

POL 51: Scientific Study of Political Science

Professor Brad Jones

Fall Quarter 2008

Assignment 2: Graphical Displays of Data w/Interpretation

The objective of this assignment is to get you thinking theoretically about observed political data. This assignment requires you to do several things and so do not delay in starting it! It is due Monday, October 20.

On the class website, there are two versions of the same data set called `congress_price_data`. The first version of the data is in a format that is compatible with the statistical software Stata; the second is compatible with Microsoft's Excel. Note that you do not need to use Stata or Excel to complete this assignment. Further, you do not even need to have these programs on your computer to complete the assignment. I ask you to address several questions and perform several tasks. Answers should be word-processed and graphs should be neatly incorporated into the document. I want you to use R in order to produce some basic plots and simple statistics for this assignment. In order to successfully do this assignment, take the following initial steps:

1. Download the R package (it is compatible with both Mac and Windows OS (as well as Unix)). Instructions to do this are on the class website.
2. Create a folder on your computer in which you will save the data set as well as the R code used for this assignment (as well as other assignments for this class). For example, you might create a directory called:

C:\pol51\R

Save your materials (data/script/etc.) here.

3. Write up your R script to read in the data. The data I want you to use are data from the Iowa Electronic Market. In this case, the winner-take-all market for control of the U.S. Congress (House and Senate). This market gives four prices. The prices are (along with the name of the variable in the data set):

Democratic House/Democratic Senate `dhdprice`
Republican House/Republican Senate `rhrsprice`
Democratic House/Republican Senate `dhrsprice`
Republican House/Democratic Senate `rhrsprice`

These data are *time-serial* in that we have daily pricing for each market. The time series begins August 24, 2008 and goes through October 10, 2008. The meaning of the price data can be summarized as follows:

"In the IEM winner-take-all market, one share of a candidate pays off one dollar if the candidate wins

and nothing if the candidate loses. A portfolio of one unit of each candidate pays exactly one dollar. A trader who buys one unit of a candidate at, say 40 cents on the dollar, wins either one dollar (a 60 cent profit) or nothing (a 40 cent loss) if the contract is held until market closing following the election. If our trader buys at 40 cents and sells at, say 60, the profit is 20 cents.” (Erikson and Wlezian, n.d.).

Apart from the pricing data, the other variable in the data set of interest is called `date`, which gives the calendar date on which the pricing data are recorded. This variable is read into R as a character variable (non-numeric). I will give you the code to convert the date variable into a factor (that is, a variable that can be used for creating plots).

4. After reading the data in, please answer/do the following questions/tasks. Please note that currently, both houses of Congress are controlled by Democrats (I refer to them as the legislative incumbent party) and the White House is controlled by the Republican party (I refer to them as the executive incumbent party).

a. What is the actual distribution of seats in the current House and Senate (between Republicans, Democrats, and Independents)? You won't find this answer in the data! What would it take in terms of seat change for the Republicans to regain the House and Senate?

b. Suppose the following hypothesis were posed:

i. When the economy is faring poorly, the *legislative* incumbent party is punished at the polls (implying the Democrats lose seats) and the Republican party is rewarded (with seat gains).

ii. When the economy is faring poorly, the *executive* incumbent party is punished at the polls (in this case, implying the Republicans lose seats) and Democratic party is rewarded (with seat gains). iii. When the economy is faring poorly, Americans tend to prefer divided government (i.e. a situation where one house of Congress is controlled by one party and the other is controlled by the opposite party).

Given the data that you have here, discuss what you would expect to observe in these pricing data under hypotheses i., ii., and iii. That is, what kind of trends, averages, etc. would it take for you to accept or reject any of these hypotheses and why? Be creative knowing that you don't actually have to do all the things you discuss here! Also, are these hypotheses directional in nature? Or bi-directional? How would you characterize them?

5. Now, provide some analysis of these data. Specifically, answer/do the following:

a. What is the mean (average), minimum, and maximum for each market? Put this in tabular form.

b. Based on these numbers, which of i, ii, or iii is most supported and why?

c. Create a time-series plot for each of the markets and combine the plots to fit on a single page (I'll give you code fragment to do this).

d. Describe/analyze the following:

i. What are the general trends, if any, in the pricing data for these markets?

ii. Is there evidence of a post-convention “bounce” or “drop” in any of the markets (note, you will have to find out when the conventions were and then “eyeball” the plot)? Specifically, can we detect a Palin

effect in these markets? If there are any bounces/drops, how long-lived are they?

iii. Overall, if you were to forecast the election outcome based on your plot and analysis of these trends, what would you predict about party control of Congress in the Nov. 4th elections?