Undergraduate Economic Review

Volume 8 | Issue 1 Article 11

2011

Does the Economy Determine the President? A Regression Model For Predicting US Presidential Elections

Roy K. Roth
Brigham Young University - Utah, roykroth@gmail.com

Does the Economy Determine the President? A Regression Model For Predicting US Presidential Elections

Abstract

There is a prevalent belief that the economy determines the President. If the economy is good, the President keeps his job, if it is bad, he is out. A large body of econometric literature has been published on this topic. This paper takes a new approach. I look not at how the popular vote changes with economic conditions, but how the electoral vote changes. I further examine how these changes affect the probability that the incumbent party stays in office. I find that economic conditions may not be as important as they have been purported to be.

Keywords

President, Elections, Econometrics, Economy

1 Introduction

Many Americans believe that the economy determines who wins the presidency. There is a large body of work dedicated to the effects that the economy has on presidential elections in the United States. This paper takes a slightly different approach, and presents important insights. I find that while economic conditions are important, they are usually overshadowed in elections where an incumbent faces a challenger.

The majority of work done in this area, including Fair (1972) and Kahane (2009), use the percentage of the popular vote as the dependent variable of their econometric models. In this paper I use two models that supplement the existing literature on this subject. The first estimates a vote-share equation with the percentage of the electoral vote, rather than the popular vote, as the dependent variable. The second estimates the probability that the Democratic candidate in any particular election will win. My approach has two clear benefits. First, these methods give clear predictions of the winner. Equations giving a prediction of the share of the popular vote sometimes fail in this area. The results of the 2000 election give an example of why this is desirable. Gore, the Democratic candidate received the majority of the popular vote, but received fewer electoral votes than Bush. Gore lost the election by a small margin. In that situation even a perfect prediction of the popular vote would be misleading. The second advantage regards significant third-party candidates. In 1992 and 1996 elections, Ross Perot played a major role as a third party candidate, receiving 18.9% and 8% of the popular vote, respectively. This could seriously affect vote-share models based on a two-party system. Electoral votes, however, are given (with a few exceptions) only when a candidate receives the largest share of the popular vote in a state, and are thus less sensitive to third-party candidates. Even Perot failed to receive a single electoral vote. In fact, during the elections covered in this paper, no third-party candidates received any electoral votes.

In this paper I use a small number of economic indicators, coupled with measures of the tenure of the political party in the White House and an indicator of whether or not the incumbent president is running for reelection to make predictions of the election winners. I then compare the predictions of this model to the actual outcomes to determine if the model is has reliable predictive power. This is of particular interest because we do not take into account other political factors such as campaign donations, endorsements, political blunders, etc. I do not claim that these are unimportant; however, we test the assumption that the state of the economy and the pursuit of reelection are the more important factors. I leave open the possibility that some of these factors, possibly including campaign donations and endorsements, are a function of economic and incumbency conditions.

The remainder of this paper is organized as follows. Section 2 presents an overview of some of the significant work already done in this area. Section 3 then presents the model I test in this paper. Section 4 describes the data I use to test my hypothesis. Section 5 presents the result of my econometric analysis. Section 6 presents my conclusions as well as suggestions for further research.

2 Background

The most well-known econometric work in elections is by Ray Fair. In his 1978 paper he begins with a basis of utility theory and builds a model in which a voter considers current economic conditions and his or her expected utility under each of the candidates; the voter then votes for the candidate under which he or she has higher expected utility. Fair then uses national data on economic indicators, most notably the growth rate of GNP and inflation as measured by

the growth rate of the GNP deflator to test the validity of his model. His predictions are quite accurate.

Kahane takes a similar approach using panel data at the state level. This presents several advantages over data at the national level, as well as several drawbacks. The first advantage is that the sample size necessarily increases. There are 50 states from which to gather data in each presidential election, as opposed to one nation. Thus there are 50 observations for each election year rather than one. Another advantage is that with more observations, there is more possibility for variation in the explanatory variables, increasing the reliability of estimators. There are also important drawbacks. For many of the desired explanatory variables, data is only available at the national level, therefore proxies are needed at the state level. While proxies can be very useful, there is always some loss of accuracy when they are used. Kahane finds that economic conditions are significant in determining the election outcomes (2009).

3 Model

I hypothesize that the economic conditions immediately preceding a presidential election have a strong predictive effect on the outcome. There are at least two ways that these conditions could affect voting behavior. The most immediate plausible assumption is that voters consider how their own economic health has been affected during the tenure of the current presidential incumbent, then vote for the incumbent candidate or party if the change is favorable, and vote against the incumbent otherwise. However, there are other possible explanations. Another possibility is that voters exhibit sociotropic voting patterns. Kinder and Kiewiet (1981) explain "Purely sociotropic citizens vote according to the country's pocketbook, not their own. Citizens moved by sociotropic information support candidates that appear to have furthered the nation's economic well-being and oppose candidates and parties that seem to threaten it." There may indeed be more avenues through which economic conditions affect voters' behavior. We do not hypothesize nor do we test how the effect enters into the voters' minds. We simply analyze the observable effects.

I test the hypothesis that an important economic indicator in presidential elections is the growth rate of real production. Production growth could affect voters in at least two ways. The most clear way is through their own wealth. Though increased production will not affect everyone equally, it is a good indicator of the aggregate economic well-being of US voters. The other way that it could enter into voting decisions is through the media. Voters, especially voters who are up to date on the news, hear media reports of the health of the economy. A common indicator reported is the growth of production. This second way could be particularly important if the idea of sociotropic voting behavior is believed. In that case, even when an individual voters are in desirable economic circumstances, reports of low production growth could lead those voters to vote for the non-incumbent candidate or party.

Inflation and unemployment are also important economic indicators, and are plausibly important in determining election outcomes. Americans cringe when inflation is higher than what is considered good and healthy. We would therefore expect that the incumbent party would be less likely to win an election when inflation has been high. However, deflation is just as bad and as dangerous to an economy as inflation. The absolute value of the inflation rate thus seems the most appropriate variable to use in this context. The unemployment rate may be the most reported economic indicator. It is also plausible that it has the largest effect on individual voters well-being (at least for those who are unemployed), and on the decisions of sociotropic voters. However, the Current Population Survey, the survey from which the unemployment rate is

determined, did not begin until 1940. Using this data would limit our already small sample size. We will therefore not include unemployment in our analysis, but will leave that work until more data becomes available.

There are of course other factors that must be taken into account. It is reasonable to believe that there is a fundamental difference in elections where an incumbent is running for reelection against a challenger, and elections where two non-incumbents are running. It is also logical that if one particular party has been in control for a long time that there is a different election atmosphere.

A key assumption of my models is that the contest between Republicans and Democrats is a zero-sum game. In other words if the economy is strong, and a Democrat is in office, then the Democrat in the next election should have an advantage. The same is true if there is a Republican in office. To account for this contest, values of the explanatory variables will multiplied by *I*, a dummy variable equal to 1 when a Democrat is in office and -1 when a Republican is in office. Taking these factors into account, I believe that an equation for vote-share can be written as

$$Prctel_i = \alpha + \beta_1 I * gProduction_i + \beta_2 I * Inflation_i + \beta_3 Reelect_i + \beta_4 Terms_i + u_i \quad \textbf{(1)}$$

where $Prctel_i$ is the percentage of electoral votes received by the Democratic candidate. The other variables have their intuitive interpretations, and are more fully specified in the data section. Similarly, a probability model can be specified as

$$P(Demwins_i) = \alpha + \beta_1 I * gProduction_i + \beta_2 I * Inflation_i + \beta_3 Reelect_i + \beta_4 Terms_i + u_i \quad \textbf{(2)}$$

We will test these models for their statistical significance as well as for the accuracy of their predictions.

4 Data

I use data for the elections from 1932 to 2008, inclusive, for a sample size of 20 observations. My analysis closely follows the work of Fair (2009), however, there are important changes. The first is the change of dependent variable. The benefits of this change have already been discussed. There is at least one significant drawback. It is that electoral votes are not a measure of the percentage of votes cast by individual voters, but instead votes cast on behalf of states. Electoral votes are assigned to states based on population. The results of any analysis therefore cannot be interpreted as the effect of economic conditions on individuals voting behavior. They can only be interpreted as the effects that those same conditions have on electoral votes received, an important distinction. As the 2000 election demonstrates, those are not always the same thing. Likewise, the probability equation says nothing about margin of victory, or votes received. Notwithstanding this weakness, electoral votes and probabilities are useful measures in this study, and will add a new approach to the extant literature on this subject.

There are several ways to measure the growth rate of production. Fair, in his original 1978 paper, uses the average growth rate of real GNP over the two year period immediately preceding the election. Kahane, using state level data, is uses a slightly different measure—change in per capita income. We use the growth rate of real per capita GDP. We use GDP because it is the primary and official estimate of United States production. It is also the measure most commonly reported, and is easily accessible on the BEA website. The time period is the first three quarters of the election year. This is the same measure used in Fair's 2009 paper. This approach provides several advantages. First, using data for the first three quarters of the election year, measured at an annual rate, provided more satisfactory results in my regressions than any

other time period. It also seems reasonable that voters base their decisions on more recent economic developments. Per capita GDP is used rather than aggregate GDP because there was a very large change in the population of the United States over the time specified. Per capita GDP therefore seemed a more appropriate measure.

Inflation is also measured the same as in Fair (2009). The measure is the absolute value of the growth rate of the GDP deflator in the first fifteen quarters of the incumbent administration, except in 1944 and 1948, where the values are zero. This was chosen for two reasons. The first is that by using the same data as Fair, we are able to see if his results hold when vote-share equations are specified differently. The second is that the time period is appropriate. The time period allows a sufficient period for changes in the price level to affect voters. The GDP deflator was chosen rather than a measure of CPI to be consistent with Fair (2009).

Political measures are straightforward. The variable *Reelect* is equal to 1 if the incumbent president is a Democrat and is running for reelection, -1 if the incumbent president is Republican and running for reelection, and zero otherwise. The variable *Terms* is similarly defined. It is equal to the number of consecutive terms that the president has been of the same party as the incumbent. It likewise takes on a positive value if the president is a Democrat and a negative value if the president is a Republican. Full explanations of the variables, as well as summary statistics are listed in the appendix as Table 1.

5 Results

I first estimate equation (1), the vote-share equation. A full listing of the estimates is listed in the appendix as Table 2. The first and second equations in Table 2 were estimated by OLS. To account for possible heteroskedasticity, robust standard errors are reported. In equation (1) the model is estimated using all of the variables listed. The resulting equation predicts 16 out of the 20 elections correctly. Equation (2) omits the inflation rate and estimates the vote-share model using only the other 3 explanatory variables. The result is slightly more accurate, predicting 17 elections correctly.

The estimated coefficient on *I*Inflation* is 1.353 in (1), though it was not statistically different than zero at any reasonable significance level. Despite this, the positive coefficient seems counterintuitive. It also contradicts the results of Fair (2009) and Kahane (2009). We therefore re-estimated equation (1) using LAD rather than OLS to account for any possible influential observations. The sample size of 20 is rather small for LAD estimation, but the predictions of this model are reliable—17 out of 20 predictions were correct—enough to provide some validity to the estimation. The coefficient estimate for inflation in this equation is -1.821; but it is likewise not statistically different than zero. It does, however, seem more intuitive, and is more in line with the published literature on the subject. We therefore cannot say anything conclusive about the marginal effect of a change in the inflation rate on *Prctel*. The same is not true for *I*gProduction*, which was highly statistically significant in all three regressions. The estimated marginal effect is also intuitive, and in line with previous work on the subject (see Fair 1978; Fair 2009; Kahane 2009).

A surprising result in all three regressions is the size of the marginal effect of the variable *Reelect*, which had a marginal effect of greater than 25% of the electoral vote in each estimated equation. However, that coefficient alone can be misleading. Due to the way in which *Terms* is specified, whenever a president is running for reelection, the value of *Terms* will be at least one in absolute value (when there are Republican incumbents the value will be negative). Because

the coefficient estimate for *Terms* is negative, the net effect of running for reelection will be smaller than coefficient estimate for *Reelect* alone would suggest. Nevertheless, the effect is large enough, together with the constant, to overshadow almost any other conditions and predict an incumbent victory. This seems somewhat extreme; however, it is historically accurate. In the elections covered in this paper there were 13 elections where an incumbent was running for reelection. Only three lost.

Estimation of the probability equation (equation (2)) led to similar results to those of the vote-share equation, but with slightly more accurate predictions. Table 3 contains a full listing of the estimates. The results were interpreted as a prediction of a Democrat win whenever $\hat{p} > .5$ and a Republican win otherwise. Three of the four estimated equations ((1), (3), and (4)) made predictions with 90% accuracy. Equation (1) was simply estimated by OLS using robust standard errors. Due to the inherent heteroskedasticity associated with the linear probability model, equation (1) was re-estimated by WLS ($\text{weights} = \frac{1}{\hat{p}(1-\hat{p})}$). However, there were 3 predicted probabilities greater than one, and one predicted as negative. The estimated probability less than zero was replaced with .001 and those greater than one were replaced with .999 to make possible WLS (specific values of .999 and .001 were chosen by the author). The estimate of this equation is listed in Table 3 as equation (2).

The OLS and WLS estimates made several predictions out of the [0,1] range. For this reason, and for robustness, the probability model was also estimated by probit and logit regressions. These estimates are listed in Table 3 as equations (3) and (4). The coefficient estimates listed are evaluated at the mean for each respective explanatory variable, with the exception of the constant term, where such a report would be meaningless. These equations, together with the OLS estimate of the LPM, were the most historically accurate, predicting 18 of the 20 elections correctly.

Importantly, the same effect found in the vote-share predictions were also found here. The marginal effect of running for reelection is again large enough to overshadow almost any other conditions. Though, once again, the effect is somewhat counteracted by the opposite effect of the variable *Terms*. These results are not found by Fair (2009) or Kahane (2009), though they do estimate a positive marginal effect of running for reelection. My findings simply estimate that effect to be much larger.

A full listing of each model's predicted winner for each election is listed as Table 4 in the appendix.

6 Conclusion

My results are mixed. The economic variable most consistently significant in our regressions was the growth rate of real production, which was statistically significant in each of our electoral vote-share equations. However, it was not significant in any of the probability equations. But, the estimated effect in each equation was positive, lending credibility to the theory that economic growth matters in elections. Inflation, on the other hand was not statistically significant in any of our estimated equations. Furthermore the sign of its estimated coefficient was not consistent across our estimations. This seems to indicate that inflation is not a key economic variable to consider. Perhaps there are other variables that would be more significant.

The strength of the models is shown in the accurate predictions that are made. The probability models were the most reliable, predicting 90% of elections in the time frame

correctly. When more data becomes available, re-estimation of these models should yield more statistically significant results.

The models presented in this paper are designed to be a supplement to the already published literature on the subject. However, this is by no means the end of the topic. There is still more work to be done in the field of understanding and predicting presidential elections in the United States. An immediate extension is to use other measures of economic growth (such as stock averages) and inflation (CPI) and see if the results are consistent with those presented in this paper. Measures of unemployment may also shed light on this issue and result in more accurate predictions, especially as time passes and more data becomes available.

The most significant contribution of this paper is that perhaps economic indicators are not as significant as they have been portrayed. They seem to have very significant effects on the share of the popular vote (see Fair 2009), however, this research indicates that their effects are less significant in actually determining the winner of an election. The advantage that an incumbent has over a challenger is usually large enough to overshadow economic conditions. As a result this research provides some justification of term-limits. Further research in this area should reinforce these findings.

Appendix

Table 1:

Descriptive Statistics

Years 1928 through 2008 (n=20)

Variable	Description	Mean	Std. Dev.	Min	Max
Demwins	Equal to one if the Democratic candidate wins the election and zero otherwise.	.6	.503	0	1
Prctel	The percentage of electoral votes received by the Democratic candidate.	54.930	30.799	2.246	98.493
gProduction	The growth rate of per capita GDP in the first three quarters of the election year. measured at an annual rate	2.028	5.136	-14.586	11.836
Inflation	The absolute value of the average rate of inflation over the first 15 quarters of the current presidential term, except for 1944 and 1948 when the value is zero.	3.054	2.370	0	7.864
Reelect	Equal to 1 if the current president is a Democrat and is running for reelection. Equal to -1 if the current president is a Republican and running for reelection. Zero Otherwise.	.05	.826	-1	1
Terms	Equal to 1 (-1) if the Democratic (Republican) party has been in the White House for one term. 2 (-2) If two terms, and so on.	.25	2.268	-3	5
I	Equal to 1 if the current president is a Democrat and -1 if the current president is a Republican.	0	1.026	-1	1

Table 2:
Estimation of *Prctel*

	(1)	(2)	(3)
Estimation Method	OLS	OLS	LAD
*I*gProduction*	2.858*** (.773)	2.759*** (.783)	3.459** (1.432)
*I*Inflation*	1.353 (1.667)	-	-1.821
Reelect	25.828*** (7.316)	28.488*** (6.588)	34.331*** (11.251)
Terms	-7.828*** (2.294)	-6.842*** (1.904)	-5.134 (3.587)
Constant	52.323 (5.854)	50.857 (4.977)	44.200 (7.805)
n	20	20	20
Standard Error	18.638	18.427	-
R^2	.711	.699	.515
F-Statistic	22.22	27.56	-
Percent Predicted Correctly	80%	85%	85%

•Robust standard errors in parentheses

•Percent predicted correctly based on if the model predicted the correct winner

•* denotes significance at the 10% level, ** denotes significance at the 5%level, and *** denotes significance at the 1% level

• The R^2 in (3) is the Psuedo R^2 from the LAD regression

Table 3:
Estimation of P(*Demwins)*

	(1)	(2)	(3)	(4)
Estimation Method	OLS	WLS	Probit	Logit
*I*gProduction*	.0228 (.0135)	.0044 (.0047)	.0531 (.0447)	.0527 (.0498)
*I*Inflation*	-.0352 (.0279)	-.0368 (.0219)	-.0288 (.0589)	-.0292 (0602)
Reelect	.5975*** (.1116)	.6075*** (.0635)	.8929*** (.3632)	.9131** (.4221)
Terms	-.1181** (.0500)	-.0883*** (.0198)	-.2196* (.1387)	-.2317 (.1505)
Constant	.5294 (.1058)	.5233 (.0198)	.1277 (.4542)	.2993 (.8462)
n	20	20	20	20
Standard Error	.3363	.0620	-	-
R^2	.6459	.9807	.6128	.6074
F-Statistic/Chi-Square Statistic	17.55	242.50	16.50	16.35
Log Liklihood	-	-	-5.212	-5.285
Percent Predicted Correctly	90%	85%	90%	90%

•Robust standard errors in parentheses in (1)

•Percent predicted correctly based on if the model predicted the correct winner

•* denotes significance at the 10% level, ** denotes significance at the 5%level, and *** denotes significance at the 1% level

• The R^2 in (3) and (4) is the Psuedo R^2

•Coefficients reported in (3) and (4) are the marginal effects computed at the mean, except for the constant term

Table 4:

Predicted Winners

		Vote Share Equations			Probability Equations			
Year	**Winner**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(4)**
1932	1	1	1	1	1	0	1	1
1936	1	1	1	1	1	1	1	1
1940	1	1	1	1	1	1	1	1
1944	1	1	1	1	1	1	1	1
1948	1	1	1	1	1	1	1	1
1952	0	0	0	0	0	0	0	0
1956	0	0	0	0	0	0	0	0
1960	1	1	1	1	1	1	1	1
1964	1	1	1	1	1	1	1	1
1968	0	1	1	0	0	0	0	0
1972	0	0	0	0	0	0	0	0
1976	1	0	1	1	1	1	1	1
1980	1	1	1	0	1	1	1	1
1984	0	0	0	0	0	0	0	0
1988	0	1	1	1	1	1	1	1
1992	1	0	0	0	0	0	0	0
1996	1	1	1	1	1	1	1	1
2000	0	0	0	0	0	0	0	0
2004	0	0	0	0	0	0	0	0
2008	1	1	1	1	1	1	1	1
Number Predicted Correctly	-	16	17	17	18	17	18	18
Percent Predicted Correctly	-	80%	85%	85%	90%	85%	90%	90%

•1 represents a Democrat, 0 represents a Republican

2012 Election Predictions

There is still almost a year until the 2012 election. Currently, it is not even clear who the Republican nominee for the presidency will be, however, I can use the most up to date data available to make a prediction of which party will win. Using data from Fair's website[1] my prediction is the same for all models. I predict a victory for President Obama. However, there is much that can happen between now and election day.

Works Cited

Fair, Ray C. "The Effect of Economic Events on Votes for President." The Review of Economics and Statistics 60.2 (1978): pp. 159-173. Print.

Fair, Ray C. "Presidential and Congressional Vote-Share Equations." American Journal of Political Science 53.1 (2009): pp. 55-72. Print.

Kahane, Leo H. "It's the Economy, and then some: Modeling the Presidential Vote with State Panel Data." Public Choice 139.3/4 (2009): pp. 343-356. Print.

Kinder, Donald R. and Kiewiet, D. Roderick "Sociotropic Politics: The American Case." British Journal of Political Science 11.2 (1981): pp. 129-161. Print.

[1] http://fairmodel.econ.yale.edu/vote2012/index2.htm