Basics for Quantitative Analysis

(2020 Final Exam)

Upload your answer sheets and R program in the I-campus

1. [20 points] Use the data set gpa1.csv.
(a) Estimate the following model.

$$colGPA_i = \beta_0 + \beta_1 hsGPA_i + \beta_2 ACT_i + \beta_3 skipped_i + \beta_4 PC_i + u_i$$

(b) Conduct the Breusch-Pagan test and report the test statistic with its p-value. Is there heteroskedasticity at the 5% significance level?

(c) Report the heteroskedasticity robust standard errors. Compared to the result with the usual OLS standard errors, is there any difference in terms of statistical significance at the 1% level? In paricular, does any varible become insignificant at the 1% level when you use the heteroskedasticity robust standard error?

(d) Conduct the feasible GLS estimation explained on page 18 of the lecture note (LectureNote_Assumtion4.pdf) and report the estimates.

2. [20 points] Use the data set fertil3final.csv. Δgfr_t, Δpe_t, Δpe_{t-1}, and Δpe_{t-2} are cgfr, cpe, cpe1, and cpe2 in the data set, respectively.
(a) Estimate the following model.

$$\Delta gfr_t = \beta_0 + \beta_1 \Delta pe_t + \beta_2 \Delta pe_{t-1} + \beta_3 \Delta pe_{t-2} + u_t$$

(b) Conduct the Breusch-Godfrey test for AR(1) serial correlation and report the test statistic with its p-value. Is there serial correlation at the 5% significance level?

(c) Report the robust standard error by the Newey-West HAC variance-covariance estimator. Compared to the result with the usual OLS standard errors, is there any difference in terms of statistical significance at the 5% level?

(d) Conduct the Cochrane-Orcutt estimation (the iterated Cochrane-Orcutt using the orcutt package) and report the estimates.

3. [20 points] Use the data set wagefinal.csv.
(a) Estimate the following model and report the estimates of β_1 and β_4.

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + \beta_2 \text{exper}_i + \beta_3 tenure_i + \beta_4 black_i + u_i$$

(b) Suppose that $sibs_i$ is uncorrelated with the error term. Show whether it satisfies the condition for IV (by using regression).

(c) Using $sibs_i$ as an IV, conduct the two-stage least squares estimation. Report the estimates of β_1 and β_4.

(d) Suppose that $sibs_i$, $meduc_i$ and $feduc_i$ satisfy the conditions for IV. Using these three variables as IVs, conduct the two-stage least squares estimation. Report the estimates of β_1 and β_4.

4. [20 points] We estimate a linear model. We assume that there is no perfect collinearity and the zero conditional mean assumption for the error term holds. You conduct the Breusch-Godfrey test for serial correlation in the error term and the p-value of the test is 0.11. You also conduct the White test for heteroskedasticity in the error term and the p-value of the test is 0.96. We assume that you can easily obtain the OLS and GLS estimators for this question.

1) Which estimator do you prefer between OLS and GLS in this case? Briefly explain your answer.

Now suppose that the p-value of the Breusch-Godfrey test for serial correlation is 0.03 while the rest are the same as given above.

2) Which estimator do you prefer between OLS and GLS? Briefly explain your answer.

3) One claims that the OLS estimator is better than the GLS estimator if the robust standard error using the Newey-West HAC estimator is used. Do you agree or not? Briefly explain your answer.

Now suppose that the p-value of the Breusch-Godfrey test for serial correlation is 0.96 and the p-value of the White test for heteroskedasticity is 0.04.

4) Suppose that you choose to use the White heteroskedasticity robust standard error. Is your inference valid? Briefly explain.

5. [20 points] Consider a linear model:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + u_t$$

where

$$u_t = \rho u_{t-1} + e_t \text{ for } 0 < \rho < 1$$

for $t = 1, 2, \cdots, n$. We assume that e_t is white noise (i.e., homoskedastic and serially uncorrelated) and e_t is uncorrelated with u_s for $s < t$. Suppose that x_t is independent of u_t.

(a) What problem will the OLS estimator of β_0, β_1, and β_2 have in this case? Explain your answer.

(b) If you use the robust standard error using the Newey-West HAC estimator, do you still have any problem? Explain your answer.

6. [20 points] Given

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i,$$

we estimate $\beta = (\beta_0, \beta_1, \cdots, \beta_k)'$ using the least squares estimation method.

(a) Write the five classical linear model assumptions learned in the lecture.

(b) If the second classical linear model assumption does not hold, what problem will the OLS estimator have?

(c) If the fifth classical linear model assumption does not hold, what problem will the OLS estimator have in terms of its property or inference? Briefly explain how to overcome the problem.

(d) If the fourth classical linear model assumption does not hold, what problem will the OLS estimator have in terms of its property or inference? Briefly explain how to overcome the problems.

(e) If the third classical linear model assumption does not hold, what problem will the OLS estimator have in terms of its property or inference? Briefly explain how to overcome the problem.