

Assignment 5: Multiple Regression Topics and Functional Forms

Due Friday, November 8, 2024

Insert Name Here

Instructions

Answer the questions within this document. Keep the questions included below. Make sure to answer the questions in sufficient detail.

- Report all regression results in table form using “**texreg**”.
- Submit your rendered .pdf in Blackboard.

Number 1: Standardized Coefficients and Model Issues

Let’s say you’re interested in the correlates of candidate evaluation in the American public. Using “nes.dta” (National Election Studies), you examine a dependent variable that is a 101-point “feeling thermometer” of evaluations toward Donald Trump in 2016 (variable name is **ft_Trump_pre**) when he ran against Hillary Clinton. Recall that President Obama was the incumbent at this time. The variable ranges from 0 to 100, where 0=cold, negative feelings, 50=neutral, and 100=warm, positive feelings.

You’re interested in the effects of three core variables:

- **partyid7**: Party identification; 7-point scale ranging from 1=Strong Democrat to 7=Strong Republican
- **libcon7**: Ideology; 7-point scale ranging from 1=extremely liberal to 7=extremely conservative.
- **better_worse_past_econ**: Economic perceptions: 5-point scale measuring whether respondent believes economy is better or worse since last year; 1=much better, 2=some-what better, 3=about the same, 4=somewhat worse, and 5=much worse.

You will also include the following controls:

- **Female**: 1=female, 0=male
 - **educ4**: 1=HS or less, 2=some college or Assoc. degree, 3=B.A., 4=Grad or more
 - **Race3**: 1=white, 2=black, 3=Hispanic; note that you do not want to include this variable as is. This is nominal variable, so you need to “dummy out” the categories and include two out of the three.
 - **income5**: Income; 1=<=\$22.5k, 2=\$22.5-\$45k, 3=\$45-\$75k, 4=\$75-\$125k, 5=>\$125k
 - **Age**: Age in years, ranging from 18-97
- A. Estimate a linear regression of Trump evaluation (using the feeling thermometer). Include in your model the three core independent variables as well as the control variables. Report your results. Interpret the results, and focus on the core three variables of interest. You can quickly cover the controls.
- B. A discussant on your paper encourages you to draw more specific comparisons between the effects of your core independent variables of interest. She suggests that you estimate

standardized coefficients for the effects of all independent variables. Report your results. What do your results suggest about the relative impacts? What are the pros and cons of using standardized coefficients in this context? Under what conditions does it make sense to use standardized coefficients?

- C. In addition, briefly discuss whether you think this model is in serious violation of any regression assumptions.
- D. Do you suspect the model may be suffering from multicollinearity, particularly among the three core independent variables? Explain. Are any of these concerns more serious in the sense of violating more core regression assumptions as discussed in Part C?

Number 2: Nonlinear Functional Form I

Consider the following hypothesis: *In comparing countries around the world, as GDP increases, levels of democracy increase as well.* Use the “world.dta” dataset to answer the following questions.

Dependent variable:

- **fhrate08_rev**; Democracy score as measured by Freedom House

Independent variables:

- **ungdpcap10k**: GDP per capita in \$10,000 increments (a 1-unit increase is a \$10,000 increase)
 - **govregrel**: Government regulation of religion, 10 point scale, higher values more regulation
 - **educ_quality**: Avg. rating of educational system, ranges from about 2 to 6
 - **frac_eth**: Ethnic fractionalization; 0=low, .5=medium, 1=high
- A. Generate a scatterplot between democracy and GDP. Include both a smoothed line and a linear line of best fit. What type of relationship does this look like?
- B. Estimate a linear regression model testing this hypothesis. GDP is your variable of interest. Also include the three control variables. *Model the linear functional form for the effect of GDP.* Report your output and interpret the results, focusing specifically on sign and statistical significance, as well as goodness of fit.
- C. Generate a partial residual plot for GDP. Combining this result with your graph from part A, what is the appropriate functional form for the effect of GDP?
- D. Estimate a new regression model with the appropriate transformation for GDP. Include the three control variables as well. Include your results. Interpret your results.
- E. Generate a post-estimation plot for the nonlinear functional form for the effect of GDP.

Interpret the results and make a substantive conclusion.

Number 3: Nonlinear Functional Form II

An expert on college readiness suggests that one's high school size has a negative effect on SAT scores. This expert argues that as school size increases, student distractions increase, class sizes increase, and there are less opportunities for small group instruction and test preparation. These factors all have negative effects on a student's ability to excel on standardized tests.

Use the "sat.dta" dataset, which is a survey of 4,137 students. Here are the relevant variables:

- **sat**: Student's SAT score
 - **hsize**: Size of student's graduating high school class in hundreds
 - **female**: 1=female, 0=male
 - **white**: 1=white, 0=non-white
 - **athlete**: 1=athlete, 0=not an athlete
- A. Generate a scatterplot between SAT score and graduating class size. Include a smoothed line and a linear line of best fit. What does the graph suggest about the expert's expectation?
- B. Estimate the following model and report your results. Interpret your results and explain whether the results support the expert's expectation.

$$sat_i = \beta_0 + \beta_1 hsize_i + \beta_2 female_i + \beta_3 white_i + \beta_4 athlete_i + u_i$$

- C. Generate a partial residual plot for **hsize**. What type of functional form does it suggest?
- D. Estimate a new model with the appropriate nonlinear functional form for the effect of **hsize**. Report your results and interpret what you find for **hsize**.
- E. Report a post-estimation graph (use margins) showing the nonlinear functional form for **hsize**.