Pocketbook Predictions of Presidential Elections—Pocketbook variables are almost always good indicators of electoral outcomes

By Patrick L. Anderson and Ilhan Kubilay Geckil

Patrick Anderson is the founder and principal of the consulting firm Anderson Economic Group LLC. He has served as the chief of staff of the Michigan Department of State and as a deputy budget director for the State of Michigan. He was an assistant vice president of Alexander Hamilton Life Insurance Co. and a graduate fellow for the Central Intelligence Agency. He holds a masters degree in public policy and a bachelors degree in political science, both from the University of Michigan at Ann Arbor.

Ilhan Kubilay Geckil is an economist with Anderson Economic Group. His work includes economic analysis, business valuation, strategy development, and forecasting. He holds an MA in economics from Michigan State University and a BA in economics from the KOC University in Turkey.

Political scientists have long known that “pocketbook issues” strongly affect the fortunes of presidents and other political leaders. Economists studying this relationship have established that certain economic factors—such as growth in income, inflation, and unemployment—directly affect the votes of the incumbent party in the presidential elections. Other institutional factors, such as the number of terms a party has occupied the oval office, also appear to affect voting patterns.

We present a set of simple “rational voter” economic models, which includes economic and institutional factors largely known in advance of the election. Such models typically explain about 75 percent of the variation of the popular vote in presidential elections since 1916. However, such time series models are bedeviled by the lack of observations, since presidential elections are only held every four years. To address this weakness, we specify and estimate a model with two methodological improvements: First, we use voting and economic data from the twenty largest states over elections since 1980, estimating parameters using pooled estimation techniques. Second, we improve the specification of the relevant variables, such as unemployment and the incidence of “limited wars,” to more accurately reflect the motivations of contemporary voters. Our pooled state model forecast for 2004 indicates that economic conditions favor a narrow re-election for the incumbent President. However, we point out how some elections cannot be entirely explained with a rational voter model.
Politicians have recognized that voters are interested in their pocketbook at least since the days of the Roman Empire. Indeed, in one recent presidential campaign, the incumbent was unseated by a challenger who made a mantra out of the phrase “It’s the economy, stupid.” Economists in recent decades have begun to quantify the relationship between electoral support and economic conditions. In some sense, this can be viewed as a rigorous response to the famous question posed by another presidential challenger: “Are you better off today than you were four years ago?”

In this paper, we start by briefly reviewing the standard approach to forecasting the results of presidential elections on the basis of pocketbook issues, as most comprehensively presented by Fair (2002). We outline how a rational voter model is consistent with public choice theory and other aspects of economic reasoning, and describe a small group of variables that has been shown to consistently explain voting behavior in presidential elections.

We note serious weaknesses in these models, most of which can be traced to a small number of observations over a very long time period.

To address these weaknesses, we suggest two methodological improvements, which allow for the use of significantly more data, explicitly estimate differences in the responses of voters from different states, and improve the specification of variables such as unemployment and war.

We then specify and estimate a model with these innovations and present results from different states. As a test, we compare the predicted response of voters in the 20 most populous states to the economic conditions in the year 2000 with their actual voting behavior.

Finally, we offer a conclusion about the limits of economic rationalism in voting behavior and suggest that some elections—such as the 1992 election, and potentially the 2004 election—are not fundamentally about economics.

The Standard Approach

Microeconomics is based on utility theory, which is founded on axioms regarding consumer behavior. Consumers are assumed to be rational in defined ways. For example, consumers are assumed to prefer more goods to fewer goods; less risk to more risk; and consumption now to consumption later. Upon these axioms the whole structure of microeconomics is based, and the rest of economics generally follows.

The advent of public choice theory has expanded the use of these axioms into the realm of political behavior. Public choice theory provides a rigorous economic understanding for what philosophers and peasants have known for thousands of years: politicians operate under their own set of incentives, which are not always consistent with those of the people they govern.

---

1 “The people that once bestowed commands, consuls ships, legions, and all else, now concerns itself no more, and longs eagerly for just two things--bread and circuses” complained the first-century Roman writer Juvenal.

2 Public choice theory is most associated with James Buchanan, who won the Nobel Prize for Economic Science in 1986 for his pioneering role in establishing contemporary public choice theory. He joined with Gordon Tullock to author the classic *Calculus of Consent* in 1962.
It should be straightforward then to apply a similar public choice theory to the voters. Such a “rational voter” model is simple to develop.

We define an economic rational voter model using axioms adapted from microeconomic theory, as well as a simplified notion of voter incentives.

**The Assumptions**

Voters hold the incumbent party responsible for economic performance.

This is a doubly-simplifying assumption. First, in the United States of America, a Republic with three branches of government and a federal system, it is not entirely within the President’s power to set economic policy for all states. Second, voters could view presidential elections as referendums on a number of issues—regional, personal, partisan, etc.

This assumption results in a model of voting as a referendum on the incumbent party’s record of economic performance during the president’s term in office.

Voters prefer more income to less income.

This straightforward axiom leads to a complicated data question, as there are a number of variables that can be used to measure current income, including real disposable income, per capita real disposable income, and median household income.

Voters prefer a less risky stream of income. We consider both inflation and unemployment as risks to income. First, a spell of unemployment can reduce the earnings of a family to below their expenses. The recurring focus of political advertising on social security benefits indicates that, at least from the point of view of candidates and their media advisors, voters care deeply about long-term income security. Second, note that inflation is associated with both economic uncertainty, and with loss of real income and wealth.

Thus, the components of the “misery index” of political fame, and the “Phillips curve” of macroeconomic interest, are components in the voters’ calculus. We discuss the proper specification for unemployment and inflation below.

---

3 We avoid here the contentious mathematical analysis of whether individual voters can expect to control policy through voting, a matter dealt with by another Nobel Prize-winning economist, Kenneth Arrow, in his “social choice” analysis. His famous “impossibility theorem” implies that, in many cases, it is impossible for straightforward voting to result in the best outcome for everyone. See Shaw (1999) for an accessible discussion of both public choice and social choice theory and full references in the endnotes.

4 Much of social research on political preferences is based on a broader, though similar, assumption that elections are a referendum on the performance of the incumbents, especially executives such as governors and presidents. For this reason, the “do you approve of the job [elected official] is doing” and “do you think the country is on the right track or wrong track” questions are a fixture in political polling.

5 We briefly consider below the role of asset returns in this question. However, we reject the notion of including stock market trends in the model, because the stock market reflects discounted expectations about the election.

6 An interesting question, which may be possible to analyze with the methodological innovations described below, is whether voters evaluate “permanent income” in any quantifiable sense. The rise of the “investor class” suggests that this should be the case, and that both returns on investment assets and long-term expected retirement benefits would be factors in voters’ economic calculus.

7 The uncertainty in income arises in both the policy changes that often follow increases in the inflation rate—wage and price controls, for example—and the negative effects on the real economy associated with growing inflation. Additional factors are the implicit taxation that occurs on assets whose nominal value has increased, and the higher nominal taxes due to “bracket creep” in many income tax systems.
Major institutional factors also govern voting behavior.

The first is war. Wars have been threats to national survival and have taken precedent in national policy for most of human history. This includes American history; one does not need to go back too far to find war bonds, de facto nationalization of industries during war time, and wage and price controls.

In this sense, the phenomenon of “limited war” is a new and unusual concept. The ancients would have scratched their heads in wonder at the United States of America which, during the first Persian Gulf War, actively debated reducing its defense budget.

Econometric models of the US economy have typically used war-time dummies to account for shifts in production during periods of war. The standard approach to predicting presidential voting behavior has been similar. We will examine in a later section whether this needs to be modified due to the advent of “limited war.”

Testing the Axioms

To test these axioms, we performed a series of exploratory data analyses, comparing the economic and institutional factors the axioms suggest will predict voting behavior with actual voting behavior.

Dependent Variable

For the dependent variable, we use the difference in the shares of the popular vote between the incumbent party and the challenging party in the general election for President of the United States. Thus, we are predicting the margin of victory or defeat, expressed as percentage, between the major party candidates.

The selection of this measure is complicated by two factors: the likely nonlinear relationship over the entire spectrum of potential outcomes and the presence of third-party candidates. We deal with the former by assuming that, within the range of outcomes for most contemporary elections (in which the incumbent party wins or loses by less than ten percent, and often less than 5 percent), the response by voters is close to linear. The latter issue is dealt with by calculating only the vote share gained by the major party candidates, and using an explanatory variable for significant third-party candidates. We discuss third-party candidates further below.

Income

A test of the income axiom, using the dependent variable outlined above, is straightforward. Of the last twenty-two presidential election years since 1916, the US economy realized annual nominal GDP growth rate more than ten percent nine times.

9 We experimented with truncating the data, which reduced the residuals in landslide election years, and could have transformed the variable using a log, logit, probit, or similar model. However, the models described below are pure linear models.

10 The third-party candidate vote totals were especially strong earlier in the 20th century, when Teddy Roosevelt’s “Bull Moose” party challenged the incumbent of his own party, and when socialist parties had especially strong followings.

11 For the equations, we will use a more specific income measure. GDP, however, is a handy and consistent series for use in the nearly century-long analysis presented here.
shows that of those nine years, the incumbent party won seven of the presidential elections.

**Figure 1:** Presidential Vote Margin vs. Annual GDP Growth Rate, 1916-2000.

**Income Risk: Inflation**

The historical record indicates that if the price level changes rapidly it is very difficult for an incumbent party to win the presidential election.

To better assess the effect of inflation on income security, we use the absolute value of the inflation rate as our measure. During the Great Depression, a sharp deflation was associated with a dramatic electoral reversal for the incumbent party, indicating the economic fundamentals about the price level—that unexpected price level changes, and changes that result in loss of income or wealth—are more damaging to individuals than a steady inflation rate.

Figure 2 shows the relationship between inflation and presidential voting. We had very high inflation rates (in absolute value, during the election year) six out of 22 times since 1916. In four out of six times, the incumbent party lost the seat. Even in the exception years, 1912 and 1948, high GDP growth helped incumbent parties to keep the seat.

**Figure 2:** Presidential Election Voting Patterns v. Inflation Rate, 1916-2000.

**Income Risk: Unemployment**

Unemployment, according to the axioms, is a threat to income, and therefore higher unemployment should directly result in lower vote margins for the incumbent.

It is straightforward to show that during times of high unemployment (such as the beginning of the Great Depression, and in 1980) the incumbent party is likely to lose and that, conversely, during times of low unemployment (the Eisenhower years, the Reagan landslide in 1984, the Clinton re-election in 1996), the incumbent party is likely to win.

However, the unemployment rate is strongly correlated with output and income growth rates, especially when these are viewed on an annual basis. Therefore, using both an income and an unemployment variable in the same equation will introduce multicollinearity, which can result in estimation problems. Furthermore, economic theory indicates that there is a “natural” rate of unemployment, and therefore implies that voters do
not penalize a president that keeps the unemployment rate near its natural rate. We return to this issue below.

**Record of Rational Voter Models**

The rational voter models that have been used in the past, such as those comprehensively described in Fair (2002) and presented recently by Anderson and Geckil (2004) have a fairly impressive ability to predict voting behavior, given that they rely solely on economic and institutional factors.

**Fair Model Specification**

The Fair model predicts the vote share of the incumbent party, using a linear model with the following variables:

- Per capita growth rate of real GDP during the entire term,
- Average inflation rate over the 15 quarters prior to the election,
- “Good news” quarters (number of quarters out of the 15 quarters before the election with per capita growth rate of real GDP of 3.2 percent or more),
- “President running” dummy (if the president is running for reelection),
- A “duration” variable (the number of terms the presidency has been held by the incumbent party),
- A party dummy variable, which varies with the party in office, and
- A (full-scale) war dummy variable, which is one in years of World War I, World War II, the Korean War, and the Vietnam War.

A model based on the Fair model was also developed by Macroeconomic Advisors (2004). Their data refinements include using annualized percent change of real disposable personal income over the three quarters immediately prior to the election quarter, rather than the GDP growth used by Fair. They also drop the “good news” quarters in favor of housing starts, indicating that investment decisions by private owners is a direct measure of economic confidence. Finally, they keep the same party and duration dummy variables. The predictive results improve slightly, even though the number of explanatory variables is reduced.12

**Anderson and Geckil Model Specification**

The authors developed a standard model (the A-G model) to model predict the difference in vote share between the incumbent and the challenging party, using a linear model with the following, smaller list of somewhat different variables (Anderson and Geckil, 2004):

- GDP growth rate during the election year,
- Absolute value of the inflation rate in the election year,
- Difference in unemployment from a “natural” 3.5 percent unemployment rate during the election year,
- A limited-war dummy (if a limited war happened during the term of the incumbent party),
- Third-party candidate dummies, for both incumbent and challenging parties.

**Results in 2000**

The results for these two models are fairly similar, so we consider each one from a different perspective. Table 1 shows the estimated Fair model, and its prediction for the 2000 election.

12 Using a short sample of 1952-2000, the equation has a smaller standard error, and a similarly high (.97) R-squared statistic.
Table 1: Fair Model and 2000 Elections

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Value</th>
<th>Coefficient x Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate</td>
<td>0.70</td>
<td>2.2</td>
<td>1.54</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.71</td>
<td>1.7</td>
<td>-1.21</td>
</tr>
<tr>
<td>Good news quarters</td>
<td>0.90</td>
<td>7.0</td>
<td>6.30</td>
</tr>
<tr>
<td>President running</td>
<td>4.00</td>
<td>0.0</td>
<td>0.00</td>
</tr>
<tr>
<td>Duration</td>
<td>-3.30</td>
<td>1.0</td>
<td>-3.30</td>
</tr>
<tr>
<td>Party variable</td>
<td>-2.80</td>
<td>1.0</td>
<td>-2.80</td>
</tr>
<tr>
<td>War variable</td>
<td>4.70</td>
<td>0.0</td>
<td>0.00</td>
</tr>
<tr>
<td>Intercept</td>
<td>48.40</td>
<td>1.0</td>
<td>48.4</td>
</tr>
</tbody>
</table>

The table indicates that the coefficients all have the expected signs, with positive coefficients for growth rate and “good news” quarters (reflecting the number of recent quarters with high growth); and negative coefficients for inflation, and having the same party stay in office beyond two terms. The war dummy variable is positive, reflecting the traditional “rally around the commander” pattern of Americans in pre-Vietnam conflicts.

For the 2000 election, using actual values, the Fair model predicts the incumbent party’s vote share to about one percent of the actual result, which was a nearly 50-50 split. We review the A-G model from a time-series perspective next, but note that it also predicts the 2000 election to about one percent of the actual result.

Results over Time: A-G Model

The estimated A-G model is shown in Table 2. The variables used in Tables 2 and 3 (discussed below) are defined as follows:

\[ VOTEDIFF: \text{Percentage point difference between the percent of votes received by the incumbent party and the percent of votes received by the major challenging party, i.e., margin of victory (Assuming that the sum of the percents of votes received by those two parties equal to 100 percent).} \]

\[ GDP_{LAST}: \text{Annual GDP growth rate during the last year of the presidential term.} \]

\[ U_{LAST}: \text{Unemployment rate of the last year of the presidential term.} \]

\[ INF_{LAST}: \text{Absolute value of inflation rate based on CPI during the last year of the presidential term.} \]

\[ INCMBTHRD: \text{A strong third party, sharing a similar philosophy with the incumbent party, challenging the incumbent party (e.g., Green Party in the 2000 presidential election).} \]

\[ CHLGTHRD: \text{A strong third party, sharing a similar philosophy with the incumbent party, challenging the major opponent party (e.g., Green Party in the 2004 presidential election).} \]

\[ WAR: \text{Limited war dummy (e.g. Vietnam and Gulf Wars).} \]
Table 2. Anderson & Geckil National Model:

Dependent Variable: VOTEDIFF
Method: Least Squares
Date: 06/14/04 Time: 13:03
Sample (adjusted): 223
Included observations: 22 after adjustments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.133712</td>
<td>0.047943</td>
<td>2.788981</td>
<td>0.0138</td>
</tr>
<tr>
<td>GDP_LAST</td>
<td>0.670068</td>
<td>0.255842</td>
<td>2.619074</td>
<td>0.0193</td>
</tr>
<tr>
<td>U_LAST</td>
<td>0.040549</td>
<td>0.412597</td>
<td>0.098277</td>
<td>0.9230</td>
</tr>
<tr>
<td>INF_LAST</td>
<td>-2.663252</td>
<td>0.502262</td>
<td>-5.302519</td>
<td>0.0001</td>
</tr>
<tr>
<td>INCMBTHRD</td>
<td>-0.088787</td>
<td>0.048604</td>
<td>-1.826753</td>
<td>0.0627</td>
</tr>
<tr>
<td>CHLGHTRD</td>
<td>0.094618</td>
<td>0.044474</td>
<td>2.127480</td>
<td>0.0504</td>
</tr>
<tr>
<td>WAR</td>
<td>-0.204056</td>
<td>0.062860</td>
<td>-3.246174</td>
<td>0.0054</td>
</tr>
</tbody>
</table>

R-squared     | 0.796343    | Mean dependent var. | 0.054051 |
Adjusted R-squared | 0.714880 | S.D. dependent var. | 0.147820 |
S.E. of regression | 0.078931 | Akaike info criterion | -1.987117 |
Sum squared residual | 0.093451 | Schwarz criterion | -1.639968 |
Log likelihood | 28.85829 | F-statistic | 9.775553 |
Durbin-Watson stat | 2.177262 | Prob(F-statistic) | 0.000178 |

The coefficient on growth is positive, as expected. The coefficients on institutional factors such as third-party challengers are also as expected. Inflation, shown as the absolute value of the rate, also has a negative coefficient.

However, a surprise lurks in the unemployment rate. After the role of income growth (represented by the GDP variable) is taken into account, as well as inflation, an increment in the unemployment rate above the “natural” rate does not have a statistically significant coefficient. Indeed, the estimated coefficient is positive.

Another surprise lurks in the limited war dummy. If we analyze the impact of “limited” wars, such as the Persian Gulf War in 1992, and the Vietnam War, the coefficient in this equation indicates that such wars produce a reduction in electoral support for the president’s party, rather than the increase found for full-scale wars during most of the century.

The A-G model was estimating for the years 1860 through 2000, producing a fairly consistent result, shown in Figure 3. For elections before 1920, the model does explain a large part of the variation in voter behavior but can’t be relied upon to suggest how voters will decide elections that are reasonably close. Furthermore, the swings upward and downward in voter preferences are not well captured by the model. On the other hand, from 1920 through 2000, this model, like the Fair model, consistently predicts the majority of the variation in electoral behavior.

Conclusion: Standard Models
We conclude that a standard rational voter model, using data from 1920-2000, is a useful guide to electoral behavior. In particular:

Some variables were not available in the same series this far back in history, so some splicing or substitution was necessary.
1) About three quarters of voter behavior can be explained by these models. To be more specific, both the Fair and A-G models have an R-squared statistic from a standard OLS regression of between .70 to .80.

2) This explanatory power is accomplished without using any sentiment variables.

3) The approach is fairly robust to minor specification differences, when the objective is the overall predictive capability, rather than “predicting” specific individual elections.

4) For most elections, about 15 out of 20, the model predicts a clear winner, and the voters elect that candidate. Using a simple batting average test, which gives points only for correctly picking the winner, even if the margin is grossly mis-estimated, produces the observation that in 18 out of the 21 past elections, these models predict correctly who will be the winner.14

5) Both models typically capture the amplitude in voter swings as well as the direction. For example, both of the models show large electoral majorities for FDR during the Depression/World War II years; anticipate close elections in 2000; and properly predict victories by small margins in years such as 1968 and 1952.

6) Both models do not correctly predict voter behavior in the years 1976 and 1992. In both cases, an incumbent president is not reelected, even though institutional and economic factors suggest that he should be.

7) The estimation of parameters, and the use of these parameters for predictions, are hampered by two structural problems: First, there are simply too few observations to estimate all the coefficients; a 1920-2000 presidential election data set includes observations from only 21 elections, which may be needed to estimate 5 or more coefficients. Second, we expect that voter’s reaction to economic conditions would change over time. This time-varying preference might be ignored over one to three decades; it cannot be ignored over an entire century.

Sentiment-Based Models

There are other models that incorporate economic information, which we do not consider comparable. These models include sentiment variables, such as approval ratings from polls. Examples include Wlezian and Erikson (1996), Wlezian (2000), and many of those summarized in Greene (1993). While adding sentiment variables improves the forecasting performance, it undermines the microeconomic motivation for the rational economic voter model. Viewed from the perspective of determining what effect pocketbook issues have on voters, such models have circular reasoning: one set of sentiment indicators (poll results, political support) predicts another expression of sentiment (the election results), with the aid of economic information that also affects sentiment. These models could be useful in at least two ways: first, interpreting how new economic

14 Depending on how one counts the 2000 election, predicted to within about one percent by both models, it might be 19 out of 20.
information translates into changes in voter sentiment over time; and second, as pure predictive models.

Perhaps the most serious effort to develop a rigorous economic model that included sentiment variables is described in Stambough and Thorson (1999). Their model uses a principal components approach to isolate the underlying economic conditions and overcome multicollinearity problems. They also use a form of pooled data, which anticipates the innovation described in this paper for a contemporary model. However, the Stambough and Thorson model includes variables such as past electoral behavior in a state, as well as indicators of political support for candidates. It therefore is not a pocketbook model, but instead one in which the pocketbook variables supplement information from polls and other expressions of political support.

Sunspot Models

In addition, there are numerous back-of-the-envelope models based on stock-market returns, consumer confidence ratings, or interpretations of poll results. These, in turn, are supplemented by even more ersatz models based on comparative candidate height and the league winning the World Series. These run the gamut from the purely entertaining to the modestly useful, from the perspective of wagering on the outcome of the election. However, for the same reasons we rejected the use of sentiment variables in our pocketbook model, we did not include sunspot activity and other events that correlate, one way or another, with electoral victories.

Innovations in Pocketbook Models

Addressing Problems in Past Models

To address some of the weaknesses of the past models, we suggest two sets of innovations: the use of pooled data from multiple states, and better-specified variables.

Use of Pooled Data

The key econometric challenge in national pocketbook models is too few observations. This, combined with structural changes in the relationship over the past century, has made it difficult to properly estimate the parameters of an equation specifying how economic conditions affect voter behavior. The best method of attacking this problem is to bring more data to the subject. This can be done by treating elections as generating a pooled set of data on election returns for 50 states. Thus, each election is a cross-section of data from one time period. Multiple elections over a handful of decades then produce a time series set of these cross sections. By limiting the elections we consider to those occurring in a specific era, we reduce (but do not eliminate) the effects of structural changes over time.

This econometric insight dovetails with the observation that voters in individual states may react quite differently to economic news.

Improved Specification

15 The estimation algorithm used was not described in the paper, but it appears some type of varying constants or coefficients were allowed for individual states.
In addition, with insights from A-G (2004) and Macroeconomic Advisors (2004), we attempt to refine our specification. From Anderson and Geckil (2004), we adopt as variables the absolute value of the inflation rate to account for the negative feedback about deflation; and a “limited war” dummy variable. For a national model, we believe the Macroeconomic Advisors selection of an improved disposable income measure enhances the equation, but the measure could not be used here for state income variables.

In addition, we use the deviation from the “natural” unemployment rate as our indicator, rather than its absolute value.

Finally, although it is clear that adding sentiment variables such as polling or recent election results would improve the forecasting accuracy of the model, we restricted the type of variables to economic and major institutional factors.

Estimating a Pooled State Data Election Model

Data
An ideal pooled data set would contain individual and different explanatory variables for each subject in the cross section, for each time period. We approach this ideal, but do not reach it.

We assemble a data set that includes separate observations on unemployment, as well as a state-level income variable, for all states. We also include dummy variables for third-party candidates.

For income growth, we used the annual change in gross state product during the election year. We note this is not precisely income, and will discuss the point further in the econometric section, below.

For third-party candidates, the use of individual state data allows for a more precise specification. States vary in the restrictions they place on third-party ballot access, so we are better able to specify when a significant third-party candidate participates by using state data than national data. For example, in three of the major states, a third-party challenger to the incumbent did not gain ballot access during the entire sample period; all states had third party candidates on the challenger side.

For state-level price data, we used the CPI for large metropolitan areas in each state, if available. For example, for California, we used the average CPI for Los Angeles, San Diego, and San Francisco. For Indiana, no metropolitan CPI was available, so we used the average of Cincinnati and Chicago.

The CPI has well-known limitations as a price level indicator, and city CPI data would only partially reflect the actual inflation rate experienced by residents across the state.

For unemployment by state, we had excellent data by state from the US Bureau of Labor Statistics. In the equation, we use the deviation from a natural unemployment rate of 3.5 percent.

---

16 Leip (2004) was the source for much of these data.
17 In particular, the overweighting of housing price data would tend to exaggerate the price changes faced by the residents that do not purchase a new home; the use of a “consumer” basket will also not match the price level effects across the entire economy. This latter bias may, however, make the variable a better indicator of how most voters perceive inflation’s effect on their lives.
18 A potential improvement to this approach would be to estimate the natural unemployment rates of individual states. However, since we are using the arithmetic difference from the natural unemployment rate as the explanatory variable, such an enhancement would largely affect the estimated constants.
For the limited war dummy, we use the same variable for all states. Although we did not include it in the general model, it is clear that the “home state” advantage for a candidate is a powerful factor, and one that could almost be considered an institutional one.\textsuperscript{19}

While this is clearly a richer data set than used in the national time-series models, there is a systemic problem with much of these data that must be considered. For the income and price data in particular, the different observations across states are strongly correlated. Even with the efforts of the state and national statistical agencies to improve state-level data, some components of these variables are shared across all states.\textsuperscript{20} Thus, state-level data bring more information to the subject, but are likely to have measurement errors that are strongly correlated across the country.

\textbf{Econometrics}

Estimating models using pooled data is tricky. Such models can potentially have additional constants for every subject in the cross-section, as well as interaction variables for each of the explanatory variables, to account for different responses in each subject.\textsuperscript{21} Attempting to estimate all the possible multiple interaction variables and varying constants would quickly use up the additional degrees-of-freedom obtained by assembling the pooled data set.

In addition, given the strong correlation of data such as income and employment across subject states, we are likely to face multicollinearity if we attempt to estimate too many parameters. This is compounded by the observation that the measurement errors in these data are also correlated across states. Indeed, trial estimation runs with a large number of state-varying parameters produced warnings of singularity, signaling that multicollinearity in the data is a serious problem.

We therefore follow the classic approach of restricting the number of coefficients we attempt to estimate. In particular, we allow for varying constants across all states, but do not estimate varying coefficients for each explanatory variable in our general pooled state model. It is possible that further econometric work would yield an approach, perhaps supplemented with a richer set of state-varying data on prices and incomes would yield a model with well-estimated behavioral coefficients that vary across states.

\textbf{Estimation Results}

The results of our estimation of the pooled state model are shown in Table 3. We allow the constant term to vary among the various states. This allowance allows for a straightforward social science intuition from the data, as some states immediately appear to be more “incumbent friendly” than others. The coefficients on the behavioral variables are restricted to be the same across all states, and can be compared directly with the A-G national model results shown in Table 2.\textsuperscript{22}

\textsuperscript{19} We do consider this when testing the model results against the 2000 election.

\textsuperscript{20} For example, income at the state level must reflect import and export activity, which must be apportioned across all states; survey data on labor market activity will invariably reflect fewer observations in individual states than in the nation as a whole; and price level data from indicative markets are invariably used, in both the government and private sectors, as national prices.

\textsuperscript{21} In addition, the presence of serial correlation and heteroskedasticity in the disturbances, which is often a problem even in non-pooled data, can create an even more difficult estimating environment.

\textsuperscript{22} The A-G national model used the unemployment rate rather than the difference from the natural rate; see the
Table 3. Pooled State Model:

Dependent Variable: VD?
Method: Pooled EGLS (Cross-section weights)
Date: 08/04/04   Time: 09:37
Sample (adjusted): 1 6
Included observations: 6 after adjustments
Cross-sections included: 20
Total pool (balanced) observations: 120
Linear estimation after one-step weighting matrix

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.019823</td>
<td>0.03388</td>
<td>0.585089</td>
<td>0.5599</td>
</tr>
<tr>
<td>GSP_GR?</td>
<td>1.194701</td>
<td>0.296542</td>
<td>4.028781</td>
<td>0.0001</td>
</tr>
<tr>
<td>UN?</td>
<td>1.525821</td>
<td>0.536932</td>
<td>2.841739</td>
<td>0.0055</td>
</tr>
<tr>
<td>CPI?</td>
<td>-2.049109</td>
<td>0.220464</td>
<td>-9.29446</td>
<td>0</td>
</tr>
<tr>
<td>WAR?</td>
<td>-0.17346</td>
<td>0.027794</td>
<td>-6.240905</td>
<td>0</td>
</tr>
<tr>
<td>INCTRTHRD?</td>
<td>0.055749</td>
<td>0.024054</td>
<td>2.317636</td>
<td>0.0226</td>
</tr>
</tbody>
</table>

Fixed Effects
(Cross)

|  _AZ--C | -0.028668  |
|  _CA--C | -0.025448  |
|  _FL--C | 0.020369   |
|  _GA--C | 0.020613   |
|  _IL--C | -0.003166  |
|  _IN--C | -0.015066  |
|  _MA--C | 0.032234   |
|  _MD--C | 0.010934   |
|  _MI--C | 0.013913   |
|  _MO--C | -0.009124  |
|  _NC--C | 0.002462   |
|  _NJ--C | 0.055089   |
|  _NY--C | 0.020478   |
|  _OH--C | 0.000172   |
|  _PA--C | -0.020849  |
|  _TN--C | 0.005245   |
|  _TX--C | -0.059317  |
|  _VA--C | 0.010893   |
|  _WA--C | -0.027395  |
|  _WI--C | -0.00337   |

Effects Specification
Cross-section fixed (dummy variables)

<table>
<thead>
<tr>
<th>Weighted Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
</tr>
<tr>
<td>S.E. of regression</td>
</tr>
<tr>
<td>F-statistic</td>
</tr>
<tr>
<td>Prob(F-statistic)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unweighted Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
</tr>
<tr>
<td>Sum squared resi</td>
</tr>
</tbody>
</table>

We observe that the additional data from a much shorter period, in a similar model allowing for varying constants across states, produced an equation almost the same explanatory power overall. In both the long-term national model, and the single-era pooled state model, the R-squared statistic indicates that about 75 percent of the variation in voting can be explained by a small number of economic and institutional variables.23

The residuals from the regression for all 20 states across the 1980-2000 do not appear to have any particular common patterns.24

Variation Across States
These results suggest that some states are historically more “incumbent friendly.” The different constants estimated for each state suggest that, for example, voters in New York, New Jersey, and Massachusetts have given incumbents a leeway of two to five percent more than the typical voter in a large state.

We further consider this result, and test the power of the model to capture how economic conditions in separate states affect their voting behavior, using a straightforward test. For the year 2000, we tabulated the model’s output for individual states. Note that these are not 20 individual state estimates, but individual state differences derived from the estimation of a national model.25

23 The R-squared statistic should be interpreted with some caution in the pooled regression, but given the relatively small number of state-varying coefficients, the caution does not undermine the observation that, within a reasonable margin, both models explain about 75% of the variation.
24 Graphs of these residuals can be obtained from the authors on request.
25 Recall that this is a national model estimated with pooled data, which allows for differences in the incumbent advantages among states. Thus, the vote

discussion below under “unemployment rate and the incumbent advantage.”
We call these the “economic vote factors” and list these data in Table 3. The factor should indicate how individual state economic conditions, based on both the state incumbent advantage and the national patterns of voter response to economic conditions, are likely to affect the vote margin in each state. We then compare these vote factors with the actual vote margins in these states. If the model had no predictive power, we would expect the correlation between these data to be around zero. In the year 2000, the correlation between the economic vote factors by state and the actual vote margins was .67.

Unemployment, Growth, and the Incumbent Advantage

We observe that the estimated coefficient on the deviation from natural unemployment is positive. This appears to run counter to our axioms, as a more risky stream of income should encourage voters to reject the incumbent party. However, a further review is required.

First, we include both an income growth variable (in the pooled equation, gross state product growth) and an unemployment variable in the equation. Thus, there are two explanatory variables with a strong correlation. This invariably leads to less precision in the estimates, and makes the equation sensitive to small differences in variable definitions. Second, repeated trial regressions using slightly different factors calculated in the table are not individual state predictions. However, if one were to extend this model in the direction anticipated in the discussion above under “Econometrics,” and develop both the data and the estimation techniques capable of separately estimating multiple coefficients for multiple states, it would be possible. See the note below about the results of specifications confirm that, once income is included in the pooled state equation, the additional explanatory power of an unemployment variable becomes small, and in some cases, statistically insignificant. This is the case with the A-G national model shown in Table 2. Perhaps for this reason, the Fair model shown in Table 1, and closely related Macroeconomic Advisors model, include an additional growth variable and omit unemployment entirely.

If an unemployment variable is included, its specification directly affects the equation’s estimated incumbent advantage. Equations that use the absolute level of unemployment and have a negative coefficient often have a high constant variable. For example, the A-G national model shown in Table 2, indicates a constant and a negative coefficient on the unemployment rate variable. If one transforms the unemployment rate by subtracting the natural rate, the constant should shift closer to zero. This more correctly indicates both a nominal incumbent advantage, and an unemployment variable that provides a very small contribution once growth is considered.

Historic State Preferences

It is important to note that these results were obtained without any sentiment variables, and also without taking into account the obvious fact that some states are more likely to vote for one party or the other. The sample period here includes elections in which both Republicans (1980, 1984 and 1988) and Democrats (1992 and 1996) won repeated elections. The simple predictive power would certainly be improved by including a state-level variable for the historic partisan

Stambough and Thorson (1999), which were derived using additional sentiment variables.
preferences of voters in each state. This may be a useful extension of the research presented here, and could be considered a primarily pocketbook model if there were only a single, historical preference variable applied across all periods for each state, rather than a group of sentiment variables.

**Prediction for 2004**

**Factors Missing in the Models**

In the discussion above, we reviewed the record of the national time-series model described by Fair (2002), as well as the variant published by Macroeconomic Advisors (2004) and a somewhat different model presented by Anderson and Geckil (2004). These models, as well as a number of variations analyzed by the authors in the research leading to this article, do surprisingly well in predicting the variation in the vote for the incumbent party, considering that they do not include any sentiment variables such as poll results. As a confirmation of this, all three models predict the very close 2000 election to within about one percent of the actual vote margin.

However, these models typically fail to predict the election results in 1976 and 1992, in each case forecasting that the relatively positive economic conditions would result in an incumbent victory. Repeated efforts to improve the model by refining the economic variables have been reported in Fair (2002), criticized by Greene (1993), and replicated by the authors. Fair’s use of “good news” variables, for example, is a straightforward attempt to make economic conditions explain what (to economists, at least) appear to be recalcitrant voter behavior.

A better explanation is necessary. We offer the following: Voters are not driven entirely by economics. There are strong axioms of microeconomics which, adapted to this problem in a manner of analysis familiar to Public Choice students, produce the theoretical grounds for a rational economic voter model. These models, in both time series and pooled cross-section-time series versions, provide a very powerful explanation for well over half of the variation in voting behavior.

For models that eschew sentiment variables, such predictive power (often resulting in R-square statistics of 75 percent or better, and “batting average” metrics of 18-out-of-21 correct predictions) should be accepted as nearing the limit of what social scientists can do with unpredictable human behavior.

Furthermore, the “wrong” elections provide us with strong clues as to what non-economic variables are most important. In 1976, Watergate, the resignation of President Nixon, and voter revulsion over the scandal were clearly dominant themes. In 1992, the end of the Cold War changed the role of the President of the United States.

As pointed out by an eminent national pollster, once the Cold War was ended, voters were no longer required to pick a President knowing his conduct could be tested by Cold War confrontations like the

---

26 If we added multiple variables of this type, and further supplemented it with political support variables, we would be following the approach in Stambough & Thorson (1999).

27 Somewhat less powerful examples are the landslides of Reagan in 1984 and Johnson in 1964. These victories were aided by economics, but were obviously motivated by other issues.
Cuban Missile crisis.\textsuperscript{28} Thus, the elections from 1992 through 2000 were conducted during an interregnum that removed from many voters minds the imperative of selecting a person with a “commander in chief” character that probably seemed vital to the generations that voted during the eras of World War II and the Cold War.

Tragically, the terrorism of September 11, 2001 ended that interregnum. Thus, future elections—particularly the 2004 election—are likely to be decided on larger factors than economics.

\textbf{National Presidential Election Prediction for 2004}

With this observation in mind, we nonetheless calculate the forecast for 2004, using the pooled state economic model.

Using estimated variables for 2004 on a state-by-state basis, and using the varying state constants and other coefficients estimated in the model, the pocketbook prediction for 2004 in major states is the following:

\begin{itemize}
  \item The popular vote winner will be the Republican Party in 2004 by a narrow 0.5 percent margin.
  \item The strongest influence on the results is the relatively strong income and employment gains for the last two years, along with a low inflation rate.
  \item The painful limited war going on in Iraq is an institutional factor that will reduce the support of the incumbent party.
  \item The only significant third party candidate is likely to be Ralph Nader, who will be on the challenging party’s side.
\end{itemize}

By comparison, Fair (2002, 2004) expects a stronger Bush victory, as do Macroeconomic Advisors (2004). Of course, all these predictions are based on economic conditions assumptions that may prove to be incorrect, and are based on models that explicitly include only economic and institutional variables.

Although these economic models, including our own, anticipate a Bush re-election, we note the conclusion drawn above: voters do not only care about economics. In particular, during times when the world order becomes more dangerous, their dominant influences appear to be related to national interest and national survival, rather than improving their current economic condition. It seems apparent, even to economists that insist on keeping poll data out of their pocketbook models, that 2004 is such a year.

\textbf{Conclusion}

Our results indicate that voters do respond to economic conditions in a manner consistent with a small set of axioms drawn from microeconomics. Furthermore, rational voter models can be specified and estimated in various forms that predict about 75 percent of the variation in vote over elections from 1916 through 2000.

Our results show that a national model using pooled state data, estimated over a shorter period of time, can avoid some of the econometric problems that have hampered past time-series models. Using this model, we find that voters in different states vary in their responses to both economic factors and institutional factors, and are significantly different in “incumbent

\textsuperscript{28} David Petts, of Petts & Blumenthal, Washington DC; private conversation.
friendliness.” As a test of our analysis, we estimate the economic contribution to the state-by-state vote in the 2000 election, and find a positive correlation of approximately 66 percent with the actual vote margins.

Our pooled state model forecast for 2004 indicates that economic conditions favor a narrow re-election for the incumbent President. However, we point out how some elections—such as the loss of George Bush to Bill Clinton in 1992, as well as the upcoming 2004 election—cannot be entirely explained with a rational voter model. In dangerous times, voters appear to place more emphasis on national survival and national interest than on current economic conditions. Thus, our research both demonstrates the power of economics to predict voting behavior, and its limitations as well.

REFERENCES


