Problem Set #5

ECONOMICS 240B
SPRING 2010

Due April 21

Review Questions (not to turn in) from Ruud's text:

Chapter 21: Exercise 21.8
Chapter 22: Exercise 22.16
Chapter 27: Exercises 27.4, 27.5
Chapter 28: Exercises 28.4, 28.5, 28.11

Questions to turn in:

1. Consider two scalar random variables y_i and x_i that are related by the simple linear model with no intercept,

$$y_i = \beta_0 \cdot x_i + u_i,$$

which is assumed to satisfy the two moment conditions

$$0 = E[u_i] = E[u_i x_i].$$

For convenience, the data are assumed to be i.i.d. draws from a joint distribution satisfying the moment conditions; the first two moments of x_i are $E[x_i] \equiv \mu$ and $E[x_i^2] \equiv \tau^2$, the (unconditional) variance of u_i is σ^2, and $E[u_i^2 x_i] \equiv \delta$ and $E[u_i^2 x_i^2] \equiv \gamma$. (All of these parameters are, in principle, unknown.)

A. Find the asymptotic distributions of the method-of-moments estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ which minimize the GMM criterion $[\bar{m}(\beta)]'\hat{A}_j\bar{m}(\beta)$ for $j = 1, 2$, where $\bar{m}(\beta)$ is the vector of sample analogues to the moment conditions and the weighting matrices for each estimator are defined as

$$\hat{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \hat{A}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

(that is, the estimators which use each moment condition separately). Be sure to cite any needed identification conditions explicitly (but assume all other needed regularity conditions hold implicitly).

B. Derive the asymptotic variance of the optimal GMM estimator based on both of the moment conditions.

C. Suppose u_i and x_i just happen to have $\delta = \sigma^2 \cdot \mu$ and $\gamma = \sigma^2 \cdot \tau^2$ (which would follow if u_i and x_i were in fact independently distributed), but that this fact is unknown (so these extra moment conditions are not imposed in the estimation of β_0). Show that, for this special case, the optimal GMM estimator has the same asymptotic distribution as one of the estimators in part **A.** above.

2. Suppose you wanted to model voting behavior for individual voters in terms of the "closeness" of an individual's views to a particular candidate's. An underlying latent (unobservable) variable y_i^* indicates the "discrepancy" between the candidate's views and the voter's, which is modelled in the usual way as a linear function of some observable covariates $\mathbf{x}_i$ and an error term ε_i :

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i,$$

where ε_i is assumed to have a known c.d.f. $F(\cdot)$ (e.g., logistic). The individual is assumed to vote for the candidate if the magnitude of the latent variable, $|y_i^*|$, is less than some unknown threshhold value K..

Given a sample of N observations on $\mathbf{x}_i$ and y_i, the latter being an indicator variable for whether the individual voted for the candidate, give expressions for the maximum likelihood estimator, nonlinear least squares estimator, and nonlinear weighted least squares estimator (with nonrandom weights $w_i = w(\mathbf{x}_i)$) for the unknown parameters $\boldsymbol{\beta}$ and K. Also, find the best nonlinear weighted least squares estimator (i.e., find optimal weights $w_i^*(\mathbf{x}_i)$) and determine whether it is asymptotically efficient.

3. Suppose each member of a population travels to work by one of two modes, automobile or public transit. Let

$$y_i = 1\{\text{person } i \text{ travels to work by automobile}\}.$$

Assume that every member of the population has access to transit service, but some people do not have access to a car. For each person, then,

$$d_i = 1\{\text{person } i \text{ has access to an automobile}\}$$

characterizes their options. It is known that

$$d = 1(\gamma_0 + v_i > 0),$$

where γ_0 is constant and v_i varies over the population. For individuals with $d_i = 1$, it is assumed that that person chooses to travel by automobile if $\mathbf{x}_i'\beta_0 + u_i > 0$, and selects public transit otherwise. Here $\mathbf{x}_i$ is a K-vector, $K \geq 1$, that varies over the population, u_i is a scalar error term, and β is a conformable parameter vector. Thus

$$y_i = 1\{\mathbf{x}_i'\beta_0 + u_i > 0 \text{ and } d_i = 1\}.$$

The first component of $\mathbf{x}$ is a constant, and the remaining components are jointly continuously distributed. A random sample of N individuals are drawn from the population. Also, it is assumed that v_i, u_i, and $\mathbf{x}_i$ are mutually independent for each individual.

A. You are told that v_i and u_i have standard logistic distributions, but only observe the sample realizations of y_i and x_i (d_i is unobserved). Propose an asymptotically efficient method to estimate $\theta_0 = (\gamma_0, \beta_0')'$, and derive its asymptotic distribution [don't worry about regularity conditions].

B. Now you observe the sample realizations of y_i, $\mathbf{x}_i$, and d_i. Propose an asymptotically normal method to estimate β_0 alone, i.e., without simultaneously estimating γ_0. Will this be efficient? Explain.

C. Now suppose you are told that v_i is not distributed standard logistic but instead is standard normal. You recompute your answer to question A. above and find that the estimate of γ_0 changes but the estimate of β_0 remains the same. Explain why.

4. Consider the following "disequilibrium regression model with observed regime", consisting of the latent bivariate regression system

$$y_1^* = \mathbf{x}'\boldsymbol{\beta}_1 + \varepsilon_1$$
$$y_2^* = \mathbf{x}'\boldsymbol{\beta}_2 + \varepsilon_2$$

where

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim \mathfrak{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right) \equiv \mathfrak{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right),$$

and the observed dependent variables are

$$d = \mathbf{1}\{y_1^* \leq y_2^*\}$$
$$y = \min\{y_1^*, y_2^*)$$

so that the smaller of y_1^* and y_2^* is observed, along with an indicator d of whether y_1^*. One seeks estimators of the parameters $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\Sigma}$. The marginal distribution of the vector $\mathbf{x}$ of regressors is unspecified and unrelated to the parameters of interest.

A. For a random sample $\{(y_i, d_i, \mathbf{x}_i')\}_{i=1}^N$ of size N from this model, derive the form of the average log-likelihood function for the unknown parameters (conditional on the regressors). For concreteness, let $n(\varepsilon_1, \varepsilon_2 | \mu, \boldsymbol{\Sigma})$ denote the bivariate normal density function for $\mathfrak{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\phi(\cdot)$ and $\Phi(\cdot)$ denoting the univariate standard normal density and cumulative, as usual. Express the average log-likelihood using this notation, and simplify your expression as much as possible.

B. Derive the form of the average log-likelihood if only d_i and $\mathbf{x}_i$ are observed (i.e., the "outcome" variable y_i is unobserved), and discuss which functions of the parameters $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\Sigma}$ are identified and consistently estimable using only d_i and $\mathbf{x}_i$. Also, give an expression for the asymptotic covariance matrix of the maximum likelihood estimator of the identifiable parameter vector (denoted δ_0).

5. Consider an "M-estimator" $\hat{\theta}_M$ of a p-dimensional parameter vector θ_0 that solves a system of p estimating equations of the form

$$0 = \frac{1}{N} \sum_{i=1}^{N} \psi(z_i, \hat{\theta}_M) \equiv \frac{1}{N} \sum_{i=1}^{N} \psi_i(\hat{\theta}_M),$$

where z_i is i.i.d. and $E[\psi_i(\theta_0)] = 0$, $E[||\psi_i(\theta_0)||^2] < \infty$. Assume that $\hat{\theta}_M$ is consistent ($\hat{\theta}_M \xrightarrow{p} \theta_0$) and has the asymptotically-linear representation

$$\hat{\theta}_M \stackrel{A}{=} \theta_0 + \frac{1}{N} \sum_{i=1}^{N} H_0^{-1} \psi(z_i, \theta_0),$$

with

$$H_0 \equiv H(\theta_0),$$
$$H(\theta) \equiv -E\left[\frac{\partial \psi_i(\theta)}{\partial \theta'} \right].$$

A "one-step" alternative to the estimator $\hat{\theta}_M$ is

$$\hat{\theta}_{OS} \equiv \tilde{\theta} + \frac{1}{N} \sum_{i=1}^{N} [\hat{H}(\tilde{\theta})]^{-1} \psi(z_i, \tilde{\theta}),$$

for

$$\hat{H}(\theta) \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \psi_i(\theta)}{\partial \theta'}$$

and for $\tilde{\theta}$ being some initial estimator of θ_0. (The one-step estimator is a single Newton-Raphson step from $\tilde{\theta}$ toward $\hat{\theta}_M$.)

Assuming $\hat{H}(\hat{\theta}) \xrightarrow{p} H_0$ whenever $\hat{\theta} \xrightarrow{p} \theta_0$, verify the "**one-step theorem**": that is, if the initial estimator $\tilde{\theta}$ is $\sqrt{N}$-consistent and asymptotically normal, i.e.,

$$\sqrt{N}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, V_0)$$

for some covvariance matrix V_0, then show that $\hat{\theta}_{OS}$ is asymptotically equivalent to $\hat{\theta}_M$, i.e.,

$$\sqrt{N}(\hat{\theta}_{OS} - \hat{\theta}_M) \xrightarrow{p} 0.$$