Economic Growth and U.S. Presidential Election Outcomes

Nicholas Kortessis

Contents

Introduction

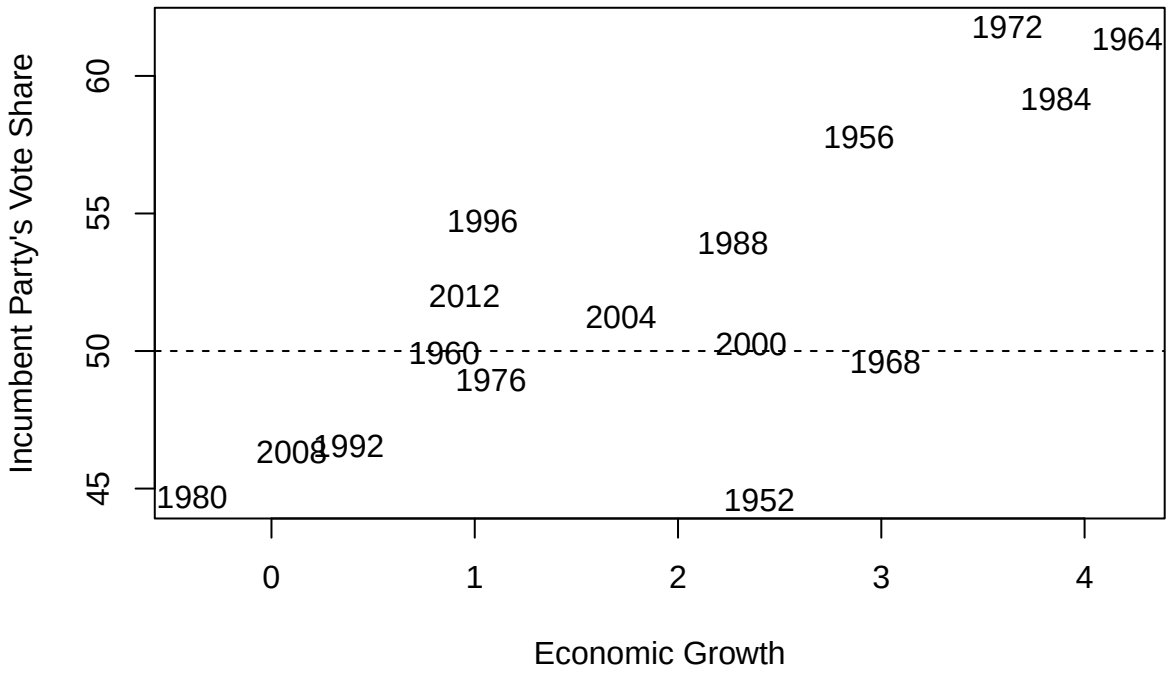
This analysis explores the relationship between economic growth and incumbent party performance in U.S. presidential elections. We use the Hibbs dataset which contains information about economic growth rates and voting percentages for presidential elections.

Data Exploration and Visualization

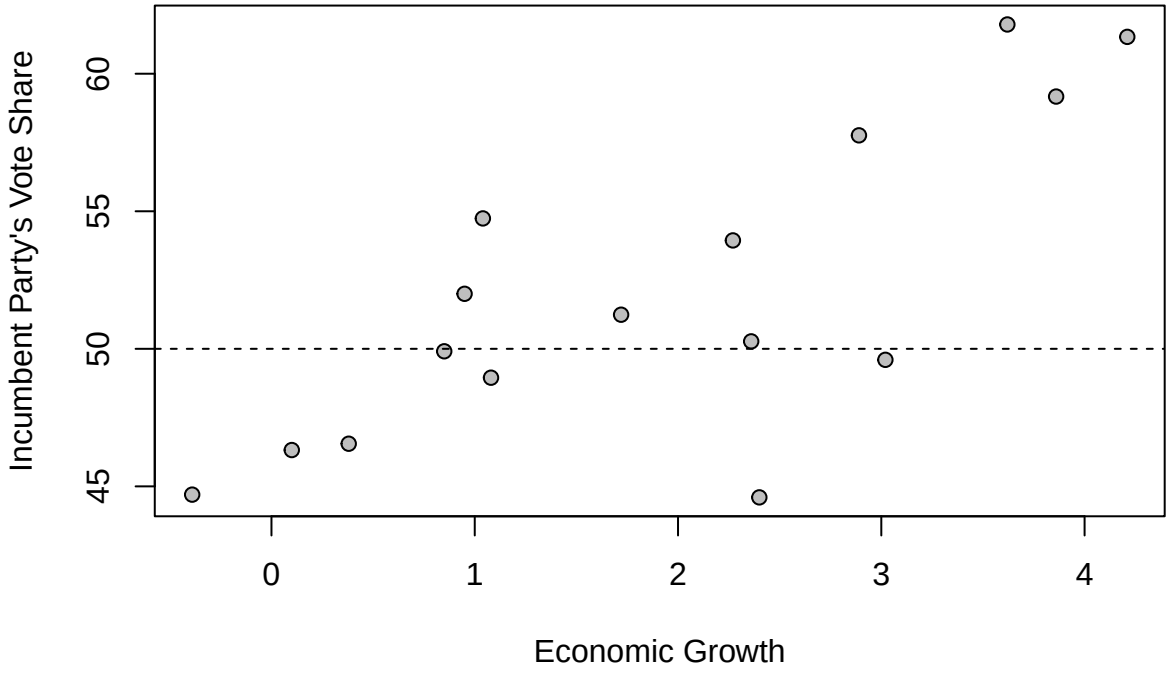
The dataset contains information about U.S. presidential elections, including economic growth rates and incumbent party vote shares.

Let's start by visualizing the relationship between economic growth and incumbent party vote share. On the x-axis is the a measure of economic growth over the 4-year presidential term in terms of percentage increase in household income. On the y-axis is the voting percentage for the incumbent party's presidential nominee. I've labeled points by the year of the election.

As an example, look at point 2012, the most recent point in the dataset. The x-axis is the percentage change in average household income during the years 2009-2012 when Barack Obama was in his first term in office. The election in 2012 was between Barack Obama (D) and Mitt Romney (R). The y-axis shows Barack Obama's vote share that election because he was the democratic party's nominee for president and the incumbent (Barack Obama) was a democrat.



The year's are informative, but make it a bit difficult to see. Here is the same figure with points. In both panels, the horizontal dashed line represents 50% vote share. Above this line, the incumbent was the popular vote and below this line the opposition party's nominee won the popular vote. (US elections don't exactly follow popular vote shares; notable examples where the popular vote winner lost the election were Al Gore (D) in 2000 and Hilary Clinton (D) in 2018. We'll ignore this complication for now and use popular vote share to predict US presidential elections.)



Linear Regression Analysis

We can try to predict vote share using economic growth as a predictor using linear regression. Doing so gives the following model fit.

```
##
## Call:
## lm(formula = vote ~ growth, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9929 -0.6674  0.2556  2.3225  5.3094
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46.2476     1.6219  28.514 8.41e-14 ***
## growth        3.0605     0.6963   4.396  0.00061 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.763 on 14 degrees of freedom
## Multiple R-squared:  0.5798, Adjusted R-squared:  0.5498
## F-statistic: 19.32 on 1 and 14 DF,  p-value: 0.00061
```

Interpreting the Results

The regression analysis shows the relationship between economic growth and the incumbent party's vote share. The coefficient for growth indicates how much the vote share changes with each percentage point increase in economic growth. And the intercept term shows what the vote share is when economic growth is 0%.

With an estimated intercept of 46.25%, we can see that voters penalize the incumbent party's candidate when their household income doesn't change. We might say voter's expect to make more money every four years, and if they don't, they vote for change.

With an estimated slope of 3.06, we say that the average voting share increases about 3 points for every 1% increase in household income. As a presidential adviser with statistics expertise, you might then be able to advise a president to prioritize the economy to ensure future political success and give this number as the evidence for their prioritization of economic issues.

Of course, both of these have uncertainty associated with them as indicated by the standard errors of the coefficients. We can use these standard errors to generate confidence intervals as in the table below.

```
##                 2.5 %    97.5 %
## (Intercept) 42.768951 49.726345
## growth       1.567169  4.553887
```

From this, we are relatively certain in our claim that voters punish incumbent parties that don't increase household wealth because the 95% confidence interval for the intercept doesn't overlap 50%. Moreover, while the effect of increasing the economy could be stronger or weaker, we are quite confident that the economy matters for presidential voting because the 95% confidence interval for the slope is quite high above 0.

But the economy isn't everything. The president might have other political issues they think are worth focusing on. To get at precisely how much the economy explains about presidential voting, we can use R^2, which tells us about the proportion of the variation in presidential voting that can be explained by differences in household income. Here, we find that the R^2 values is about 55%. This tells us that focusing on the economy is important, but 45% of the variation in the voting is about other issues than household income. For example, social issues are often important, geopolitical issues sometimes matter, and of course candidates matter. A more complicated model that includes these factors might be more predictive of US presidential voting.

Prediction Intervals

The best fit line tells us about the average voting preferences under some amount of economic growth. To make predictions about specific elections, we need to use another part of the model. Remember that the model for linear regression is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where ϵ_i is the residual and assumed to be normally distributed with mean 0 and some standard deviation, σ^2. Another way to write this that is exactly equivalent is

$$Y_i \sim \text{Normal}(\mu = \beta_0 + \beta_1 X_i, \sigma^2)$$

meaning that the voting share is normally distributed at some economic growth X_i with mean given by the point at that line and some variance. The standard deviation is estimated by the model and is identified in the model summary above as "residual standard error" with an estimated value of 3.763.

Thus, we have the following fit for our model

$$Y_i \sim \text{Normal}(\mu = 46.25 + 3.06 X_i, \sigma = 3.76).$$

This is a normal distribution just like any other. We can use this to make predictions. For example, if the economic growth is 2%, then the distribution of vote shares is

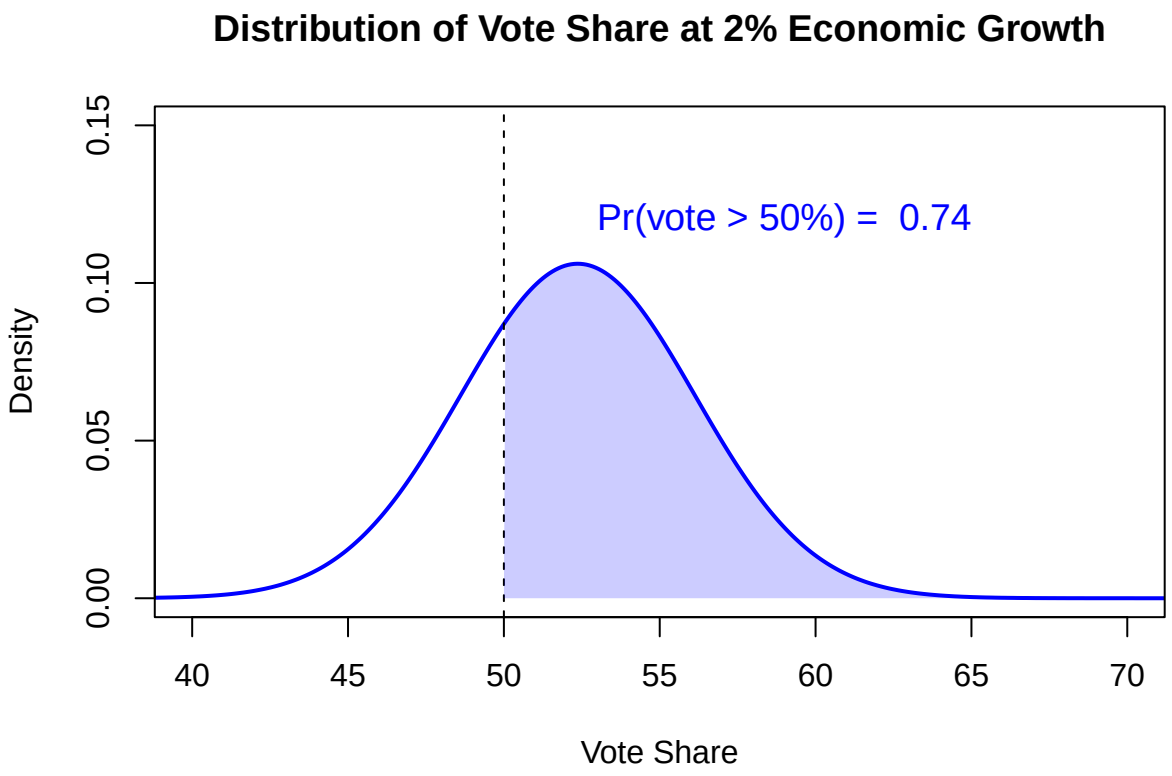
$$Y(X = 2) \sim \text{Normal}(\mu = 46.25 + 3.06 \times 2, \sigma = 3.76)$$

or

$$Y(X = 2) \sim \text{Normal}(\mu = 52.37, \sigma = 3.76)$$

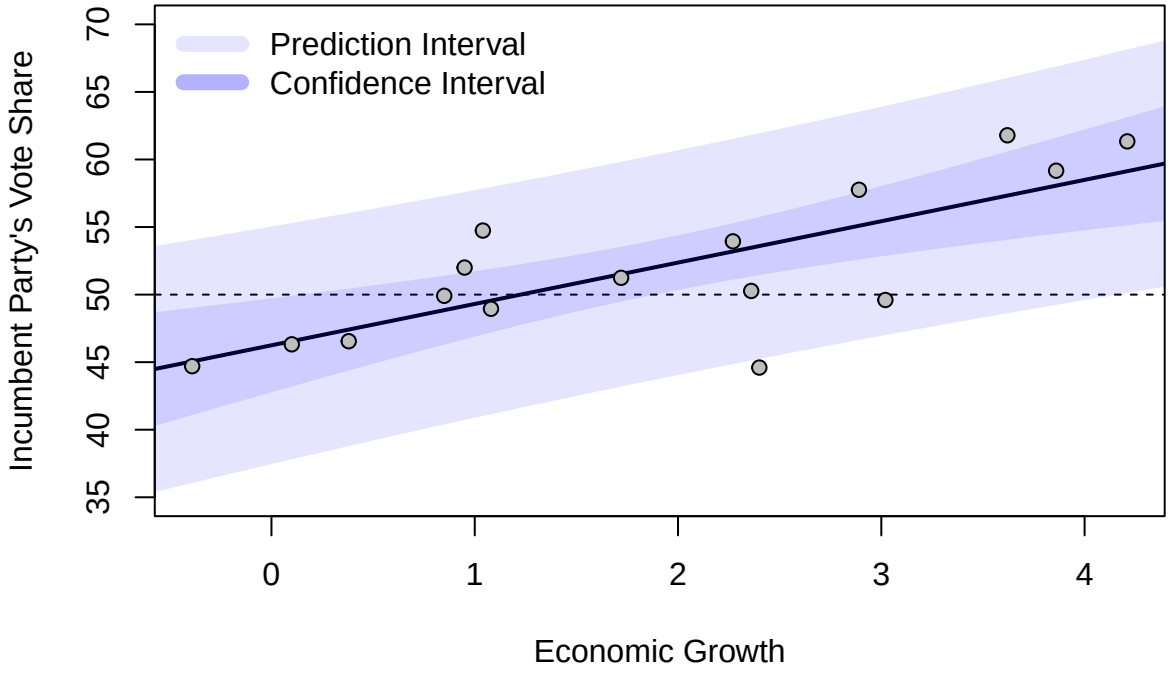
Clearly the average of this distribution is above 50%, but what is the probability that it is above 50%?

Here is a figure of the distribution identifying the fraction of the distribution above 50%.



Visualizing the Regression Model

We can visualize the regression model, the confidence interval of the line, and the prediction interval for the entire dataset. Below is a figure. The 95% confidence interval gives the region where 95% of the best fit lines fall. The 95% prediction interval gives the region where 95% of the elections (i.e., individual data point) fall. The prediction interval is wider than the confidence interval because it includes the uncertainty in the mean response and the variability of individual responses around that mean.



Conclusion

This analysis demonstrates a clear relationship between economic growth and incumbent party performance in U.S. presidential elections. The linear model suggests that:

1. Higher economic growth tends to benefit the incumbent party
2. The 50% vote threshold (needed for victory) typically requires positive economic growth
3. The relationship is statistically significant, but there is still considerable variation not explained by economic factors alone.

The confidence intervals show the uncertainty in the estimated mean response, while the prediction intervals show the wider range of potential outcomes for individual elections.