responsible for macroeconomic performance, commonly measured by the rate of growth of real per capita disposable income (RDI), the unemployment rate (UE), the rate of consumer price inflation (CPI), and the rate of growth in total employment (EG). The first three variables are fairly standard in the existing literature. Employment growth figured prominently in the 2004 election, and had been found by Lewis-Beck and Tien (2004) to add to the predictive performance over that of a model with RDI. Higher unemployment and inflation rates are presumed to be costly to voters, so that these variables are usually expected to have negative effects on the incumbent party's vote share. However, net debtors may benefit from unanticipated increases in inflation, so that a positive relation between this variable and the incumbent party's vote share is also plausible. The unemployment rate is a noisy measure of macroeconomic labor market conditions as people enter and leave the labor force depending upon job opportunities. Consequently, the employment growth rate may have a stronger influence on voter decision than the unemployment rate. The final model specification selects from these four candidate variables based on statistical significance.

Lags on the economic variables may be short, reflecting myopic behavior by the voters, or long, indicating that voters take into account economic performance over a

substantial portion of the current President's four-year term. A systematic search by Strumpf and Phillippe (1999) favored lags capturing two or more years of cumulative economic events, and this was confirmed in the present study. Cumulative growth rates of real disposable income and employment growth, and averages of unemployment and inflation rates were found to have maximum quantitative effects and greatest statistical significance with lags extending back two years prior to the second quarter of the election year (e.g. the third quarter of 2002 through the second quarter of 2004 for the observations for the 2004 Presidential election). The finding that voters are not myopic in the use of economic information undermines the possibilities for short run manipulation of the economy for political gain as posited by the political business cycle model. The decision not to include economic data through the third quarter of the election year allows real time forecasts to be produced by the end of July in the election year.

Strumpf and Phillippe (1999), Eisenberg and Ketcham, (2004) and Holbrook (1991), investigated the importance of state level economic conditions as well as national economic variables on Presidential elections. Although some state variables showed statistical significance in determination of Presidential voting, these were strongly dominated by comparable national variables. The President is seen by the voters as influencing the economy at the national level while differences in state-specific economic performance are largely independent of macroeconomic conditions (Eisenberg and Ketcham 2004). Consistent with this reasoning and empirical evidence, state-specific macroeconomic variables are not included in the models estimated in this study.

The influence of national economic factors on voter behavior is conditioned by underlying demographic characteristics of the electorate. Party affiliation is a strong

determinant of Presidential voting (Stanley and Niemi 2004), and party allegiances are influenced by the electorate's perception of the general economic policies of the two parties. The Democratic coalition, for example, reflects the perception by economically less advantaged groups in society that they will benefit from social programs favored by that party. Therefore, partisan tendencies of each state are hypothesized to reflect average levels of educational attainment, racial composition, age distributions, the share of state populations living in non-metropolitan areas, and poverty rates. Additional unobserved partisanship is captured by state fixed effects.

Two variables measure educational attainment: the proportion of each state's adult population with a high school degree (POPHS) and the proportion with a Bachelor's degree (POPBA). The other demographic variables are the percentage of each state's population that is African American (POPAN), the percentage that is 65 years of age and older (POP65), the percent living in non-metropolitan areas (POPNM), and the percent with incomes below the official poverty line (POV). The demographic factors for each state are measured for the year prior to the election.

Most election forecasting models include one or two incumbency variables: one to capture the advantage of personal incumbency, and another that reflects the disadvantage faced by a party that has held the Presidency for two or more terms. The election of 1992 (George H.W. Bush vs. Clinton) was the only election since 1972 in which the party of an elected incumbent President had held the office for more than one term. Therefore, there is insufficient data to distinguish between these two effects on election outcomes. Including both variables in this model would be nearly redundant.

Personal incumbency does not carry the same advantages in a Presidential race as it does in Congressional contests. A Congressional incumbent may receive the advantages of name recognition, fundraising for campaigns, and a gerrymandered district, and these account for the high percentage of members of the House of Representatives that are returned to office every two years. A Presidential candidate, on the other hand, can only be a running incumbent once, and does not accrue the same advantages of incumbency listed above for members of Congress. On theoretical grounds, therefore, personal incumbency is not included in the model.

The second incumbency variable is based on the findings of Fair (2004) and Abramowitz (2001, 2008) that incumbent parties lose some support once they have been in office beyond one term. Bartels and Zaller (2001) offer an explanation for the importance of this variable as incumbent party fatigue. When a party is newly elected into office, the first round of political leaders, administrators, and advisors of the highest quality eagerly join the new administration. As a party remains in office beyond one term, the first string of officials moves on to other activities, and they are replaced with less able successors. Second term administrations are more often plagued by scandals, and a lame duck administration will be less effective in implementing its legislative agenda. Meanwhile the challenging party will tire of being out of power and will be energized and unified in an effort to recapture the White House. Party in-fighting and domination by extreme ideologies will give way to moderation in an effort to regain political power. This effect is measured by a dummy variable (TC) that equals one when the incumbent party has held the Presidency for two or more terms.

The dependent variable is the Democratic candidate's share of the two-party vote. Since the macroeconomic variables are expected to affect the incumbent party's vote share, the time-in-office variable and all national economic variables must be interacted with a (-1, +1) dummy variable indicating whether the White House is occupied by a Republican (-1) or a Democrat (+1).

This voting model is estimated with data over the past nine Presidential elections from 1972 through 2004. Aggregate time series studies generally require longer time series, which faces the risks of structural shifts due to changes in institutions or in the composition of the electorate. 1972, for example, was the first Presidential election in which eighteen year olds were allowed to vote. For each election year the dependent variable and the demographic characteristics are collected for each of the 50 states plus the District of Columbia, for a total of 459 observations. Data sources are reported in the appendix.

### **III. Methodological Issues**

The use of pooled data necessitates the choice among fixed effects, random effects, or a common intercept across all states. Given the construction of the dependent variable as the Democratic percentage of the two-party vote, state-specific effects are interpreted as the relative partisan bias of each state towards Democratic candidates. Clearly partisanship varies substantially across states, ruling out the common intercept model. The random effects model would be appropriate if these effects are uncorrelated with the explanatory variables in the model. The Hausman (1978) test of this condition is based on a statistical comparison of fixed effects and random effects estimates, and sharp

differences in coefficient estimates for some demographic variables rules out the use of the random effects model. Although the fixed effects estimator uses less information than the random effects model and is therefore less efficient, Monte Carlo comparisons of these two models find the fixed effects estimator retains good unbiasedness and efficiency properties (Baltagi 2008, page 41). In addition, Baillie and Baltagi (1999) present simulation evidence that the fixed effects model produces forecasts with mean squared errors that are competitive with random effects predictions.

One goal of this study is to uncover the state characteristics that underlie the partisan differences. The overall partisan bias of each state is hypothesized to be a reflection of social and demographic characteristics of each state's population. For example, the overwhelming tendency for Washington, D.C. to vote for the Democratic candidate may reflect the large proportion of the District's voters who are African American. To the extent that social and demographic variables capture these underlying sources of partisanship, forecast errors should be reduced by the inclusion of these variables in a model that also includes fixed effects, incumbency factors, and macroeconomic variables.

Some studies of Presidential elections allow for possible endogeneity of economic performance variables (Allesina, Londregan, and Rosenthal 1993). Concern over endogeneity is grounded in the theory of political business cycles, which contends incumbent Presidents will initiate policies that produce temporary accelerations in economic growth, timed to curry favor with the voters. This model has been strongly criticized on institutional, theoretical and empirical grounds. Although voters may hold the incumbent President, or his party, responsible for economic performance during his

term of office, it is not clear that the President has effective, unilateral control over the economy or the tempo of the business cycle. Business fluctuations are affected by monetary policy as well as fiscal policy, and the lags in enactment of fiscal policy and its impact on output growth are uncertain, variable, and dependent upon actions by the Congressional branch of government. Theoretical and empirical critiques of the political business cycle model are summarized in Peltzman (1990). Given these considerations, endogeneity of economic performance variables is not seen as a concern in the present analysis.

All models are estimated by the least squares – fixed effects procedure. Temporal autocorrelation is tested using 2dridge's (200?) t-statistic in a panel regression of residuals on lagged residuals. Standard errors are calculated using White's (1980) procedure to be robust to heteroscedasticity across states or over time. Although the dependent variable is constrained to lie between zero and one hundred, the use of the logistic transformation does not appear to be necessary. Most observations on the dependent variable fall well within these bounds, and other studies that have considered this issue found no substantive reason to favor the logistic over the linear models (Strumpf and Phillippe 1999; Campbell 1992).

#### **IV. Model Estimates and Forecast Analysis**

Least squares estimates of three alternative models are presented in Table 2. The baseline model 1 includes only the state fixed effects and the variable that indicates whether the incumbent party has held the Presidency for two or more terms (TC). The

other three models build on the baseline model by adding economic variables (model 2), as well as the demographic variables (model 3).

Before discussing the substantive implications of the model estimates, the forecast evaluation statistics are presented in Table 3. This evaluation emphasizes state-specific forecasts, using both in-sample and out-of-sample statistics. For the 2000 and 2004 elections, which were problematic for many forecasting models (see Table 1 and the symposium in *PS Political Science and Politics* edited by Bartels and Zaller 2001), the implied electoral vote forecast is also discussed. Since Presidential elections are decided by the electoral vote based on individual state outcomes, not the national popular vote, this is an appropriate focus for the forecast evaluation.

This model does not produce forecasts of the national popular vote, which is the most common basis for evaluating forecast performance. Forecasts of the Democratic vote percentage by state must be multiplied by the corresponding total number of votes in each state in order to convert to a prediction of the national popular vote. Since the total number of votes is not known prior to the election, this conversion to a national vote would not be a true forecast, and a comparison with national model forecasts would not be valid. In any case, since the Presidential race is the aggregation of 51 individual state and District of Columbia races, forecasting records at the state level are more informative about the outcome of the election.

The root mean squared error (RMSE) is computed from the popular vote in each state. For each election and each state the error is computed as the difference between the observed and the predicted vote percentage going to the Democratic candidate. These

errors are summed, squared, and averaged over all states to obtain the MSE for each election, and subsequently the RMSE.

For each election and each state the predicted Democratic candidate vote percentage is compared with fifty percent to determine each state's predicted winning candidate. This prediction is matched with the actual outcome, and the proportion of correctly forecasted outcomes is computed for each election.

For every year the addition of national economic variables (model 2) reduces the within-sample RMSE substantially relative to the baseline model 1 (Table 3, top section). The economic variables reduce the overall RMSE averaged over the nine years by almost one-third. The inclusion of both national economic and socio-economic variables in Model 3 reduces the overall RMSE by forty percent relative to the baseline model, and every election except 1992 contributes to this further reduction over model 2. In his forecast analysis Peltzman (1990) found that in only three of the ten years of his study were the RMSEs lowered considerably by the inclusion of economic factors, leading to his conclusion that only under unusual economic conditions does the economy have much effect on Presidential voting. In this model, however, national economic variables and state socio-economic factors substantially affect voting in every year. The smallest reduction in RMSE of fourteen percent occurs for the 1980 election, and in five of the nine election years this reduction exceeds forty percent. These results indicate much greater effects of socio-economic voting on election outcomes than is implied by Peltzman's model.

The summary RMSEs averaged over the nine elections indicates the forecast errors in this study are comparable to those of other models that use state level data. The

RMSE of 3.75% for Model 3 may be compared with Peltzman's (1990) RMSE of 4.53% over ten elections, Campbell's (1992) mean absolute percentage error of 3.02%, and Strumpf and Phillippe's (1999) median absolute error of 2.2%. Although some differences arise due to alternative summary statistics and sample periods, all state level models, including this one, produce overall forecast errors of similar magnitudes.

The second section of Table 3 records the within-sample proportions of correctly predicted state election wins for each Presidential election year. The 1976 election, with Gerald Ford running as an unelected incumbent, poses the greatest forecasting challenge for any of the three models. Even the most complete model 3 correctly predicts only two-thirds of the individual state outcomes. The 1972 and 1984 elections are of the opposite extreme, with the outcomes predicted with 86 percent accuracy with the simplest baseline model, and 98% correctly predicted outcomes with models 2 and 3. For the remaining seven elections the economic and demographic variables add to the predictive success of the model, especially in the last four election years. In every election year except 1976 more than 85 percent of the state races are correctly predicted by the most complete model 3.

Out-of-sample forecasts are presented for the 2000 and 2004 elections as a further model validation exercise. Each of the three models is re-estimated using data that excludes the forecasted election year, and the RMSEs and proportions correct are computed and displayed in the bottom panel of Table 3. These two elections posed particular problems for most models designed to forecast the national popular vote (Table 1 and Bartels and Zaller 2001).

The out-of-sample predictions for 2000 and 2004 show progressive improvement as more information is included in the model. For the 2000 election the RMSE declines from 10.4 percent for model 1 to 4.5 percent for model 3. In addition, model 1 predicts the actual state outcomes only 61 percent of the time, versus 84 percent and 92 percent for models 2 and 3, respectively. The complete model does not predict the outcome of only four state elections: Arizona, Missouri, Ohio, and West Virginia. In all four cases the model incorrectly predicted a win for the Democrats, which would have produced a forecast of an Electoral College majority for that party. This illustrates the difficulty of predicting the Electoral College winner, especially in this close election.

In the 2004 election the demographic variables contribute substantially to forecast accuracy, but the national economic variables alone do not improve the forecasts relative to the basic model 1. In fact, the RMSE for model 2 actually exceeds that for model 1 (5.2% vs. 4.6%), while the RMSE for model 3 is reduced to 3.5%. Models 1 and 2 have the same record of successful state election predictions at 72.5%, generally consistent with the errors in national popular vote forecasts that relied on economic plus incumbency factors (see Table 1). The complete model performs substantially better in this contest, predicting 94% of the state elections correctly, with misses for only three states: Republican wins were incorrectly predicted for Maine, Oregon, and New Hampshire. Therefore, President Bush was predicted to receive 310 electoral votes compared with his actual vote of 286. The amounts to an electoral vote error rate of 5.2%, which is in the range of errors for other 2004 electoral vote forecasts summarized by Jones (2008).

The forecast analysis establishes the validity of the socio-economic variables in explaining state level voting behavior. Forecast error statistics are comparable to those in the previous literature, and sufficiently small in absolute terms to provide confidence in the model specifications. The analysis also shows gains in predictive power with the addition of economic and/or demographic variables to the baseline model, allowing substantive interpretations of the individual coefficient estimates in Table 2.

## **V. Discussion of Model Estimates**

The models presented in Table 2 are distillations of more general specifications grounded in principal-agent theories of voting behavior, informed by empirical specifications in other models of Presidential election outcomes, and augmented by survey studies of demographic determinants of partisan leanings. These foundations are summarized in the theoretical discussion of Section II.

The final specifications exclude variables that failed to show statistical significance in the complete model 3, and some of these excluded variables are of interest for their surprising lack of explanatory power. Among the national economic variables, the final specification includes only the inflation rate (INFL) and the employment growth rate (EG), both measured over the eight quarters ending with the second quarter of the election year. The rate of growth of real disposable income and the unemployment rate did not contribute significantly to the explanation of state Presidential election outcomes. The unemployment rate is a noisy measure of aggregate labor market conditions, due to withdrawals from and entries into the labor force with changing business cycle conditions. The employment growth rate, also emphasized by Lewis-Beck and Tien

(2004), is found to be a more significant indicator of labor market conditions as they affect Presidential voting. This variable also captures real economic growth effects on Presidential voting, with inclusion of real disposable income not adding to the explanatory power of the full model.

The final list of state demographic factors includes the percentage of the population residing in nonmetropolitan areas (POPNM), the percentage of the population with at least a Bachelor's degree (POPBA) or with at least a high school diploma (POPBS), and the percentage 65 years of age and older (POP65). Other demographic factors not demonstrating significant explanatory power were the poverty rate and the proportion of the state population that is African American. The lack of significance for these two variables undermines the conventional view of the Democratic coalition. To the extent that African Americans and those below the poverty line have partisan leanings towards Democratic candidates, these effects are already captured by the other demographic and national economic variables in the model.

Since the result on the racial variable is quite surprising, several alternatives to model 3 were estimated with the percent of the state population that is African American included. When added directly to model 3, this variable carries a positive coefficient that is significant at the 10 percent level. However, when the sample is truncated at the year 2000 to prepare the out-of-sample forecast for 2004, the coefficient loses significance (p-value of 0.27). To allow between-states variation to enter into the estimation, the model was re-estimated with random effects. While this did produce a statistically significant coefficient on the African American variable, the random effects specification is rejected by the Hausman test (1978). Significant differences between the coefficients estimated by

the fixed effects and random effects procedures imply that the random effects estimator is inconsistent due to correlations between the demographic variables and the state-specific effects contained in the error term.

There is no robust evidence, therefore, that this racial variable affects Presidential election outcomes by state, when controlling for other demographic factors in the model. Excluding the nonmetropolitan population percentage, for example, produces a large and statistically significant coefficient on the racial variable, but this exclusion must be interpreted as a model misspecification. It is well established from voter surveys that African Americans disproportionately favor Democratic candidates (for example, Erikson, et al. 1989), but compositional effects mean that aggregate and individual voting behavior can differ. For example, Glaser (1994) presents evidence that Southern whites who live in areas with higher concentrations of black residents tend to exhibit greater opposition to government programs directed towards advancement of African Americans. Therefore, African American support for Democrats in states with high concentrations of black populations may be offset by stronger negative attitudes by white towards Democratic candidates who favor affirmative action or civil rights programs.

Consistent with the result for the poverty rate, Stanley, et al. (1986) found no significant effect of being poor on party identification, and the study by Klingman and Lammers (1984) showed no relation between state poverty indexes and their measure of state ideology. Erikson, et al. (1989), on the other hand, state that being poor has been an important determinants of Democratic party identification since the Reagan era. Although the literature on their party identification is mixed, this study finds no evidence that the

poor demonstrate allegiance to one party over the other in their voting behavior, except as captured by their membership in the other demographic groups in the model.

Across all three models Presidential candidates from a party that has occupied the White House for two or more terms are severely handicapped. In the full model such a candidate starts with a 4.8% disadvantage in each state election, on average. This strong effect is comparable to those found in other national Presidential voting models.

Abramowitz (2004) and Holbrook (2004) estimate the penalty for more than one term of party incumbency to be more than five percentage points. Fair's (2004) duration variable, which increases by one point for every additional term of party incumbency, has an estimated negative effect of 3.6 percentage points on that party's vote share. Model 1 with the term-of-office variable and fixed effects only produces an R-squared of .62 and shows signs of significant first order serial correlation.

Adding inflation and employment growth variables increases the R-squared to 0.82, but there is still evidence of first order autocorrelation in model 2. Given this problem the discussion of parameter estimates is deferred to the complete model.

In model 3 86% of the variation in state-level voting is explained, and the autocorrelation test indicates no significant problem. The two national economic variables have strongly significant explanatory power, with expected signs on the estimated coefficients. A one percentage point rise in the rate of inflation is expected to lower each states vote for the incumbent party's candidate by 0.37 percentage points, approximately half the impact estimated by Fair (2004). Given the low rates of inflation experienced by the United States since the mid 1980s, this variable has not played a major role in any Presidential election since 1984.

Employment growth, on the other hand, has had a strong impact on election outcomes, with a one percentage point increase in the rate of growth adding 2.9 percentage points to the incumbent party's expected share of state votes. With a standard deviation of 1.1, a one percentage point change in the employment growth rate is not an unusually large change to observe from one election to the next. Lewis-Beck and Tien (2004) also include employment growth, measured as the growth over the first three and one-half years of the incumbent's term in office. They estimate a half percent rise in incumbent party vote share for a one percentage point increase in employment growth. When converted into a one percent annual rate of employment growth, their estimate implies a 1.75 percentage point increase in the incumbent party's vote share. Clearly differences in sample, the cumulative lags in construction of this variable, and the set of control factors can account for this difference in estimated effects.

. In the complete model the four demographic variables show statistical significance at the .05 level or lower. States with higher population percentages over 65 and those with higher proportions living in non-metropolitan areas also tend to favor Democratic candidates. The partisan effects of these demographic categories may reflect perceptions that these groups stand to gain from the social programs of the Democratic party. Democratic candidates have more effectively capitalized on concerns of the elderly population about Social Security and Medicare, garnering a greater share of votes from the elderly voters than do the Republicans.

Erikson, et al. (1989) also found a metropolitan bias in favor of Republicans, contrary to the general perception that urban voters tend to vote Democratic (Sperling and Helburn 2005; Cook 2005). It is important in interpreting these conflicting points of

view to distinguish between urban and metropolitan residents. The latter includes suburban residents who are generally seen as a part of the Republican base.

The education variables work in opposite directions from each other. States with greater proportions of high school graduates tend to favor Republican candidates. Equivalently, states with larger populations disadvantaged by less than a high school education favor Democrats, a result consistent with the theory of Democratic partisanship expressed above. However, Model 3 estimates show that a higher proportion of college graduates biases the state towards Democratic candidates, a result that is contrary to this theory of partisanship. Contrary to this last result, Erikson, et al. (1989) found college educated voters tended to favor Republican candidates. However, they did not include a separate category of high school degree holders, which could account for their different results.

The relative impact of each demographic factor is measured by the effect of a one standard deviation change in that variable on the Democratic vote percentage. By this measure the population proportion that lives in non-metropolitan areas has the greatest effect on vote shares, followed by the percentage of the population with a Bachelor's degree. States with non-metropolitan population percentages that are one standard deviation above the norm tend to favor Democrats by eleven percentage points. Similarly, states with college degree population percentages that are one standard deviation above average show Democratic biases of 4.6 percentage points.

## **VI. Conclusion and 2008 Prediction**

This study has established the importance of state socio-economic factors in an economic model of Presidential voting. The model is estimated using a pooled data set spanning nine Presidential election years and 51 states plus the District of Columbia, allowing a prediction of the national electoral vote. The final model combines a terms-in-office effect, national economic variables, and state-specific socio-economic variables. In-sample forecast analysis establishes the validity of the model specification, with ninety percent of state election outcomes correctly predicted over the nine years of the sample. Out-of-sample predictions for the state level elections of 2000 and 2004 are correct in over ninety percent of the cases, even though these two elections proved problematic for other forecasters of the national popular vote.

Several notable results emerge from the complete model estimates. Parties that hold office for more than one term experience a 4.8 percentage point disadvantage in each state election. Of the national economic variables inflation and employment growth over the two years prior to the election have statistically significant effects on state election outcomes, displacing the unemployment rate and the growth in real disposable income from the model.

Six socio-economic variables are introduced to represent partisan bias at the state level. States with larger proportions of economically disadvantaged groups are hypothesized to favor Democratic candidates. Of the six demographic factors included in this study, neither state poverty rates nor percent African American showed robust statistical significance when controlling for other socio-economic factors. States with higher proportions of elderly populations, residents of non-metropolitan areas, and

persons without high school degrees show significant leanings towards Democratic candidates.

Previous state-level studies have led to skepticism about the use of economic voting models for predicting electoral vote outcomes (Strumpf and Phillippe 1999; Peltzman 1990), and recent attempt for forecast the electoral vote have relied on non-explanatory models (Jones 2008). However, within sample predictions of state outcomes achieve a 90 percent success rate over the nine elections of this sample. An out-of-sample prediction for the 2004 race misclassified only three states, predicting that President Bush would receive 310 electoral votes instead of his actual count of 286.

A real-time forecast of the 2008 electoral vote, generated on July 31 2008, is displayed in Table 4. Consistent with contemporaneous forecasts of the national popular vote, this model predicts a strong win for the Democratic candidate with 363 electoral votes versus 175 for the Republican. Most large states will go for Senator Obama, with the exceptions of Texas and Florida, although the latter is predicted to show a very close 50.3% vs. 49.7% vote. Close races that are predicted to go to Obama are Arkansas (50.0), Georgia (50.9%), Louisiana (50.9%), and Tennessee (50.8%), with a total of 41 electoral votes. If all of these ended up going for Senator McCain, Barack Obama would still retain the Electoral College majority.

## References

- Abramowitz, Alan (2001) "The Time for a Change Model and the 2000 Election," *American Politics Quarterly* 29: 279-282.
- Abramowitz, Alan (2004) "Terrorism, Gay Marriage, and Incumbency: Explaining the Republican Victory in the 2004 Presidential Election," *The Forum* 2(4): Article 3 (<http://www.bepress.com/forum>).
- Abramowitz, Alan (2008) "It's About Time: Forecasting The 2008 Presidential Election With The Time-For-Change Model," *International Journal of Forecasting* 24(2): 209-217.
- Alesina, Alberto, John Londregan, and Howard Rosenthal (1993) "A Model of the Political Economy of the United States," *American Political Science Review* 87(1): 12-33.
- Baillie, Richard T., and Baltagi, Badi H. (1999) "Prediction from the Regression Model with One-Way Error Components," in Hsiao, Cheng, ed. *Analysis of Panels and Limited Dependent Variable Models*. Cambridge: Cambridge University Press. 255-67.
- Baltagi, Badi H. (2008) *Econometric Analysis of Panel Data*. Chichester, UK: Wiley.
- Bartels, Larry M. and John Zaller (2001) "Presidential Vote Models: A Recount," *PS: Political Science and Politics* 34(1): 8-20.
- Burden, Barry C. (2004) "An alternative Account of the 2004 Presidential Election," *The Forum* 2(4): Article 2 (<http://www.bepress.com/forum>).
- Campbell, James E. (1992) "Forecasting the Presidential Vote in the States," *American Journal of Political Science* 36(2): 386-407.

- Campbell, James E. (2004) "Introduction – The 2004 Presidential Election Forecasts," *PSOnline* 733-735 ([www.apsanet.org](http://www.apsanet.org)).
- Cook, Charles E. (2005) "Did 2004 Transform U.S. Politics," *The Washington Quarterly* 28(2) 173-186.
- Eisenberg, Daniel and Jonathan Ketcham (2004) "Economic Voting in U.S. Presidential Elections: Who Blames Whom for What," *Topics in Economic Analysis and Policy* 4(1) 1-23 (<http://www.bepress.com/bejeap/topics/vol4/iss1/art19>).
- Erikson, Robert S., Thomas D. Lancaster, and David W. Romero (1989) "Group Components of the Presidential Vote, 1952-1984" *Journal of Politics* 51(2): 337-346.
- Fair, Ray C. (2004) "A Vote Equation and the 2004 Election," 1-10  
<http://fairmodel.econ.yale.edu/vote2004/vot1104a.pdf>
- Glaser, J. M. (1994) "Back to the Black-Belt - Racial Environment And White Racial-Attitudes in The South," *Journal of Politics*, 56 (1): 21-41
- Hausman, J. A. (1978) "Specification Tests in Econometrics," *Econometrica* 46 (6): 1251-1271.
- Hillygus, Sunshine and Todd Shields (2005) "Moral Issues and Voter Decision Making in the 2004 Presidential Election" *PS Political Science and Politics* 38 (4): 665-665.
- Holbrook, Thomas M. (1991) "Presidential Elections in Space and Time" *American Journal of Political Science* 35(1): 91-109.
- Jones, Randall J. Jr. (2008) "The State Of Presidential Election Forecasting: The 2004 Experience," *International Journal of Forecasting* 24(2): 310-321.

- Klingman, David and William W. Lammers (1984) "The General Policy Liberalism Factor in American State Politics," *American Journal of Political Science* 28(3): 598-610.
- Lewis-Beck, Michael S. and Charles Tien (2004) "Jobs and the Job of President: A Forecast for 2004," *PSOnline* 753-758 ([www.apsanet.org](http://www.apsanet.org)).
- Lewis-Beck, Michael S. and Charles Tien (2008) "Forecasting Presidential Elections: When to Change the Model," *International Journal of Forecasting* 24: 227-236.
- Peltzman, Sam (1990) "How Efficient is the Voting Market," *Journal of Law and Economics* 33: 27-63.
- John G. Sperling, John G. and Helburn, Suzanne W. (2005) *The Great Divide: Retro vs. Metro America*. PoliPointPress.
- Stanley, Harold W. and Richard G. Niemi (2004) "Partisanship, Party Coalitions, and Group Support, 1952-2000," in Weisberg, Herbert F. and Clyde Wilcox, eds., *Models of Voting in Presidential Elections: The 2000 U.S. Election*. Stanford: Stanford University Press. 123-140.
- Strumpf, K Coleman S. and John R. Phillippe, Jr. (1999) "Estimating Presidential Elections: The Importance of State Fixed Effects and the Role of National versus Local Information," *Economics and Politics* 11(1) 33-50.
- White, H. (1980) "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48: 817-838.
- Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data* Cambridge, MA: MIT Press.

Table 1. Model forecasts of 2004 popular vote share.

Forecaster	Predicted Vote Percent for Bush	Data of forecast
Abramowitz	53.7	July 31
Campbell	53.8	September 6
Wlezien and Erikson	51.7 to 52.9	August 27
Holbrook	54.5	August 30
Lewis-Beck and Tien	49.9	August 27
Lockerbie	57.6	May 21
Norpoth	54.7	January 29
Fair	57.7	October 29

Note. The first seven forecasts are taken from Table 1 in Campbell (2004). The forecast by Fair is from Fair (2004).

Table 2. Model Estimates, 1972-2004 Presidential election years.

Variable	model 1		model 2		model 3	
	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
C	45.94	128.868	46.916	197.644	21.963	4.361
INFL			-0.475	-8.284	-0.366	-6.733
EG			3.455	23.299	2.913	20.214
TC	-2.173	-4.881	-4.053	-10.315	-4.815	-12.584
POPNM					0.565	4.231
POPBA					0.766	5.901
POPHS					-0.215	-2.287
POP65					0.804	2.487
R-squared	0.62		0.823		0.864	

Notes. The dependent variable is the Democratic Party candidate's percentage share of the two-party vote. National level variables are INFL=CPI inflation rate; EG=total civilian employment growth rate (both computed over the two year period ending in June of each election year); TCID=1 if incumbent party has been in office for two terms or more. State specific variables are POPNM=percent of population that resides in nonmetropolitan areas; POPHS=percent of the adult population with a high school degree; POPBA=percent of the adult population with a Bachelor's degree; POP65=percent of population 65 years of age and older. All models are estimated with state fixed effects; t-statistics are computed from standard errors robust to heteroscedasticity in either temporal or cross sectional dimensions. The autoc test reports the t-statistic on the one-period lagged residual in the pooled regression of residuals on their lagged values.

Table 3. Forecast error analysis: 1972 – 2004 Presidential election years.

	Model 1	Model 2	Model 3
RMSE of state popular vote percentages			
1972	10.438	6.330	5.834
1976	5.982	5.549	4.617
1980	5.481	5.106	4.738
1984	6.641	2.881	2.561
1988	4.130	3.555	3.248
1992	4.863	2.405	2.859
1996	8.251	3.835	3.393
2000	6.000	4.211	3.414
2004	4.061	4.543	3.055
average	6.205	4.268	3.747
Proportion of state outcomes correctly predicted			
1972	0.863	0.980	0.980
1976	0.588	0.667	0.667
1980	0.882	0.922	0.882
1984	0.863	0.980	0.980
1988	0.804	0.882	0.863
1992	0.765	0.902	0.922
1996	0.549	0.882	0.863
2000	0.706	0.922	0.961
2004	0.784	0.745	0.941
average	0.756	0.876	0.895
Out of sample analysis			
2000			
RMSE	10.403	5.475	4.524
Proportion correct	0.608	0.843	0.922
2004			
RMSE	4.572	5.208	3.527
Proportion correct	0.725	0.725	0.941

Table 4. 2008 electoral vote forecast

State	% Democrat	Democratic Win	Electoral Votes	Dem votes	Rep votes
Alabama	47.9	0	9	0	9
Alaska	43.2	0	3	0	3
Arizona	45.2	0	10	0	10
Arkansas	50	1	6	6	0
California	58.4	1	55	55	0
Colorado	53.3	1	9	9	0
Connecticut	59.5	1	7	7	0
Delaware	60.1	1	3	3	0
D.C.	98.7	1	3	3	0
Florida	49.7	0	27	0	27
Georgia	50.9	1	15	15	0
Hawaii	64.4	1	4	4	0
Idaho	36.3	0	4	0	4
Illinois	58.5	1	21	21	0
Indiana	46.6	0	11	0	11
Iowa	54	1	7	7	0
Kansas	43.2	0	6	0	6
Kentucky	49.2	0	8	0	8
Louisiana	50.9	1	9	9	0
Maine	56.6	1	4	4	0
Maryland	62.1	1	10	10	0
Massachusetts	69	1	12	12	0
Michigan	57.2	1	17	17	0
Minnesota	60.6	1	10	10	0
Mississippi	42.6	0	6	0	6
Missouri	53	1	11	11	0
Montana	50	0	3	0	3
Nebraska	40.3	0	5	0	5
Nevada	49	0	5	0	5
New Hampshire	51.9	1	4	4	0
New Jersey	58.7	1	15	15	0
New Mexico	51.6	1	5	5	0
New York	63.2	1	31	31	0
North Carolina	49.4	0	15	0	15
North Dakota	42.3	0	3	0	3
Ohio	53.8	1	20	20	0
Oklahoma	41.9	0	7	0	7
Oregon	56.7	1	7	7	0
Pennsylvania	57.7	1	21	21	0
Rhode Island	65.8	1	4	4	0
South Carolina	48	0	8	0	8
South Dakota	46.3	0	3	0	3
Tennessee	50.8	1	11	11	0
Texas	46.3	0	34	0	34
Utah	36.3	0	5	0	5
Vermont	58	1	3	3	0
Virginia	51.7	1	13	13	0
Washington	57.9	1	11	11	0
West Virginia	53.3	1	5	5	0
Wisconsin	56.6	1	10	10	0
Wyoming	41.3	0	3	0	3
		<b>31</b>		<b>363</b>	<b>175</b>

## Appendix – Data Sources

### POPULATION 65 AND OVER

Source: Bureau of the Census

- URL 1970 – 1999: <http://www.census.gov/popest/archives/>
- URL 2000-2004: <http://www.census.gov/popest/datasets.html>
- URL 2008: <http://www.census.gov/popest/age.html>
- o go to: [http://www.census.gov/popest/states/files/SC-EST2004-AGESEX\\_RES.csv](http://www.census.gov/popest/states/files/SC-EST2004-AGESEX_RES.csv)

### POPULATION 25 AND OLDER WITH HIGH SCHOOL OR COLLEGE DEGREE

Source 1: FreeLunch.com operated by economy.com

Consolidated data from U.S. Census Bureau: SUMMARY FILE 3

- URL: <http://www.economy.com/freelunch/default.asp>

Note: FreeLunch.com has provided an interpolation of the Census' decennial data smoothed near the decennial years. Data for 1972-1996 were from this source.

Source 2: Bureau of the Census, Current Population Survey

- URL 1999-2004: <http://www.census.gov/population/www/socdemo/educ-attn.html>

Source 3: Department of Education, National Center for Education Statistics

- URL 2008: <http://nces.ed.gov/programs/digest/d07/tables/xls/tabn011.xls>

Note: This cite provided educational attainment data for 2000 and 2005; the required observations for 2007 were obtained by extrapolating the rate of change in attainment between 2000 and 2005 forward by two years.

### POPULATION NON-METROPOLITAN

Source: Bureau of Economic Analysis

- URL: <http://www.bea.doc.gov/bea/regional/reis/>
- URL 2008: <http://www.bea.gov/regional/reis/drill.cfm>

o **Step 1:** Personal income and population summary estimates (CA1-3)

o **Step 2:** Population, State Metro/Non-Metro Portions, Years 1969-2003

### POPULATION AFRICAN AMERICAN

Source: Bureau of the Census

- URL 1970 – 1999 (excludes 1980): <http://www.census.gov/popest/archives/>  
(Population by State, age, sex, race.)
- URL 2000 – 2003: <http://www.census.gov/popest/datasets.html>
- URL 2008: <http://www.census.gov/popest/states/>

o Go to: <http://www.census.gov/popest/states/asrh/files/SC-EST2003-race6.csv>

Calculated proportion of total population categorized as black or African American.

## POVERTY RATE

Source: Bureau of the Census

- URL 1980 - 2003: <http://www.census.gov/hhes/www/poverty/histpov/hstpov21.html>

- URL Decennial 69, 79, 89: <http://www.census.gov/hhes/poverty/census/cph1162.html>

- URL Regional 1959,1960, 1969 – 2003:

-URL 2008: <http://www.census.gov/hhes/www/poverty/poverty.html>

<http://www.census.gov/hhes/www/poverty/histpov/hstpov9.html>

Note: For the years 1970 – 1978, a linear interpolation of the decennial data adjusted for yearly regional changes was used to construct the poverty data

## UNEMPLOYMENT RATE

Source: Bureau of Labor Statistics

-URL: <http://www.bls.gov/cps/home.htm>

-URL 2008: <http://www.bls.gov/data/#unemployment>

## INFLATION

Source: Bureau of Labor Statistics

-URL: <http://bls.gov/cpi/home.htm> (National CPI)

-URL 2008: <http://www.bls.gov/data/#prices>

Note: From CPI for all items - all urban consumers; 1967 base year.

## EMPLOYMENT GROWTH RATE

Source: Bureau of Labor Statistics

- URL: <http://bls.gov/cps/home.htm>

-URL 2008: <http://www.bls.gov/data/#employment>

Note: From seasonally adjusted national employment level for population 16 and older.

## REAL DISPOSABLE INCOME

Source: Bureau of Economic Analysis

- URL: <http://www.bea.doc.gov/bea/dn/nipaweb/index.asp>

-URL 2008: <http://www.bea.gov/national/index.htm#gdp>

Note: Data are quarterly percentage change in constant 2000 dollars.