Assignment 7: Diagnostics

Due Monday, December 9, 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: Heteroscedasticity

A. Using the "heterosc" data, estimate the following regression model and report your results:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

Briefly discuss the results in terms of direction and statistical significance.

B. Save your residuals, and plot those residuals against $\widehat{y}_i$ and each of the x variables. Include the four graphs. Briefly interpret and discuss those graphs. Do any of the graphs suggest evidence of heteroscedasticity? Do any of the x variables appear to be inducers of heteroscedasticity? Explain your reasoning for your interpretations.

C. Run a "global" Breusch-Pagan-Godfrey (BPG) test for heteroscedasticity. Include your results. What do you conclude?

D. Run covariate-specific BPG tests for each of the three independent variables. Include your results. What do you conclude?

E. Re-estimate the model using robust standard errors. Include your results. Discuss whether and to what extent your results change. In particular, discuss how the standard errors change and whether the statistical significance of each variable changes.

Number 2: Multicollinearity

In this problem, you'll use the "collinearity" data. Assume that there has been lots of theorizing about y_i. You have brought new data to bear on this theory and have specified the following model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + u_i$$

Past work has posited that x1 and x3 have negative effects on y, while the remaining five variables have hypothesized positive effects on y. x6 and x7 are particularly controversial, according to some scholars, and the prevailing wisdom is that these variables have a positive impact on y.

A. Estimate the regression model. Include your results. How do the results accord with the theoretical expectations discussed above?

B. Produce a correlation matrix of the x variables. Include your results. Discuss some of the high correlations. Do any of these correlations shed light on any unexpected results from part A?

C. Generate variance inflation factors (VIFs) and sqrt(VIF) for each independent variable. Include your results. Interpret the results. What numerical guideline for sqrt(VIF) suggests severe multicollinearity? Which effects suffer from severe multicollinearity?

D. Having made a diagnosis, discuss some steps (at least three steps) you would take to manage the problem of multicollinearity. In answering the question, imagine how you would think about these steps if this were your dissertation research or an article you are planning on submitting to a journal.