POLS 602 Homework 4

Professor Paul Kellstedt
Nicholas Ray, Supplemental Instructor (SI)

2023-11-07

This homework assignment is due November 14th via email to the SI. The assignment is worth 100 points total, with each problem and its subsequent parts labeled with their respective worth. You are allowed to collaborate on this assignment and use online resources, so long as you turn in your own work and clearly attribute others' ideas. In your email, please include the .pdf file with your answers and any code used to produce your answers and the .pdf.

Problem 1: Homoskedasticity (25 points)

1.1 (10 points)

What is the assumption of homoskedasticity and what are the consequences of violating it?

1.2 (10 points)

Read in `data1.csv` and estimate the model $y_i = x_i + z_i + \varepsilon_i$. Evaluate the assumption of homoskedasticity for this model using both a figure [`plot(...,which=3)`] and a formal test [`lmtest::bptest(...)`]. Discuss if the residuals appear to be homoskedastic or not and why.

1.3 (5 points)

Briefly, do you have any concerns about the model you ran given your evaluation of homoskedasticity above?

Problem 2: Autocorrelation (50 points)

2.1 (10 points)

What is the assumption of no autocorrelation and what are the consequences of violating it?

2.2 (10 points)

Read in `data2.csv` and estimate the model $y_i = x_i + z_i + \varepsilon_i$. Evaluate the assumption of no autocorrelation for this model using both a figure [`stats::acf(...)`] and a formal test [`lmtest::bgtest(...)`]. Discuss if the residuals appear to be autocorrelated or not and why.

2.3 (5 points)

Briefly, do you have any concerns about the model you ran given your evaluation of autocorrelation above?

2.4 (25 points)

Conduct a simulation to evaluate the effects of serial correlation (autocorrelation in the residuals) on a regression model. More specifically, simulate the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where $\beta_0 = 1$ and $\beta_1 = 2$. Use the function `stats::arima.sim(list(ar=p),N)` to generate ε_i with various levels of autocorrelation such that $\rho \in (0, 0.25, 0.5, 0.75, 0.99)$. For each level of ρ, conduct 100 simulations with 1000 observations for each model. Answer the following two questions using your results:

1. How does $\hat{\beta}_1$ vary across values of ρ? Compare bias across different values of ρ [i.e., $mean(\hat{\beta}_1) - \beta_1$].
2. How does $se(\hat{\beta}_1)$ vary across values of ρ? Compare the standard deviations of your estimates to their standard errors across different values of ρ [i.e., $\frac{sd(\hat{\beta}_1)}{mean(se(\hat{\beta}_1))}$]. If this ratio is bigger (smaller) than 1, the standard errors underestimate (overestimate) the true variability of the estimates.

Problem 3: Functional Form Concerns (25 points)

3.1 (10 points)

Read in `data3.csv`. Estimate the model $y_i = x_i + z_i + \varepsilon_i$. Plot the relationship between y_i and x_i from your model including the slope. Briefly, how well does this model appear to fit the observed relationship?

3.2 (15 points)

Estimate a new model with a transformed x_i that better fits the observed relationship. Make a well-formatted table to display the original model and your updated model together. Do you reach different conclusions in your new model?