Jeff Grogger
Harris School

Winter 2022
University of Chicago

PP 420: Problem Set 2
(due Tuesday, February 8, by 11:59 pm, via Canvas)

A. Analytical problems

1. Suppose you have

$$y = X\beta + \varepsilon$$
$$= X_1\beta_1 + X_2\beta_2 + \varepsilon, \tag{1}$$

and wish to test the hypothesis $H_0: \beta_2 = 0$. To do this, you estimate (1) under the null and retrieve the residuals e_1. You then estimate an auxiliary regression of e_1 on X. Denote the R^2 from this regression by R^2_{aux}. Unless otherwise indicated, you can take assumptions (a)-(d) from Theorem 3 (Topic 3) as given.

(a) Show that $nR^2_{aux} \xrightarrow{d} \chi^2_{n-K_2}$.
(b) Describe precisely how you would carry out a test of H_0. Under what conditions would you accept/reject the null?
(c) Provide some intuition for your test in terms of the quantity $X_2'e_1$.
(d) Now suppose you replace assumption (c)(ii) with (c)(ii') $E(\varepsilon\varepsilon'|X) = \Delta$, where Δ is a diagonal matrix with positive elements on the diagonal. Is nR^2_{aux} still asymptotically $\chi^2_{n-K_2}$?
(e) Suppose you accidentally regressed e_1 on X_2 alone. What can you say about the actual size of this test compared to the size of the test based on nR^2_{aux} from the regression of e_1 on X? Why is it important to include X_1 in the auxiliary regression, when you are interested in testing hypotheses about the coefficients of X_2?

2. Suppose you estimate the model (in deviation form)

$$y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

where the OLS estimates satisfy

$$\begin{bmatrix} b_2 \\ b_3 \end{bmatrix} \xrightarrow{d} N\left(\begin{matrix} \beta_2 \\ \beta_3 \end{matrix}, \begin{bmatrix} \sigma_2^2 & \sigma_{23} \\ & \sigma_3^2 \end{bmatrix}\right).$$

Explain how you would test the hypothesis $H_0: q = 1$ where $q = \beta_2/\beta_3$. Propose your test statistic and justify its distribution under the null. Explain how you deal with nuisance parameters. Explain intuitively your criteria for rejecting the null, then state your precise criterion. What is the probability that you falsely reject the null, given your approach?

B. Computational Problems

1. Using the pp420_data.dta data set, regress earnings on education, age, and age squared. Now carry out the following regressions:

(a) Regress earnings on age and age squared; retain the residuals
(b) Regress education on age and age squared; retain the residuals
(c) Regress the first set of residuals on the second. Compare the slope coefficient (and standard error) to the education coefficient (and standard error) from the original regression.
(d) Regress earnings on the second set of residuals. Compare the results to those you obtained in part (c) and explain.

2. Again, regress earnings on education, age, and age squared.
(a) What is the standard error of the regression? What is the standard error of the education coefficient? The age coefficient?
(b) Print the covariance matrix.
(c) What is the covariance between the education coefficient and the age coefficient?

3. Based on the same regression, assume that the GM assumptions hold, and carry out tests of the following hypotheses:
(a) H_0: $b_{educ}=0$
(b) H_0: $b_{educ}=2b_{age}$
Carry out this test by hand, using the estimated covariance matrix, then use a test command (e.g., testparm in Stata or linearHypothsis in R) to carry it out. Compare the results.
(c) H_0: $b_{age} = b_{agesq} = 0$
First carry this out by computing two regressions and using a calculator. Then use the testparm command (or similar if you are using a program besides Stata). Compare the results.

4. Here you analyze some features of convergence.
(a) First, draw some decidedly non-normal data, in the form of 1000 observations from an exponential distribution with mean 1.5. Plot the distribution of these data by means of a histogram. Refer to these raw data as your population.
(b) Now draw 10,000 random samples of size n=3 from the population. Compute the sample average $\bar{x}$ for each. Plot the distribution of $\bar{x}$ using a histogram. Repeat the exercise for sample sizes n=6 and n=20. What principle from the lecture is being illustrated here?
(c) Repeat the same exercise as in (b), but now plot the distributions of $\bar{z} = \sqrt{n}(\bar{x} - \mu)$. What principle is being illustrated here?