

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 are $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0.25 \end{pmatrix}\right)$.
 - 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. If the data come in groups, also known as *clusters*, one might expect that there could be correlations within clusters but not across clusters. Here, correlations across observations within clusters are of unknown form, but correlations across observations in different clusters are assumed to be zero. For this question, assume there are 2 clusters.
 - a. What would the GLS weighting matrix look like in this case?
 - b. How could one estimate the variance of the OLS estimated parameter vector? (This is called the “clustered” covariance matrix.)
 - c. What happens if one of the clusters is very big, say, 95% of the observations?
 - d. Why can’t we use this method to model arbitrary cross-observation correlations by saying that there is just one cluster, and using the clustered covariance matrix?
4. 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, and how could I estimate the marginal effect of education on earnings, and what is the difference in what these mean?
5. Why are goodness-of-fit tests based on the residual sum-of-squares and Wald-type tests both distributed as chi-squares?
6. 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.
7. 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. $\chi_{10}^2 + 0.13$?
8. 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?
9. If the discrepancy vector d is just a scalar, can I construct a test statistic equal to $d/se(d)$? What would be the distribution of this test statistic (and under what conditions)?
10. Suppose that I have a nonlinear equation $Y = f(X, \beta) + \varepsilon$, $E[X' \varepsilon] = 0_K$, where there are

K parameters and X is an $N \times K$ matrix. How could I estimate it by MoM? Would this give me the same answers as choosing parameters by minimising the sum of squared errors? Why is it unnecessary to make assumptions about the variance of epsilon?

11. Consider Jacks and Pendakur (2011).
 - a. Are they convincing in their argument that freight prices are endogenous and have only a small effect on trade volumes? Why or why not? Be specific.
 - b. Is their instrumented regression better than their OLS regression? Why or why not? Better for what? (Hint: think about testing vs estimation of a magnitude.)
 - c. They allow for an MA(2) process in the disturbance term for each country. Should they have allowed for more, eg., an MA(40) process?
 - d. Jacks and Pendakur have pretty low precision in their IV estimates. Should they have used the (endogeneity-corrected) random effects estimator instead of the (endogeneity-corrected) fixed effects estimator, so as to get lower variance?
12. Consider Dunbar, Lewbel and Pendakur (2012).
 - a. What makes their estimation “structural estimation”? What is the structure?
 - b. In their model, budget shares are linear in the log of expenditure. Could they estimate the model by linear regression of budget shares on the log of expenditure and z 's, and then use the resulting estimates to construct estimates of the structural parameters?
 - c. Suppose the SAP restriction does not really hold, but we estimate the model given SAP anyways. What do the resulting estimates mean?
 - d. How would you test they hypothesis that resource shares are equal across people? Derive the Wald test for this hypothesis, and describe what additional information (other than in the tables) would be needed to implement the tests. (Hint: remember that a household with 2 children has 4 people, even though only 3 resource shares are estimated.)
13. Suppose you wanted to know the variance of a LAD (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 estimate.
14. Let

$$Y = X\beta + \varepsilon,$$

$$X = w + u,$$

$$Z = w\delta + t,$$

$$w \sim N(0, \sigma_w^2 I_N), \quad u \sim N(0, \sigma_u^2 I_N), \quad \varepsilon \sim N(u\gamma, \sigma_\varepsilon^2 I_N), \quad t \sim N(0, \sigma_v^2 I_N)$$
 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.