ECOM40006/90013 Econometrics 3

Week 5 Extras

Question 1: Reviewing the econometric toolkit

A significant amount of our journey in econometrics will involve many of the same steps being repeated, but for different models. Before we start doing this, it'll be a good idea if we take some time to review some of the techniques that we've got on hand. In what follows, knowledge of *convergence in probability* and *convergence in distribution* is assumed.

(a) Core concepts. Briefly review the concepts of:

(i.) The Continuous Mapping Theorem (CMT)
(ii.) Slutsky's Theorem
(iii.) Khintchine's Weak Law of Large Numbers (WLLN)
(iv.) The Lindeberg-Lévy Central Limit Theorem (CLT)

Briefly discuss some of the other weak laws of large numbers and central limit theorems: what are their main points and how do they differ?

(b) Supporting concepts. Define:

(i.) Big O notation
(ii.) Little o notation
(iii.) Stochastic order notation, O_p and o_p

(c) Block matrices and sums. Consider a $n \times k$ matrix X, which is a matrix of data that holds n observations each of k regressors. The matrix X can also be written

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}'$$

so that x_i is a $k \times 1$ vector that holds the data on the regressors for observation i. (*Note that despite being conventional notation, it can be seriously confusing. That's why this question exists.*) Furthermore, $n \geq k$. A very common sight in econometrics is to switch between the large matrix X and summations of its individual block elements, x_i. One of the most common examples is

$$X'X = \sum_{i=1}^{n} x_i x_i'.$$

Verify this result. Then, consider the $n \times 1$ vector $y = [y_1, y_2, \dots, y_n]'$ and show that

$$X'y = \sum_{i=1}^{n} x_i y_i.$$

Question 2: The OLS estimator

In lectures, Chris demonstrates a number of results in a very rigorous manner. We will explore the rigorous steps in the next set of extras. But for now, let's not bite off more than we can chew and just get ourselves more familiar with the ***basic idea*** *of how these models are going to work. We can always do the rigorous stats later.*

Consider the standard multiple regression model

$$y = X\beta + u$$

where $X = [x_1, \ldots, x_n]'$ is a non-stochastic $n \times k$ matrix with full column rank, β a $k \times 1$ vector of parameters, and y and u are both $n \times 1$ where $u \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2 I_N)$. Further assume that $\mathbb{E}(x_i x_i') = Q$, where Q is a positive definite matrix.

(a) Find an expression for the *sum of squared errors* $u'u$ in terms of y, X and β.

(b) Solve the following minimization problem to obtain the OLS estimator

$$\hat{\beta} = \arg\min_{\beta} u'u$$

in terms of the data y and X.

(c) Substitute $y = X\beta + u$ into the OLS estimator and hence show that

$$\hat{\beta} = \beta + (X'X)^{-1}X'u.$$

Use this result to show that $\hat{\beta}$ is an unbiased estimator of β.

(d) Use the result from (c) to show that $\hat{\beta}$ is also a *consistent* estimator of β. To do this, you will need to appeal to the (weak) Law of Large Numbers. You may find that multiplying things by n/n might come in useful.

(e) Show that the expression from part (c) can be written as

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}X'X\right)^{-1} \frac{1}{\sqrt{n}}X'u.$$

In particular, show also that

$$\frac{1}{\sqrt{n}}X'u \xrightarrow{d} N(0, \sigma^2 Q).$$

(Hint: use the Central Limit Theorem.)

(f) Use the Continuous Mapping Theorem and Slutsky's Theorem to show that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

and via rearrangement, that

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \frac{\sigma^2}{n}Q^{-1}\right).$$

Question 3: The Delta method

Nonlinearity is almost a feature of life when dealing with econometrics. Since linear expressions are easier to work with, the Delta method appeals to the use of a linear approximation to obtain the covariance matrix of a nonlinear function of random variables. One of the immediate uses of the Delta method is for nonlinear hypotheses in particular. For now, suppose that one has an estimator $\hat{\beta}$ that is consistent for β, with asymptotic distribution

$$\sqrt{n}(\hat{\beta} - \beta) \sim N(0, \Sigma),$$

or

$$\hat{\beta} \sim N\left(\beta, \frac{\Sigma}{n}\right).$$

(a) Explain why it is the case that for a continuous function g, $g(\hat{\beta}) \xrightarrow{p} g(\beta)$ and thus the random vector $g(\hat{\beta}) - g(\beta)$ has mean zero. There is a specific theorem that comes in very useful in explaining this.

(b) The information above tells us that $\hat{\beta}$ is a random variable. This also means that $g(\hat{\beta})$ is a random variable. What is the first-order Taylor series approximation of this function around the true value β? Assume for now that β is a scalar. Then, using your answer, find $\text{Var}(g(\hat{\beta}))$. (*Hint*: $\text{Var}(X + a) = \text{Var}(X)$ for a constant.)

(c) Now, repeat part (b), but assume that β is a random vector, so vector notation must be used accordingly.

Question 4: Variance estimators and confidence intervals

One thing that an econometrician is often quite fond of is the classic 95% confidence interval. In this question, we'll derive explicitly what this looks like when we're dealing with the standard OLS estimator. To begin with, consider the linear model from the week 5 extras.

(a) Show the following (we will need these results for later):

(i.) The rank of the residual maker M_X is $n - k$. (*Hint: The rank of an idempotent matrix is equal to its trace.*)

(ii.) The sum of squared residuals $e'e$ can be written $u'M_X u$.

(iii.) The 'normalized' sum of squared residuals $\frac{1}{\sigma^2}e'e$ is distributed as χ^2_{n-k}. (*Hint: check out the chi-squared properties in the week 4 extras, part II. One of them might be useful for this question.*)

(iv.) Hence show, using the property of the χ^2_q distribution with q degrees of freedom $\mathbb{E}(\chi^2_q) = q$, that the following estimator is unbiased for σ^2:

$$\hat{\sigma}^2 = \frac{e'e}{n - k}.$$

(b) Let ℓ_k be a $n \times 1$ vector with a one on the k^{th} element and zero everywhere else. In this case, it can be shown that

$$\hat{\beta}_k \sim N(\ell_k' \beta, \mathrm{Var}(\ell_k' \hat{\beta})) \sim N(\beta_k, \mathrm{Var}(\hat{\beta}_k)),$$

where $\mathrm{Var}(\hat{\beta}_k) = \ell_k' \mathrm{Var}(\hat{\beta}) \ell_k$. We're going to use this to verify the following:

(i.) The covariance $\mathrm{cov}(\hat{\beta}_k, e) = 0$ and hence $\hat{\beta}_k$ and e, by virtue of being jointly normally distributed, are independent. (*Note: The working is the same as a previous question in the week 4 extras. But because this is important, you get to do it again. Yay!*)

(ii.) The t-statistic for $\hat{\beta}_k$ is t-distributed:[1]

$$t = \frac{\hat{\beta}_k - \beta_k}{\widehat{\mathrm{sd}(\hat{\beta}_k)}} \sim t_{n-k},$$

where $\widehat{\mathrm{sd}(\hat{\beta}_k)} = \sqrt{\hat{\sigma}^2 \ell_k' (X'X)^{-1} \ell_k}$ is the estimated standard deviation of $\hat{\beta}_k$ using only the available data on X and y.

(iii.) Let q_x be the probability that observing the t-statistic in (c) is less than x. Construct an interval around 0 in which t will fall in with 95% probability, then rearrange it for $\hat{\beta}$ instead of t. This will be your 95% confidence interval for the OLS estimator. (*Hint: The t-distribution is symmetric around 0. So if t is distributed as t_{n-k}, so will $-t$.*)

Question 5: The Generalized Least Squares estimator

Standard OLS estimation is not the end of the line for econometric analysis. There are many little things that we can do to OLS that makes it unreliable. One such case is endogeneity (which you'll have covered in Econometrics 2). Another case is that of *heteroskedasticity*, which doesn't affect the consistency of OLS estimates, but makes them much less efficient.

(a) Suppose you have two estimators for β, A and B. You know that both estimators are consistent for β, namely that $A \xrightarrow{p} \beta$ and $B \xrightarrow{p} \beta$. However, you know that $\mathrm{Var}(A) < \mathrm{Var}(B)$. Which estimator would you prefer and why?

(b) Now suppose that you have estimators C and D. It turns out that D is unbiased so $\mathbb{E}(D) = \beta$. However, C is biased and so $\mathbb{E}(C) \neq \beta$. You also know that $\mathrm{Var}(C) < \mathrm{Var}(D)$. Discuss the conditions under which you would pick C as an estimator over D.

(c) Consider a linear model $y = X\beta + u$, where $u|X \sim N(0, \Omega)$.[2] You are also aware that

$$\Omega = \mathrm{diag}(f(x_1), f(x_2), \ldots, f(x_N)),$$

so that there exists heteroskedasticity in the model. Since the matrix is diagonal, it is also reasonable to define a new matrix $\Omega^{-1/2}$ such that $\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$ is the inverse covariance matrix.

[1] Note: $\widehat{\mathrm{sd}(\hat{\beta}_k)}$ replaces σ with $\hat{\sigma}$ so you will need to find a way of getting σ back into the t-statistic. From there, you will need to use your past results in this question and the properties of the t-distribution to show the final result.

[2] You can interpret $u|X$ as "the distribution of u, given X.

(i.) The Generalized Least Squares (GLS) method suggests that the data should first be transformed into

$$y^* = X^*\beta + u^*$$

where $y^* = \Omega^{-1/2}y$, $X^* = \Omega^{-1/2}X$, $u^* = \Omega^{-1/2}u$. Show that $\text{Var}(u^*|X) = I_N$.

(ii.) Derive the GLS estimator, obtained by OLS estimation on the transformed data.

(iii.) Derive the conditional variance of the GLS estimator $\text{Var}(\hat{\beta}_{GLS}|X)$.

(iv.) Show consistency of the GLS estimator. Assume that $\frac{1}{N}X^{*\prime}X^* \xrightarrow{p} P$ and is invertible.

Question 6 (bonus): OLS and block matrices

We can write the standard multiple regression model with k parameters as

$$y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \cdots + \beta_k x_{i,k} + u_i.$$

We know that the full model with n observations can be expressed in matrix notation as $y = X\beta + u$, but sometimes we are interested in the behavior of certain parts of the model. For example: what variables are exogenous or not in an instrumental variables regression? To illustrate, let's split the model into two parts: (i) the intercept and (ii) everything else. In other words, we can write

$$y = X\beta + u = \begin{bmatrix} \ell & X_s \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_s \end{bmatrix} = \ell\beta_1 + X_s\beta_s + u$$

where $\ell = (1, 1, \ldots, 1)'$ is a $n \times 1$ vector of ones that corresponds to the intercept term β_1. The term β_s is the matrix block containing the remaining slope coefficients.

(a) Show that

$$X'X = \begin{bmatrix} n & \ell' X_s \\ X_s'\ell & X_s'X_s \end{bmatrix}, \qquad X'y = \begin{bmatrix} \ell' y \\ X_s' y \end{bmatrix}.$$

(b) The OLS estimator $\hat{\beta}$ can be split into blocks just like what we've done above, so $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_s)'$. As we found in question 1, it is the solution to the equation $X'X\hat{\beta} = X'y$. Use your results from part (a) to show that

$$\hat{\beta}_1 = \bar{y} - \bar{X}_s'\hat{\beta}_s \qquad \text{where } \bar{y} = \frac{1}{n}\ell' y \quad \text{and} \quad \bar{X}_s' = \frac{1}{n}\ell' X_s.$$

(c) Let $M_1 = I_N - \ell(\ell'\ell)^{-1}\ell' = I_N - P_1$ be the residual maker associated with a vector of ones. Use your results from parts (a) and (b) to show that

$$\hat{\beta}_s = (X_s' M_1 X_s)^{-1} X_s' M_1 y.$$

To get you started, you should check one of the block matrix expressions from part (b) and substitute $\hat{\beta}_1$ in. Then, you should look for factors to take out so that you can use the expression for M_1 above. From there, rearrangement will give you the answer.

(d) Let's use a different method to get the estimator $\hat{\beta}_s$ above. Show that the linear model $y = X\beta + u$ can be transformed to give

$$M_1 y = M_1 X\beta + M_1 u \implies y^* = X_s^* \beta_s + u^*$$

where $y^* = M_1 y$, $X_s^* = M_1 X_s$ and $u^* = M_1 u$. Then, show that the least squares estimator $\hat{\beta}_s$ from this transformed model is precisely

$$\hat{\beta}_s = (X_s' M_1 X_s)^{-1} X_s' M_1 y.$$