Econometrics I

Midterm Examination

Fall 2006

Answer Sheet

Please answer all of the questions and show all of your work. If you think that a question is ambiguous, clearly state how you interpret it before providing an answer. All question parts have equal weight.

1. Consider the linear regression specification

$$y_i = X_i\beta + \varepsilon_i,$$

where ε_i has the following distribution for individual i. With probability π_i the disturbance term is drawn from a normal distribution with mean 0 and variance a and with probability $1-\pi_i$ the disturbance term is drawn from a normal distribution with mean 0 and variance b, where $a > b > 0$ and $\pi_i \in (0,1)$ for all i. The parameter vector β is of dimension K. You have access to a random sample consisting of N observations with information on y and X. Assume that the rank of $X'X$ is K. The parameter values β, a, b, and $\pi_i, i = 1, ..., N$ are unknown.

1. Is $\hat{\beta} = (X'X)^{-1}X'y$ an unbiased estimator of β? Why?

   Answer: The marginal density of the disturbance is given by

$$f_i(\varepsilon) = \sigma_1^{-1}\phi(\frac{\varepsilon}{\sigma_1})\pi_i + \sigma_2^{-1}\phi(\frac{\varepsilon}{\sigma_2})(1-\pi_i),$$

   where σ_j is the standard deviation associated with normal density j. The mean of ε_i is zero, no matter what the form of the dependence of π_i on X_i. Thus $E(\varepsilon_i|X_i) = 0$ for all i. Mean independence implies unbiasedness of the OLS estimator.

2. If $\pi_i = \pi$ for all i, is $\hat{\beta}$ the best linear unbiased estimator (BLUE) of β? Why?

   Answer: A mixture of normal random variables is itself normal. The variance of ε_i is given by

$$\pi_i\sigma_1^2 + (1-\pi_i)\sigma_2^2.$$

   If $\pi_i = \pi$, then all observations have the same variance, the covariance matrix of ε is diagonal and there is homoskedasticity. Thus, the Gauss-Markov theorem applies and $\hat{\beta}$ is BLUE.

3. If $\pi_i = \pi$ for all i, is $\hat{\beta}$ the maximum likelihood estimator of β?

   Answer: Following up the previous point, in the case of a constant π all disturbances are i.i.d. normal. The m.l. estimator for β is identical to the OLS estimator in this case. Thus besides being BLUE, the covariance matrix associated with $\hat{\beta}$ attains the Cramer-Rao lower bound. Thus the estimator is "best" even outside of the class of linear estimators in this case.

4. If $\pi_i = \pi$ for all i, provide a discussion of whether it would be possible to separately identify the parameters π, a, and b.

Answer: Not without additional information. The unbiased estimator of the common variance of ε is given by

$$s^2 = \frac{\sum_{i=1}^{N}(y_i - X_i\hat{\beta})^2}{N-K}.$$

But while we have an unbiased estimator of σ_ε^2, there is no way for us to determine the values of the three components σ_1^2, σ_2^2, and π. There are a continuum of combinations of values of these parameters that can yield the (derived) population parameter σ_ε^2.

2. An individual has "desired" hours of work given by

$$h^* = \beta_1 + \beta_2 w + \beta_3 Y + \varepsilon, \tag{0.1}$$

where w is the hourly wage rate, Y is nonlabor income, ε is a random variable which is independently and identically distributed in the population as $N(0, \sigma_\varepsilon^2)$, and β is an unknown parameter vector. You have access to information from a random sample containing N observations.

1. Say that the sample contained information on $\{h_i^*, w_i, Y_i\}_{i=1}^N$. Define the maximum likelihood estimator of β.

   Answer: With a normal, i.i.d. disturbance term, the maximum likelihood estimator of β is identical to the OLS estimator. Thus

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

   where $X = (1_{N\times 1}\ w\ Y)$.

2. More realistically, say that the sample contained information on $\{h_i, w_i, Y_i\}_{i=1}^N$, where

   $$h_i = \begin{cases} h_i^* & \text{if} \quad h_i^* > 0 \\ 0 & \text{if} \quad h_i^* \leq 0 \end{cases}$$

   Write down the conditional expectation of $h^*|w, Y, h^* > 0$. Based on this, what can you conclude concerning the consistency of the OLS regression estimates of (0.1) obtained from the subsample of individuals with positive hours observations.

   Answer:

   $$E(h^*|h^* > 0, X) = X\beta + E(\varepsilon|h^* > 0, X).$$

   Now

   $$\begin{aligned} E(\varepsilon|h^* &> 0, X) = E(\varepsilon|X\beta + \varepsilon > 0) \\ &= E(\varepsilon|\varepsilon > -X\beta) \\ &= \sigma_\varepsilon \frac{\phi(\frac{-X\beta}{\sigma_\varepsilon})}{1 - \Phi(\frac{-X\beta}{\sigma_\varepsilon})} \\ &= \sigma_\varepsilon \frac{\phi(\frac{X\beta}{\sigma_\varepsilon})}{\Phi(\frac{X\beta}{\sigma_\varepsilon})} \end{aligned}$$

   where the last step follows from symmetry of the mean-zero normal distribution (I didn't expect you to recall this exact expression, naturally). From the second line, it is enough to conclude that the expected value of the disturbance term is a function of X within the selected sample of individuals working positive hours. Thus mean independence does not hold within this subpopulation, and OLS applied to data from this subpopulation is neither unbiased nor consistent for the population β vector.

3. Describe the maximum likelihood estimator of β and σ_ε^2 based on the sample information in part 2.2.

   Answer: This is simply the Tobit estimator case. We have the log likelihood given by

$$\ln L = \sum_{i:h_i^*>0} \ln(\sigma_\varepsilon^{-1}\phi(\frac{h_i^* - X_i\beta}{\sigma_\varepsilon})) + \sum_{i:h_i^*\leq 0} \ln(1 - \Phi(\frac{X_i\beta}{\sigma_e})).$$

   We mentioned in class that $\ln L$ is globally concave in the parameters

$$(\frac{\beta}{\sigma_\varepsilon}, \frac{1}{\sigma_\varepsilon}).$$

   Since the log likelihood satisfies standard regularity conditions, the m.l. estimator is the unique root of the first order conditions associated with β/σ_ε and $1/\sigma_\varepsilon$. Once these values are found, the invariance of the maximum likelihood estimator property implies that the maximum likelihood estimator of β and σ_ε is given by

$$\hat{\beta} = \frac{\widehat{(\beta/\sigma_\varepsilon)}}{\widehat{(1/\sigma_\varepsilon)}},$$

   and

$$\hat{\sigma}_\varepsilon = \left(\widehat{(1/\sigma_\varepsilon)}\right)^{-1}.$$

4. Consider the even more realistic situation in which the sample information include Y for all individuals, h^* and w for those who work, and neither h^*nor w for those who do not. Describe the maximum likelihood estimator of all of the identified parameters in this case.

   Answer: This is a truncated data case, in which a right hand side variable is also missing in the subpopulation of those individuals for whom $h_i^* \leq 0$. Thus, the probability of this event cannot be computed given these data. Only the individuals with positive hours can be used to estimate the model parameters. The conditional log likelihood is given by

$$\ln L_C = \sum_{i:h_i^*>0} \ln(\frac{\sigma_\varepsilon\phi((h_i^* - X_i\beta)/\sigma_\varepsilon)}{\Phi(X_i\beta/\sigma_\varepsilon)}).$$

   The problem satisfies standard regularity conditions, and the m.l. estimator is the solution to the set of first order conditions associated with the partial derivatives of $\ln L_C$ with respect to the parameters β/σ_ε and $1/\sigma_e$.

3. Once again, consider the linear regression model

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

where

$$E(\varepsilon|X) = 0$$

and

$$E(\varepsilon_i\varepsilon_j|X) = \begin{cases} \sigma^2 & \text{if} \quad i = j \\ \alpha & \text{if} \quad |i-j| = 1 \\ 0 & \text{if} \quad |i-j| > 1 \end{cases}$$

1. Assuming that σ^2 and α are unknown, is it possible to compute a Feasible Generalized Least Squares (FGLS) estimator in this case? If so, describe your estimator.

   Answer: The covariance matrix of ε has the form

$$\psi = E(\varepsilon\varepsilon'|X) = \begin{bmatrix} \sigma^2 & \alpha & 0 & \cdots & 0 \\ \alpha & \sigma^2 & \alpha & \ddots & \vdots \\ 0 & \alpha & \sigma^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \alpha \\ 0 & \cdots & 0 & \alpha & \sigma^2 \end{bmatrix}.$$

   Then the GLS estimator is defined by

$$\tilde{\beta} = (X'\psi^{-1}X)^{-1}X'\psi^{-1}y.$$

   The FGLS estimator is a "plug in" estimator that substitutes a consistent estimator for ψ in this expression. The covariance matrix ψ is characterized by two parameters. Since the OLS estimator is unbiased and consistent, as $N \to \infty$ the OLS residuals converge to the population disturbances. That is,

$$\hat{r}_i = y_i - X_i\hat{\beta},$$

   and $\operatorname{plim} \hat{r}_i = \varepsilon_i$. Now

$$\begin{aligned} E(\varepsilon^2) &= \operatorname{plim} N^{-1} \sum \varepsilon_i^2 \\ &= \operatorname{plim} N^{-1} \sum r_i^2. \end{aligned}$$

   Then a consistent estimator of σ^2 is $\hat{\sigma}^2 = N^{-1} \sum r_i^2$. Once again, using the analogy principle, we know that

$$E(\varepsilon_i\varepsilon_{i-1}|X) = \alpha.$$

   Since the sample average of the corresponding OLS residuals converges to this population moment, a consistent estimator of α is

$$\hat{\alpha} = (N-1)^{-1} \sum_{i=2}^{N} r_i r_{i-1}.$$

   Use these two values to compose $\hat{\psi}$, and use this consistent estimator for ψ to define the FGLS estimator.

2. Modify the assumption on the conditional covariance function so that

$$E(\varepsilon_i\varepsilon_j|X) = \begin{cases} \sigma^2 & \text{if} \quad i = j \\ \alpha_s & \text{if} \quad |i-j| = 1,\ s = \min[i,j]. \\ 0 & \text{if} \quad |i-j| > 1 \end{cases}$$

   over all (i, j) pairs. Can you define a FGLS estimator for β in this case? Why or why not?

   Answer: No, you cannot because the number of parameters is increasing linearly in the sample size. The natural analog estimator for α_1, for example, is

$$r_1 r_2.$$

As sample size becomes indefinitely large, it is true that

$$\operatorname{plim} r_1 r_2 = \varepsilon_1 \varepsilon_2.$$

However, increases in sample size do not increase the number of draws from the relevant ε distributions, and hence the sample size of observations used to estimate α_1 remains at 1 as $N \to \infty$.

4. We saw that if an individual had a Cobb-Douglas utility function

$$U(x_1, x_2, x_3) = a_1 \ln(x_1) + a_2 \ln(x_2) + (1 - a_1 - a_2) \ln(x_3),$$

and if she faced prices $p_1,\ p_2,\ p_3$, and had income Y, then her expenditures on good j are given by

$$e_j \equiv p_j x_j^* = a_j Y.$$

Assume that you have access to a random sample of N observations with information on expenditures on each good and income for N individuals, where

$$e_{1,i} + e_{2,i} + e_{3,i} = Y_i \text{ for all } i.$$

There is heterogeneity in the population in that each individual's preference vector $a = (a_1\ a_2)'$ is an independent and identically distributed random variable with mean vector $\bar{a}$ and covariance matrix Σ_a.

1. Define the Best Linear Unbiased Estimator of $\bar{a}$ using "single equation" estimators. That is, define the BLUE of $\bar{a}_1$ using information only on e_1 and Y, and define the BLUE of $\bar{a}_2$ using information only on e_2 and Y.
Answer: In equation j, we can write

$$\begin{aligned} e_{ij} &= \bar{a}_j Y_i + (a_{ij} - \bar{a}_j) Y_i \\ &\Rightarrow\ e_{ij} = \bar{a}_j Y_i + \varepsilon_{ij}, \end{aligned}$$

where

$$\varepsilon_{ij} = (a_{ij} - \bar{a}_j) Y_i.$$

Then

$$\begin{aligned} E(\varepsilon_{ij}|Y_i) &= Y_i E(a_{ij} - \bar{a}_j) \\ &= 0. \\ VAR(\varepsilon_{ij}|Y_i) &= E(\varepsilon_{ij}^2|Y_i) \\ &= Y_i^2 \sigma_{jj}, \end{aligned}$$

where σ_{jj} is the j,j element of Σ_a.
Then the covariance matrix of the $\varepsilon_{\cdot j}$ is

$$E(\varepsilon_{\cdot j}\varepsilon_{\cdot j}'|Y) = \sigma_{jj}\psi,$$

where

$$\psi = \begin{bmatrix} Y_i^2 & 0 & \cdots & 0 \\ 0 & Y_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & Y_N^2 \end{bmatrix},$$

and ψ contains no unknown parameters. Then the GLS estimator is BLUE, since it "transforms" the heteroskedastic linear regression model into a homoskedastic model to which we can apply the Gauss-Markov theorem. Thus the BLUE for $\bar{\alpha}_j$ is

$$\widehat{\bar{\alpha}}_j = (Y'\psi^{-1}Y)^{-1}Y'\psi^{-1}e_j.$$

We saw that in this case, we can write

$$\widehat{\bar{\alpha}_j} = N^{-1}\sum_{i=1}^{N}(e_{ij}/Y_i).$$

Due to the adding up constraint (on expenditures), we only have to estimate the model for two of the goods. Then an unbiased estimator of the third, good 3, say, is simply $\widehat{\bar{a}}_3 = 1 - \widehat{\bar{a}}_1 - \widehat{\bar{a}}_2$.

2. Is a more efficient estimator of $\bar{a}$ available? If so, describe it and how it would be computed given the sample information available to you.

   Answer: As above, the 3 equation system is singular due to the adding up constraint, so without loss of generality we consider estimation of the two equation system involving goods 1 and 2. Then the two equation system can be written as

$$e = X\alpha + \varepsilon$$

   where

$$\begin{aligned} e &= \begin{bmatrix} e_{\cdot 1} \\ e_{\cdot 2} \end{bmatrix}, \\ X &= \begin{bmatrix} Y & 0 \\ 0 & Y \end{bmatrix} = I_2 \otimes Y \\ \alpha &= \begin{bmatrix} \bar{a}_1 \\ \bar{a}_2 \end{bmatrix}, \\ \varepsilon &= \begin{bmatrix} \varepsilon_{\cdot 1} \\ \varepsilon_{\cdot 2} \end{bmatrix}. \end{aligned}$$

   We know that to obtain efficiency gains from system estimation, two conditions must hold. First, off diagonal elements of $E(\varepsilon\varepsilon'|X)$ must be present. In this case,

$$E(\varepsilon_{ir}\varepsilon_{js}) = \begin{array}{ccc} Y_i^2\sigma_{ii} & \text{if} & i=j \text{ and } r=s \\ Y_i^2\sigma_{ij} & \text{if} & i=j \text{ and } r\neq s \\ 0 & & \text{otherwise.} \end{array}$$

   Thus there are off diagonal elements in the covariance matrix for the system, thus there are potential gains from system GLS estimation.

   However, the second condition is that the regressors be different across equations. In this case, the same regressor, Y, appears in each. The most efficient estimator is then the one that applies OLS to each expenditure function after that function has been transformed so that the disturbances are homoskedastic within equations. Thus

$$\tilde{e} = \tilde{X}a + \tilde{\varepsilon},$$

where

$$\begin{aligned}
\tilde{e} &= \begin{bmatrix} e_{\cdot 1}/Y \\ e_{\cdot 2}/Y \end{bmatrix}, \\
\tilde{X} &= \begin{bmatrix} Y/Y & 0 \\ 0 & Y/Y \end{bmatrix} = I_2 \\
\tilde{\varepsilon} &= \begin{bmatrix} \varepsilon_{\cdot 1}/Y \\ \varepsilon_{\cdot 2}/Y \end{bmatrix},
\end{aligned}$$

where now

$$E(\tilde{\varepsilon}\tilde{\varepsilon}'|\tilde{X}) = \Sigma_a \otimes I_N.$$

Each of the two equations in the transformed system have the same "regressor" - 1. Thus the system GLS estimator is the same as the OLS estimator applied to each transformed equation separately, and SUR provides no efficiency gains over GLS equation by equation.