Comprehensive Exam in Econometrics
Suffolk University
August 2007

Answer all questions. You have four hours to take this test. Time yourself wisely.

Question A1

Consider the classical linear regression model, $y = X\beta + \varepsilon$ where X is non-stochastic. Further let $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2\Omega$ where Ω is a symmetric positive definite matrix. The data represent a cross section of T observations.

a. Let $X\beta = X_1\beta_1 + X_2\beta_2$ where X_1 and X_2 are T x k_1 and T x k_2 matrices and $k = k_1+k_2$. Consider $\Omega = I_T$, where I_T is an identity matrix. Discuss the consequences of erroneously (i) excluding X_2 when it should have been included, (ii) including X_2 when it should have been left out.
b. You suspect heteroscedasticity in the above cross sectional study. Discuss the structure of the Ω matrix under heteroscedasticity and the implications of just ignoring it.
c. Further you believe that heteroscedasticity may be caused by the variations in X_2; in particular, $\sigma_i^2 = \sigma^2 e\,\mathrm{xp}(X_{2i}{}'\gamma)$. Describe the appropriate Breusch Pagan test for heteroscedasticity in detail. If detected, describe the GLS (WLS) procedure that takes care of the problem.

Question A2

a. Consider a classical linear regression model, $y_i = x_i'\beta + \varepsilon_i$, with a well behaved error term with $E(\varepsilon_i) = 0;\ E(\varepsilon_i^2) = \sigma^2;\ E(\varepsilon_i\varepsilon_j) = 0\ \ \forall\ i \neq j$. Further let ε_i be normally distributed with $f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right)$. Derive the MLE estimates of the parameters β, σ^2 and compute their variance-covariance matrix.
b. How would you test the significance of all or a subset of the above explanatory variables.
c. The response (dependent) variable in many regressions is an integer leading to what are called the count data models. For instance, it is interesting to study if the number of visits to a doctor (y) depends on the age and the type of medical insurance of the patient where, for each i, $y_i = 1, 2, \ldots \infty$. The Poisson distribution is widely used to study such models where, for $\lambda_i = \exp(x_i{}'\beta)$, the probability is given by $P(Y_i = y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}$. Briefly describe the likelihood function for the Poisson model used to estimate the unknown parameters, β.

(B1) Consider the following simple linear model

$$y_i = \beta x_i + u_i, \quad i = 1, \ldots, n$$

where x is a scalar and u has zero mean and variance σ_u^2. Suppose that $Cov(x, u) \neq 0$ and there is a variable z that is available to be used in estimation. Assume that x, y, and z all have zero means and variances σ_j^2 for $j = x, y, z$. Also, remember that the asymptotic bias of an estimator $\hat{\theta}$ for an unknown parameter θ is defined as $\text{plim}(\hat{\theta}) - \theta$ and the correlation coefficient between two variables a and b is defined as $Corr(a, b) = Cov(a, b)/(\sigma_a \sigma_b)$.

(a) Derive the asymptotic bias of the OLS estimator $\hat{\beta}$.

(b) What are the conditions for z to be a valid instrument?

(c) Suppose that there is a seemingly small, but nonzero, correlation between z and u. Derive the asymptotic bias of the instumental variable (IV) estimator $\tilde{\beta}$ and determine the effect of having a *weak instrument*, i.e., an instrument that is weakly correlated with the endogenous variable.

(d) If there is a seemingly small, but nonzero, correlation between z and u, would you still prefer IV to OLS? On what grounds and under what condition?

(B2) Consider the following simple production function

$$\log(Q_{it}) = \beta_{0i} + \beta_{1i} \log(L_{it}) + u_{it}, \quad i = 1, \ldots, n \text{ and } t = 1, \ldots, T$$

where Q_{it} is the output of firm i in year t and L_{it} is labor input in year t.

(a) Assume that $E(X_{it} u_{it'}) = 0$ for $t, t' = 1, \ldots, T$ where $X_{it} = [1 \ \log(L_{it})]$ and $E(u_i u_j^\top) = \sigma_{ij} \mathrm{I}_T$ for $i, j = 1, \ldots, n$ where $u_i^\top = [u_{i1}, \ldots, u_{iT}]$ for $i = 1, \ldots, n$.

(i) How would you estimate the system? Explain.

(ii) How would you test for $E(u_{it} u_{jt}) = 0$ for $i \neq j$? Do not derive the tests, just state the null hypothesis and the test statistic(s). What is the implication of failing to reject this null hypothesis?

(b) Now consider the following

$$\log(Q_{it}) = \beta_0 + \beta_1 \log(L_{it}) + u_{it}$$

where $u_{it} = s_i + \varepsilon_{it}$ and s_i is the firm's efficiency which is assumed to stay constant over time. Assume that $E(X_{it} \varepsilon_{it'}) = 0$ for $t, t' = 1, \ldots, T$ and $E(X_{it} s_i) = 0$ for $t, = 1, \ldots, T$ where X_{it} is defined above. How would you estimate the system *efficiently*? Explain, including all additional assumptions that may be necessary.