Final Exam Study Questions for ECON 836

1. All study questions for the midterm.
2. Suppose you have estimated a model with a very large number of cases, and that your estimates

   and their variance are given by $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, V\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.25 \end{pmatrix}$.

   a. What is the value of the test statistic for $H_0 : \beta_1 = \beta_2$, and how is it distributed? What if the covariance between the estimated parameters was -0.25?
   b. What is the value of the test statistic for the hypothesis that the first coefficient equals twice the second coefficient? How is it distributed?
3. Assume that religion affects educational attainment, and also affects the level of earnings conditional on education. How could I estimate the total effect of religion on earnings including its effect that runs through education?
4. Say the IV estimator for a coefficient is unbiased, but has a standard error of 1. If I tell you that the OLS estimator has a bias no greater than 0.5 and a standard error of only 0.5, how should you think about this trade-off? Is there an argument to use the biased OLS estimator in this case?
5. If a variable is insignificant in a regression, should you throw it out and re-estimate?
   a. Why or why not?
   b. What if two variables are insignificant in a regression?
6. Imagine that you estimate a model where earnings is a function of a bunch of covariates that are inter-related (they cause each other), and nothing in the model is statistically significant.
   a. Does that mean that these covariates do not influence earnings?
   b. If you believed that they really did influence earnings, what might you do?
7. Derive the following moment estimators:
   a. the method-of-moments estimator for a linear model satisfying the classical conditions is the the OLS estimator.
   b. the method of moments estimator for a linear model with exactly as many instruments as endogenous variables.
   c. The GMM estimator with more than 1 instrument for each endogenous variable. Show that this estimator is the 2SLS estimator when disturbances are homoskedastic.
8. In "EASI Made Easier",
   a. Pendakur uses instrumental variables to correct for endogeneity. Why is there endogeneity? Why are those variables good instruments?
   b. When you implemented your model, it was exactly identified. Were there more instruments that you could have used? Would it have been a good idea to use them?
   c. When you implemented the model, you found that the data violated the Slutsky symmetry restrictions. What should you take from that? What does it mean?
   d. Why is Slutsky symmetry a joint hypothesis involving multiple restrictions? Couldn't one just test each equality restriction separately?
   e. Prices are not observed at the individual level. Slutsky symmetry is about price responses in demand. Does this make you worry about the test of symmetry? How so? How could you address these worries?
   f. The EASI deflator is affine in the Stone index, rather than equal to the stone index. What is the importance of this difference?

9. Consider a test of a multiple linear hypothesis that 4 variables can be excluded from a regression. Suppose also that these 4 variables are independent of each other, and of the other regressors in the model. Prove that the Wald test statistic for the multiple hypothesis is equal to the sum of the Wald test statistics for each hypothesis singly.

10. What is the 95% quantile cutoff for a random whose distribution is the following (a k% quantile cutoff is the value of the random variable such that k% of the density is below that value):

   a. $N(0.13,0.01)$?

   b. $0.1N(0,1) + 0.13$?

   c. $0.10t_{25}+0.13$?

   d. $\chi^2_{10}+0.13$?

11. Consider the confidence region for a pair of parameters given OLS estimates of those parameters. Why is this region not a rectangle with the same limits as the confidence intervals for each of the parameters?

12. Suppose you wanted to know the variance of a LAD (Least Absolute Deviations, i.e., conditional median) estimator, but didn't know the formula for it. Describe a Monte Carlo experiment that would give an estimate of the variance of the estimated LAD coefficients.

13. Let
$$Y = X\beta + \varepsilon,$$
$$X = w + u,$$
$$Z = w\delta + t,$$
$$w \sim N(0,\sigma^2_w I_N), \quad u \sim N(0,\sigma^2_u I_N), \quad \varepsilon \sim N(u\gamma,\sigma^2_\varepsilon I_N), \quad t \sim N(0,\sigma^2_v I_N)$$

where X is a just one column. Let $\hat{\beta}$ denote the OLS estimator, and define residuals e as
$$e = Y - X\hat{\beta}.$$

   a. Derive the bias of $\hat{\beta}_{OLS}$.

   b. Show that Z is a valid instrument (both exogenous and relevant).

   c. Derive the bias of the two-stage least squares estimator.

14. Imagine that you estimate a model where earnings depends on many covariates that are inter-related (they cause each other), and nothing in the model is statistically significant.

   a. Does that mean that these covariates do not influence earnings?

   b. If you believed that they really did influence earnings, what might you do?

15. Consider Norris and Pendakur

   a. Do they need instruments (elements of v2 not included in v1) to identify their model? Suppose that v2 was empty. Would their approach work? Write stata code to test whether or not v2 was only in the probit and not in the regression?

   b. Do you think that v2 have enough variation to make their estimates plausible? What would be the ideal v2 (if you had 'dream data').

   c. Suppose they were interested in measuring inequality rather than poverty. What effect would their approach have on the measurement of inequality?

   d. Write out the value of the likelihood for an observation for a given v1,v2 and parameter vector for their heteroskedastic selection model.

   e. (2nd part of that paper) Slutsky symmetry is an implication of maximization. If symmetry is not true, then consumers are not maximizing.