Cornell University
Department of Economics

Econ 620 – Spring 2008
Instructor: Professor Kiefer
TA: Jae Ho Yun

Problem Set #5

(Due: April 18, Friday)

1. (Final 2002) You are interested in estimating a salary equation for professional horn players(sadly, there aren't many of these). Your specification is

$$y_i = x_i'\beta + \delta t_i + \varepsilon_i$$

where y_i is the log wage, x includes observed individual characteristics (years at Eastman School of Music, years in new orleans...) and t is talent. Unfortunately, you do not observe talent, but you observe an indiciation of talent, a score in a recent contest of horn players, t^*. You are also lucky enough to have some observable characteristics not included in the x, namely z.

   (a) Is OLS regression of y on x and t^* appropriate?

   (b) Give an alternative estimator.

   (c) how would you test whether OLS is appropriate, or whether some alternative technique is required?

2. Probit: Your dependent variables are indicators d_i, (0 or 1) and your model is $\Pr(d_i = 1) = \Phi(x_i\beta)$ is a scalar. Suppose you fit this model by nonlinear least squares.

   (a) How to get NLS estimator? Describe the procedure.

   (b) What is the variance of d_i conditioning on x_i?

   (c) Suppose you use your estimator from a and your formula from b to "correct" for heteroskedasticity and rerun your nonlinear least squares. Is your second-round estimator "better" than your first-round estimator from a)? In what sense?

   (d) We know that Maximum likelihood is the most efficient. How to get ML estimator? Describe the procedure.

3. (Final 2003) Consider one equation in a simultaneous system:

$$\begin{aligned} y_1 &= Y_2\gamma + X_1\beta + \varepsilon \\ &= Z\delta + \varepsilon \end{aligned}$$

where X_1 is the $N \times K_1$ matrix of included exogenous variables. Let X be the $N \times K$ matrix of all the exogenous variables throughout the system. You decide to fit the model by $2SLS$ and hence construct $\widehat{Y}_2$ (a. How?) and do the second state regression to obtain your coefficient $\widehat{\delta}$. Upon thinking things over, you decide that if instruments are good for Y_2, they are probably good for y_1 as well. You find $\widehat{y}_1$ (the same way you found $\widehat{Y}_2$) and regress $\widehat{y}_1$ on $\widehat{Y}_2$ and X_1 to obtain the coefficient vector $\widehat{\delta}^*$. b) Show that $\widehat{\delta} = \widehat{\delta}^*$.

4. (Final 2007) Consider the equation $y = X_1\beta + Y\gamma + \varepsilon$ where X_1 is $k_1 \times N$ and Y is $G \times N$. Y is potentially endogenous and you have a $k \times N$ matrix of instruments X, which includes X_1. Write the equation $y = Z\delta + \varepsilon$.

   (a) What is the minimum value of k needed to identify δ?
   Now assume that Y is not actually endogenous.

   (b) What is the asymptotic variance of the OLS estimator $\widehat{\delta}$?

   (c) What is the IV estimator $\widehat{\delta}_{IV}$ and its asymptotic variance?

   (d) What is the asymptotic variace of $\widehat{\delta}_{IV} - \widehat{\delta}$?

   (e) What is the rank of this covariance matrix?

   (f) What is the asymptotic covariance between $\widehat{\delta}$ and $\widehat{\delta}_{IV} - \widehat{\delta}$?

5. Consider the system

$$y_1 = \alpha_1 + \beta_1 x + \gamma y_2 + \varepsilon_1 \quad (1)$$
$$y_2 = \alpha_2 + \beta_2 f(x) + \varepsilon_2 \quad (2)$$

   where the errors are jointly normally distribute with mean zero and covariance matrix V. Investigate the identification of equation 1, when a) $f(x) = 1 + x$; (b) $f(x) = 1 + x + x^2$; (c) $f(x) = 1 + \exp(x)$. How could you estimate the model?

6. (Final 2006) You specify a linear regression model $Ey = X\beta$. Oddly, your X matrix is $n \times 2$ with the pattern $x_i = (1 \;\; 0)$ for i odd and $(0 \;\; 1)$ for i even. You have neglected to include a constant term. Assume $E(\varepsilon) = 0$ and $V(\varepsilon) = I_n$.

   (a) What are the variances and covariance of the OLS estimator for β?

   (b) You have read an article about GMM and now conclude that OLS is for wimps. After some experimentation, you realize that the moment condition $E(y - X\beta) = 0$, with empirical counterpart $n^{-1}\sum(y_i - x_i\beta)$ isn't going to get you very far. So, you split the sample into two groups, odd and even observations and solve two equations, $(n_{even})^{-1}\sum(y_i - x_i\beta)$ and $(n_{odd})^{-1}\sum(y_i - x_i\beta)$ (the sums are over the observations in each of the groups) for each of the two groups, to get estimators β^*. Compare the variance of β^* with that of the OLS estimator.