Microeconometrics **EXERCISE QUESTIONS** **S.C. AHN**
Spring 2003

Q1. (30 pts.) Consider the following statements independently. State clearly whether you agree or do not agree and make a short comment on each statement.

(1) For a panel data model, it is impossible to check whether the random-effects GLS estimator is consistent or not.
(2) The 2SLS estimator is consistent as long as the instrumental variables used are uncorrelated with the model error terms.
(3) For a SUR model, the OLS estimators obtained estimating all of the equations in the model jointly are more efficient than the OLS estimators obtained estimating each equation separately.
(4) The 3SLS estimators are always more efficient than the 2SLS estimators, except the cases in which all of the equations in a system are exactly identified.
(5) For the systems of simultaneous equations with autocorrelated errors, the standard errors of the 2SLS or 3SLS estimators are inconsistent. Thus, it is impossible to make appropriate statistical inferences about such models.
(6) I always use GLS to estimate the regression models with binary dependent variables.

Q3. (20 pts) Consider:

$y_1 = \gamma_{21}y_2 + \beta_{21}x_2 + \varepsilon_1$;
$y_2 = \gamma_{12}y_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_2$;
$y_3 = \gamma_{13}y_1 + \gamma_{23}y_2 + \beta_{13}x_1 + \beta_{23}x_2 + \varepsilon_3$,

where the y's and the x's are all $T\times 1$. Here, the y's are endogenous while the x's are predetermined. The first equation is identified.

(1) (5 pts.) Check the identification of the second equation.
(2) (5 pts.) Check the identification of the third equation, assuming $\beta_{13} = 2$.
(4) (10 pts) Consider 2SLS estimation of the <u>first</u> equation. (i) Explain how to obtain the 2SLS estimate of $\delta_1 = (\gamma_{21},\beta_{21})'$. Explain your estimation procedure in a manner that those who are familiar with OLS can understand. (ii) Explain how to estimate the covariance matrix of your 2SLS estimates.
(5) (10 pts.) Let $\hat{\delta}_1 = (\hat{\gamma}_{21}, \hat{\beta}_{21})'$ be the 2SLS estimator for the first equation. Assume:

$$\begin{pmatrix} \hat{\gamma}_{21} \\ \hat{\beta}_{21} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}; Cov\begin{pmatrix} \hat{\gamma}_{21} \\ \hat{\beta}_{21} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

Compute an appropriate statistic for testing H_o: $\gamma_{21}\beta_{21} = 1$.

Q4. (30 pts.) The exponential probability density function is given by $f(x;\theta_o) = \theta_o e^{-\theta_o x}$, where

$x > 0$. Let $\{x_1, \ldots, x_T\}$ be a random sample from an exponentially distributed population.

(1) (10 pts.) Write down the log-likelihood function and show that MLE of θ ($\hat{\theta}_{ML}$) = $1/\bar{x}$.

(2) (10 pts.) With $\bar{x} = 2$ and T = 200, estimate var($\hat{\theta}_{ML}$).

(3) (10 pts.) With $\bar{x} = 2$ and T = 200, compute a Wald statistic for H_o: $\theta + 2\ln(\theta) = 1$.

Q5. (20 pts.) The Diamondback baseball club hires a consulting firm to estimate the demand for baseball tickets and how it is affected by a variety of factors. The consulting firm obtains from all of the Major League (ML) teams data on paid attendance at each individual baseball games, and does an OLS regression of attendance on the following explanatory variables: (i) ticket price; (ii) the number of games won last season; and (iii) the price of hot dog in the stadium. Assume that these regressors are nonstochastic.

Now consider the following complications independently. For each, indicate whether it would lead you to use a different method than the consultants' regression. If so, indicate what method you would then use. For each case, assume that the regressors were observed for every game.

(1) The data only indicate whether the actual attendance of each game was above or below 10,000.

(2) The paid attendance is reported as 10,000 for the games with actual attendance below 10,000.

(3) Some owners of ML teams refused to provide any information because they were in bad mood when they were asked.

(4) Some owners of ML teams refused to provide the information about attendance (but provide the information about the regressors) because their attendance rates were embarrassingly low.

Key:

Q1: (1) Do not agree. Use the Hausman test based on the difference between the within and GLS estimators. (2) Do not agree. Zero correlation between instruments and model error terms is not enough. Instrumental variables should be correlated with regressors. (2) Do not agree. They are numerically the same. (3) Do not agree. 3SLS = 2SLS, asymptotically, when the errors are cross-equationally uncorrelated. (4) Do not agree. Use GMM estimators such as 2SIV or 3SIV. (5) Do not agree. Use probit or logit. Why? See class notes.

Q2: (1) $X_1 = X_2$ or $\sigma_{12} = 0$; (2) $X_1'y_2 = 0$.

Q3: (1) Not identified; (2) Not identified.

(3) (i) Let $Z_1 = (y_2,X_2)$ and $X = (X_1,X_2,X_3)$. Then, $\hat{\delta}_1 = [Z_1'P(X)Z_1]^{-1}Z_1'P(X)y_1$. This estimator can be obtained by the following two steps. In the first stage, regress y_2 on X and get the fitted value $\hat{y}_2$. Then, in the second stage, regress y_1 on $\hat{y}_2$ and X_2. (ii) Let $\hat{\sigma}_{11} = T^{-1}(y_1-Z_1\hat{\delta}_1)'(y_1-Z_1\hat{\delta}_1)$. Then, $Cov(\hat{\delta}_1) = \hat{\sigma}_{11}[Z_1'P(X)Z_1]^{-1}$.

(4) Use the Wald statistic: $W_T = 1/11$.

Q4: (1) $l_T = T[\ln(\theta)-\theta\bar{x}]$; $\hat{\theta}_{ML} = 1/\bar{x}$.

(2) $\frac{\partial^2 l_T}{\partial\theta} = -\frac{T}{\theta^2} \rightarrow -H_T(\theta) = \frac{T}{\theta^2}$.

$\rightarrow \text{var}(\hat{\theta}_{ML}) = [-H_T(\hat{\theta}_{ML})]^{-1} = \hat{\theta}_{ML}^2/T = 1/(T\bar{x}^2) = 1/800$.

(3) $w(\theta) = \theta + 2\ln(\theta) - 1$; $W(\theta) = 1 + 2/\theta$.

$\rightarrow W_T = [1/2+2\times\ln(1/2)-1]^2/[(1+4)^2/800] = 113.85941$.

Q5: (1) Do probit instead.

(2) This is the case where you observe $y_t = \max\{y_t^*,10000\}$. The likelihood function is given:

$$\Sigma_{y_t>10,000}\{-.5\ln(2\pi)-\ln(\sigma)-.5(y_t-x_t'\beta)^2/\sigma^2\}+\Sigma_{y_t=10,000}\Phi[(10,000-x_t'\beta)/\sigma].$$

(3) This is the case of selection in which selection is random (the case which we do not have to worry about). One might want to define and estimate a probit equation for "mood" that discribe the data selection rule. However, this rule is nothing to do with actual attendance (as long as owners' current temporal moods are not correlated with

attendance at past games). So, I simply would do OLS on the data with observed attendance only.

(4) This is the case in which selection is not random. I would first estimate a probit equation for say, “embarrassingly low attendance” (ELA). Then, I’d estimate the attendance equation by the Heckman's two step procedure. (Not tobit, because we do not know the threshold values (L) for the observability of attendance). This two step method is necessary because ELA should be correlated with actual attendance.