Analyzing Mathematical Models by Measuring Congressional Approval using 2006 Pre-Election Survey Data

R. Lee McNish and Scott J. Basinger

Department of Political Science, University of Houston, Houston, TX 77204

The purpose of this project is to analyze different methods of measuring causal effects and investigate the various pros and cons of each mathematical model used. By using survey data from a 2006 pre-congressional election study, we intend to look at the different possible models, build our own models to measure a public grading of Congress (on an A to F scale), and compare the results. This will help to determine the best course of action when building models, especially concerning the problem of whether it is better to run a simple regression model or take the risk of using bad controls.

Selection Bias

• Determining the causal effect of treatment requires the following question: "What would happen if individual *i* had chosen not to be treated and was (or vice versa)?"

• For public views on whom politicians are interested in serving and its effect on congressional grading we have:

$$E[Y_i|D_i=1] - E[Y_i|D_i=0]=$$
$$E[Y_{1i}-Y_{0i}|D_i=1] +$$
$$E[Y_{0i}|D_i=1] - E[Y_{0i}|D_i=0]$$

Observed Treatment Effect = Average Treatment Effect on Treated + Selection Bias

A major problem with this model is that $E[Y_{0i}|D_i=1]$ is hypothetical and cannot be measured.

• The treatment effect in this model will be greater or smaller than the actual treatment effect depending on whether selection bias is positive or negative.

Short Regression

• Regression is a restatement of the previous model.

• For the effect of public views on politician interest on congressional grading:

$$Y_i=\alpha+\rho D+\eta_i$$

Where Y_i is the observed treatment effect, α is a base grading coefficient, ρ is the average treatment effect, D is the effect of views on politician interest, and η_i is the error term.

• The regression formula provides clearer results, and while it does not eliminate selection bias, it can now be manipulated through the inclusion of controls in order to reduce variance and better determine a causal effect.

Long Regression

• Adding control variables can reduce variance, provided the controls affect and dependent variable and are not affected by the independent variable. For a good control, we use an approval rating of Congress in handling its responsibilities.

• The long regression formula is:

$$Y_i=\alpha+\rho D+A\gamma+\eta_i$$

Where A is an "approval" term, γ is the observed effect of approval.

Bad Control

• There are two possible ways to implement poor control variables: using a variable that is too weak, or a variable that is affected by the independent variable, the latter of which is known as over-control.

• Our bad control is public grading of how ethical Congress is:

$$Y_i=\alpha+\rho D+\gamma C+\eta_i$$

Where C is a variable detailing views on congressional ethics.

Construction of Variables

1. Grade variable: Please grade Congress on its ability to

Lee2- Act in a bipartisan manner

Lee3- Conduct business carefully and deliberately

Lee5- Tackle important issues

Lee8- Carry out productive discussion

Lee9- Make its workings and activities open to the public

2. Interest variable: Please note whom you believe your Congressman(cp16)/Congress as a whole(cp17) is most interested in

A- Themselves

B- Their district

C- The country as a whole

3. Approval variable: Please provide a measure of your current approval of Congress (strongly disapprove-strongly approve)

4. Ethical grading variable: Please grade Congress on its ability to

Lee1- Oversee the activities of the Executive Branch

Lee4- Represent America's diverse interests

Lee7- Hold its members to high ethical standards

Lee12- Control the influence of special interest groups

Data

A. Short Regression (Grades by interest) **Adj. R^2**

$$Y_i = 0.9928 + 0.2166D + \eta_i$$
$$(0.0189)$$ **0.1105**

B. Long Regression (Grades by interest and approval)

$$Y_i = 0.6171 + 0.1303D + 0.482A + \eta_i$$
$$(0.0176) \quad (0.0282)$$ **0.3072**

C. Over-Control (Grades by interest and ethical grading)

$$Y_i = 0.3912 + 0.0505D + 0.7434C + \eta_i$$
$$(0.0131) \quad (0.0201)$$ **0.6163**

Data are from a 2006 public opinion survey held before the mid-term election by the Center on Congress at Indiana University. The sample size is 1051 for model A, 1032 for model B, and 1034 for model C.

Standard errors are in parentheses.

Results

The data show that there is a correlation between what people believe politicians are interested in and how they grade Congress (A). Controlling for approval of Congress in handling its responsibilities results in a decrease in the standard error for the interest variable, showing it to be a good control (B). Controlling for ethical grading produces a nearly negligible interest variable, and is itself a high coefficient, raising alarms that this may be a bad control (C).

Conclusion

Extreme care must be taken when picking control variables. A good control can reduce variance, and therefore produce a more accurate causal coefficient. However, a bad control can generate less reliable results than not using any controls at all.

Special thanks to the University of Houston Office of Undergraduate Research for funding and helping us to do this project.