

## Introduction to Econometrics

### Midterm Examination

Fall 2005

Answer Key

*Please answer all of the questions and show your work. Clearly indicate your final answer to each question. If you think a question is ambiguous, clearly state how you interpret it before providing an answer. All question parts have equal weight, and there are 7 in total. Be sure to write your name on your answer books!*

1. Consider the following relationship:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where  $x$  is an exogenous variable and  $i$  is a member of the population of interest. Define the matrix

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix},$$

and the vectors

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

and write

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

for the relationship between  $Y$  and  $X$  in a random sample of size  $N$ . You can assume that

$$E(\varepsilon|X) = 0 \text{ for all } X$$

and

$$E(\varepsilon\varepsilon'|X) = \sigma_\varepsilon^2 I_N,$$

where  $I_N$  is the identity matrix of dimension  $N$ .

You are given the following information from a random sample of size  $N = 100$ .

$$\begin{aligned}X'X &= \begin{bmatrix} 100 & 40 \\ 40 & 200 \end{bmatrix}, \\ \sum_{i=1}^N y_i &= 40 \\ \sum_{i=1}^N y_i^2 &= 150 \\ \sum y_i x_i &= 60.\end{aligned}$$

(Recall that the inverse of a  $2 \times 2$  matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is equal to

$$(ad - bc)^{-1} \times \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Compute each estimate or say why you don't have enough information to do so:

- (a) The Ordinary Least Squares (OLS) estimate of  $\beta$ ,  $\hat{\beta}$ .

*Answer: We already have been given  $X'X$ . We find that the inverse of this matrix is*

$$\begin{aligned}(X'X)^{-1} &= (20000 - 1600)^{-1} \begin{bmatrix} 200 & -40 \\ -40 & 100 \end{bmatrix} \\ &= \begin{bmatrix} .0109 & -.0022 \\ -.0022 & .0054 \end{bmatrix}.\end{aligned}$$

*Now*

$$X'y = \begin{bmatrix} 40 \\ 60 \end{bmatrix}.$$

*Then*

$$\begin{aligned}\hat{\beta}_0 &= .0109 * 40 - .0022 * 60 \\ &= .304 \\ \hat{\beta}_1 &= -.0022 * 40 + .0054 * 60 \\ &= .236.\end{aligned}$$

- (b) The estimated covariance matrix of  $\hat{\beta}$ .

*Answer: To compute this, we need an estimate of  $\sigma^2$ , the variance of the disturbance term. To compute the unbiased estimator  $s^2 = (N - 2)^{-1} \sum_{i=1}^N r_i^2$ , we need to know*

$$\begin{aligned} \sum_{i=1}^N r_i^2 &= \sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\ &= \sum_{i=1}^N [y_i^2 + \hat{\beta}_0^2 + \hat{\beta}_1^2 x_i^2 - 2y_i \hat{\beta}_0 - 2y_i \hat{\beta}_1 x_i + 2\hat{\beta}_0 \hat{\beta}_1 x_i] \\ &= 150 + 100(.304)^2 + (.236)^2 200 - 2(.304)40 - 2(.236)(60) + 2(.304)(.236)40 \\ &= 123.480, \end{aligned}$$

so that  $s^2 = 123.480/(100 - 2) = 1.26$ . Then the estimated covariance matrix of  $\hat{\beta}$  is

$$\begin{aligned} \hat{\Sigma}_{\hat{\beta}} &= 1.26 * (X'X)^{-1} \\ &= \begin{bmatrix} .014 & -.0028 \\ -.0028 & .007 \end{bmatrix}. \end{aligned}$$

2. Consider the linear relationship

$$y_i = \alpha x_i, \tag{1}$$

where  $y_i$  and  $x_i$  are both expressed as deviations from their respective sample means (thus  $\sum_{i=1}^N y_i = \sum_{i=1}^N x_i = 0$ ). In both of the questions that follow we assume that  $x_i$  is measured without error in the sample information available to you.

- (a) First consider the case where  $y_i$  is measured with error, where the measured value of  $y$  is given by

$$y_i^* = y_i + u_i,$$

where  $u_i$  is independently and identically distributed (i.i.d.) with mean 0 and variance  $\sigma_u^2$ . The information available to you is  $y_i^*$ , which is measured as a deviation from the sample, and  $x_i$ ,  $i = 1, \dots, N$ . Find unbiased estimators of  $\alpha$  and  $\sigma_u^2$ . Is your estimator of  $\alpha$  best linear unbiased?

*Answer: After substitution, we have*

$$y_i^* = \alpha x_i + u_i.$$

*Since  $u_i$  is i.i.d., it is mean independent of  $x_i$ . Thus the OLS estimator of  $\alpha$ ,*

$$\hat{\alpha} = \frac{\sum_{i=1}^N y_i^* x_i}{\sum_{i=1}^N x_i^2}$$

is unbiased. Since the disturbance term  $u$  is homoskedastic, the OLS estimator is BLU by the Gauss-Markov theorem. An unbiased estimator for the variance of the measurement error is

$$\hat{\sigma}_u^2 = (N - 1)^{-1} \sum_{i=1}^N (y_i^* - \hat{\alpha}x_i)^2.$$

- (b) Now return to the specification given in (1) - in particular, assume that  $y_i$  is measured without error. Assume that individuals in the population differ in their value of  $\alpha$ . Say that  $\alpha$  is i.i.d. with mean  $\bar{\alpha}$  and variance  $\sigma_\alpha^2$ . From the sample information  $\{y_i, x_i\}_{i=1}^N$ , derive unbiased estimators for  $\bar{\alpha}$  and  $\sigma_\alpha^2$ . (Hint: Recognize that we can always write  $\alpha_i = \bar{\alpha} + \varepsilon_i$ , where the  $\varepsilon_i$  have mean 0 by construction).

*Answer: We can write*

$$y_i = \bar{\alpha}x_i + \varepsilon_i,$$

*where*

$$\varepsilon_i = (\alpha_i - \bar{\alpha})x_i,$$

*which has mean 0 and conditional (on  $x$ ) variance  $x_i^2\sigma_\alpha^2$ . Then the OLS estimator*

$$\hat{\bar{\alpha}} = \frac{\sum_{i=1}^N y_i x_i}{\sum_{i=1}^N x_i^2}$$

*is an unbiased estimator of the mean of  $\alpha$  due to the mean independence property*

$$\begin{aligned} E(\varepsilon|x) &= E((\alpha - \bar{\alpha})x|x) \\ &= E((\alpha - \bar{\alpha})|x) \times x \\ &= 0 \end{aligned}$$

*since  $\alpha$  is i.i.d.*

*We can form an unbiased estimator of  $\sigma_\alpha^2$  from*

$$\hat{\sigma}_\alpha^2 = \sum_{i=1}^N \frac{(y_i - \hat{\bar{\alpha}}x_i)^2}{x_i^2}.$$

3. In class we discussed the linear probability model at length, in which a binary (i.e., 0-1) variable  $d$  was the dependent variable. In matrix notation, we wrote

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

where  $X$  was  $N \times K$  dimensional matrix, with  $N \gg K$  and rank  $K$ ,  $d$  and  $\varepsilon$  were  $N \times 1$  vectors and  $\beta$  was a  $K \times 1$  vector of unknown regression coefficients.

- (a) Derive the distribution of  $\varepsilon_i$  conditional on  $X_i$  for this model and show that  $E(\varepsilon_i|X_i) = 0$ .

*Answer: The disturbance can take two values in this case. The probability distribution is*

$$\begin{array}{ccc} d_i & \varepsilon_i & \text{prob}(\varepsilon_i) \\ 1 & 1 - X_i\beta & X_i\beta \\ 0 & 0 - X_i\beta & (1 - X_i\beta) \end{array} .$$

*This follows since*

$$E(d|X) = P(d = 1|X) = X\beta.$$

*It is easy to see that*

$$\begin{aligned} E(\varepsilon|X) &= (1 - X_i\beta)(X_i\beta) - X_i\beta(1 - X_i\beta) \\ &= 0. \end{aligned}$$

- (b) Derive the covariance matrix of the *OLS estimator* of the linear probability model (remember that the actual covariance matrix is a function of the true value of  $\beta$  and the sample  $X$ , and does not involve an estimate of  $\beta$ ).

*Answer: We know that OLS is unbiased, and*

$$\hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon,$$

*so that the covariance matrix is given by*

$$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X = (X'X)^{-1}XE(\varepsilon\varepsilon'|X)X(X'X)^{-1},$$

*where*

$$E(\varepsilon\varepsilon'|X) = \begin{bmatrix} (X_1\beta)(1 - X_1\beta) & 0 & \cdots & 0 \\ 0 & (X_2\beta)(1 - X_2\beta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (X_N\beta)(1 - X_N\beta) \end{bmatrix}$$

- (c) In practical situations we do not know  $\beta$ , so we must estimate the covariance matrix of the OLS estimator. It is natural to replace the unknown  $\beta$  with  $\hat{\beta}$ . What numerical issues might arise when computing the estimated covariance matrix of  $\hat{\beta}$  in this case?

*Answer: When we replace  $\beta$  in the matrix  $E(\varepsilon\varepsilon'|X)$  with  $\hat{\beta}$ , there is nothing to guarantee that  $X_i\hat{\beta} \in (0, 1)$  for all  $i$ ,  $i = 1, \dots, N$ . In this case the estimated covariance matrix of the OLS estimator is not well-defined (i.e., it is not positive definite, which is the basis requirement of a covariance matrix).*