UCLA Fall 98 Ec203a

QUANTITATIVE METHODS

CORE EXAM Section I

Choose two out of the following three questions.

1. Let $X_1, X_2, \ldots, X_N$ be iid random variables with probability density function

$$f(x;\theta) = 2x\theta \exp[-x^2\theta] \quad x > 0 \quad \theta > 0$$

   (a) Find a one-dimensional sufficient statistic for θ.

   (b) Calculate the maximum likelihood estimator $\hat{\theta}$ and its asymptotic distribution.

   (c) Suppose $N = 200$ and the maximum likelihood estimator $\hat{\theta} = 0.4$. Test the hypothesis $\theta = 0.3$ against the hypothesis $\theta \neq 0.3$ at the 5% significance level using the Wald test.

   (d) Same as in c) using the Lagrange multiplier or Score test.

   (e) Same as in c) using the Likelihood Ratio test.

2. Let $X_1, X_2, \ldots, X_N$ be a random sample from the density

$$f(x;\theta) = \theta(1+x)^{-(1+\theta)}$$

for $x > 0$ and $\theta > 0$.

   (a) Find the maximum likelihood estimator for θ.

   (b) Find the maximum likelihood estimator for $1/\theta$.

   (c) Find the Cramer–Rao bound for unbiased estimators of $1/\theta$.

   (d) Is the minimum variance unbiased estimator for θ equal to the maximum likelihood estimator?

   (e) Suppose the maximum likelihood estimator for θ is equal to 1, and the number of observations is 100. Test the hypothesis $\theta = 1.1$ at the 5% level using a likelihood ratio test.

   (f) Repeat the test using a Wald test.

3. $(Y_1, Z_1), (Y_2, Z_2), \ldots, (Y_N, Z_N)$ are pairs of independent and identically distributed random variables, with common density

$$f_{YZ}(y, z|\lambda) = 2\lambda^2 \exp(-y\lambda - 2z\lambda),$$

for positive y and z and zero elswhere.

(a) Find a one–dimensional sufficient statistic for λ.

(b) Find the maximum likelihood estimator $\hat{\lambda}_{ML}$ for λ.

(c) Is the maximum likelihood estimator the minimum variance unbiased estimator?

(d) Find the normal approximation to the sampling distribution of $\hat{\lambda}$.

(e) Show that X_i, defined as the minimum of Y_i and Z_i, has an exponential distribution with parameter $3 \cdot \lambda$.

(f) Find the maximum likelihood estimator for λ based on $X_1, X_2, \ldots, X_N$.

Section Two

Answer two of the following three questions. Use a separate bluebook for this section of the examination.

Q-1. Given $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$ jointly multivariate normal, the conditional distribution of $\mathbf{x}_1$ given $\mathbf{x}_2$ is given by

$$f(\mathbf{x}_1|\mathbf{x}_2) = \mathcal{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

Use this result to analyze the following problems.

Suppose that three variables x, y, z are multivariate normally distributed, with means μ_x, μ_y, μ_z respectively. Denote the variances and covariances of these variables as σ_{ij}, for $i, j \in \{x, y, z\}$. Let the capital letters X, Y, Z denote n observations on each of the variables.

A. Denote the conditional mean of y as

$$E(y|x, z) = \alpha + \beta x + \gamma z.$$

Determine the constants α, β, γ as functions of the underlying mean and variance parameters μ_i, σ_{ij}, for $i, j \in \{x, y, z\}$.

B. Let $\epsilon = y - E(y|x, z)$. Derive the (conditional) variance of ϵ, σ_ϵ^2.

C. What is the variance of the least squares estimator $\hat{\beta}$? Express your answer in terms of σ_ϵ^2 and matrices involving X and Z.

Next consider the following simple regression model

$$y_t = a + bx_t + u_t.$$

Maintain the assumption that (x, y, z) are trivariate normal as above.

D. What is the conditional distribution $f(y_t|X)$? You may express your answer in terms of the parameters α, β, and γ (as well as other parameters).

E. What is the conditional mean of the simple least squares estimator $E(\hat{b}|X)$?

F. What is the (conditional) variance of u_t, σ_u^2?

G. Show that $\sigma_\epsilon^2 = V(\epsilon|X, Z) \leq \sigma_u^2 = V(u|X)$.

H. Give two conditions under which $\sigma_\epsilon^2 = \sigma_u^2$.

I. What is the conditional distribution of $\hat{b}$?

J. Show that $V(\hat{b})/\sigma_u^2 \leq V(\hat{\beta})/\sigma_\epsilon^2$.

K. What is the conditional (given X) distribution of the t-statistic for the hypotheses that $b = \beta$?

L. For each of the following conditions, describe the likely effect on the hypothesis $b > \beta$. That is, discuss the size and the power of the t-test you analyzed in (K) relative to this hypothesis.

   i. $\gamma > 0$, $\sigma_{xz} < 0$.
   ii. $\gamma < 0$, $\sigma_{xz} < 0$.
   iii. $\gamma = 0$, $\sigma_{xz} < 0$.

Q-2. Consider two, correlated AR(1) series x_t and z_t:

$$x_t = \rho_x x_{t-1} + \eta_{1t},$$
$$z_t = \rho_z z_{t-1} + \eta_{2t},$$

where $|\rho_x|$ and $|\rho_z|$ are less than one, and $\eta_t = (\eta_{1t}, \eta_{2t})$ is bivariate white noise, with covariance matrix

$$E(\eta_t\eta_t') = \Sigma_\eta = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

A. Derive the means, and variances σ_{xx} and σ_{zz} of x_t and z_t. Also derive the covariance between x_t and z_t, σ_{xz}.

B. Derive the first-order autocovariances, $\gamma_x(1)$, $\gamma_z(1)$, and the first order cross autocovariance $\gamma_{xz}(1)$ & $\gamma_{zx}(1)$ for x_t and z_t.

C. Suppose that the true model for y_t is given by

$$y_t = x_t\beta + z_t\delta + \varepsilon_t,$$

Here, ε is white noise, uncorrelated at all leads and lags with x_t. Instead of estimating the correctly specified regression above, you omit z_t from the regression.
Derive the probability limit of the least squares estimator $\mathbf{b}$ of β in the misspecified regression $y_t = \beta x_t + u_t$. Under what parameter restrictions does $\mathbf{b} \xrightarrow{p} \beta$? Under what conditions is $\text{plim}(\mathbf{b}) > \beta$, $\text{plim}(\mathbf{b}) < \beta$?

D. Let $\hat{u}_t = y_t - x_t\mathbf{b}$. Derive the limiting form of the first-order serial correlation $\gamma_{\hat{u}}(1)$ of $\hat{u}_t$ That is, what is the probability limit of $T^{-1}\hat{\mathbf{u}}'_{-1}\hat{\mathbf{u}}$. Under what conditions does $\hat{u}_t$ display positive, first-order serial correlation?

E. Under what conditions, if any, will $\hat{u}_t$ display negative first-order serial correlation? Note that this part of the problem is easier if you express $\gamma_{\hat{u}}(1)$ in terms of correlations ρ_x, ρ_z, and ρ_{xz}.

F. Based upon the results for u_t you have derived above in (D), comment on the ability of tests for residual serial correlation to offer indirect evidence of the omitted variable z_t.

Q-3. Again consider a short regression:

$$y_t = \beta^* x_t + u_t,$$

when the true model includes a variable z_t:

$$y_t = \cancel{\beta x_t + \delta z_t + \varepsilon_t} = \beta x_t + \delta z_t + \varepsilon_t$$

Assume that all variables have been centered at their respective sample means.

A. Derive the probability limit for the least squares estimator $\mathbf{b}$ in the short regression.

B. Derive the probability limit of the sample average of the sqared fitted residuals $\hat{u}_t^2$ from the short regression.

C. Let $\tilde{u}_t^2$ denote the difference between $\hat{u}_t^2$ and its sample mean. Evaluate each of the following auxiliary regressions. In each case, derive the probability limit of the regression coefficients to determine whether or not the auxiliary regression would provide evidence of the omitted variable z_t.
In each case, you may assume that the the cross products of the regressors in the auxiliary regressions converges to some probability limit that you do not need to specify. For example, if you are evaluating the regression $\tilde{u}_t^2 = \gamma_1 h_{1t} + \gamma_2 h_{2t} + \eta_t$, you may assume that

$$T^{-1}\mathbf{H}'\mathbf{H} \xrightarrow{p} \mathbf{Q}_H,$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2]$.

i. $\tilde{u}_t^2 = \gamma_1 x_t + \eta_t$,
ii. $\tilde{u}_t^2 = \gamma_1 x_t + \gamma_2 x_t^2 + \eta_t$,
iii. $\tilde{u}_t^2 = \gamma_1 z_t + \eta_t$,
iv. $\tilde{u}_t^2 = \gamma_1 z_t + \gamma_2 z_t^2 + \eta_t$.

D. Suppose that x_t and z_t are uncorrelated. How does this change the properties of these auxiliary regressions?

E. Suppose instead that both x_t and z_t are symmetrically distributed, and that they are independent of one another. How does this affect the properties of the preceeding auxiliary regressions?

F. Use these results to coment on the inclusion of linear and quadratic terms in tests for heteroskedasticity.

Section Three — Use a separate bluebook when answering this Section.

Answer any <u>two</u> of the following <u>three</u> questions.

1. Consider the following model:

$$y_{1t} = \alpha_{12} y_{2t} + \alpha_{13} y_{3t} + \beta_{11} x_{1t} + \beta_{12} x_{2t} + \varepsilon_{1t} \tag{1}$$
$$y_{2t} = \alpha_{21} y_{1t} + \beta_{22} x_{2t} + \varepsilon_{2t} \tag{2}$$
$$y_{3t} = \beta_{31} x_{1t} + \beta_{32} x_{2t} + \beta_{33} x_{3t} + \varepsilon_{3t} \tag{3}$$

The joint distribution of all of ε_{jt}'s, conditional on all of the x's, $j = 1,2,3$, $t = 1,\ldots,N$, is normal with mean zero and

$$Cov(\varepsilon_{js}, \varepsilon_{kt}) = 0, \text{ if } j \neq k \text{ or } s \neq t;$$
$$Var(\varepsilon_{jt}) = \sigma_j^2, j = 1,2,3.$$

(a) Which equations are identified in the above system? For each identified equation, which restrictions on the covariances of the disturbances are necessary for the equation to be identified? Briefly explain and justify your answers.

(b) Carefully describe how you would implement two-stage least squares estimators for the equations which are identified. Also, characterize, as precisely as possible, the asymptotic distributions for these estimators.

(c) Describe the form of the three stage least squares estimator for the parameters in equations that are identified.

(d) Show that the asymptotic means and variances of estimators in (b) and (c) are the same.

(e) Someone has suggested that, as far as estimation is concerned, equation (3) can be eliminated from the above system of equations and that y_{3t} can be reclassified as exogenous in the first two equations. Carefully evaluate this suggestion.

2. Suppose you are asked to estimate the following Cobb-Douglas production function:

$$y_i = \alpha x_{1i}^{\beta_1} x_{2i}^{\beta_2} \varepsilon_i \tag{4}$$

where y_i is the annual output of a highly perishable good, e.g., tomatoes, produced by the i^{th} farm, x_{1i} is the amount of labor (person-hours) that used, x_{2i} is the amount of land (acreage) under cultivation, ε_i is a farm-specific stochastic component that has an unconditional zero mean and constant variance, and α (or $\log\alpha$), β_1, and β_2 are parameters to be estimated. Assume you have data on a random sample of N farms that are geographically dispersed across the U.S. and across different markets for y.

(a) Suppose that you assume that ε_i represents the i^{th} farmer's entrepreneurial skills which is another input into the annual production of y.

(i) With this added assumption, discuss the statistical properties of estimating the $\log\alpha$, β_1, and β_2 using a least squares estimator. Explain your answer and the role that the above characterization of what ε_i represents affects your answer.

Suppose you are also given data on the prices of land and farm labor that prevailed in the markets in which the farms in your sample are located. Denote these prices as p_{1i} and p_{2i}. Assume that competitive conditions prevail in the markets for the two observed factors and that none of the farms in your sample have any power in the (factor) markets in which they operate.

(ii) How might you use this additional price data to obtain unbiased or consistent estimators of $\log\alpha$, β_1, and β_2? Provide a careful explanation of the estimation method(s) you would use to exploit this data, what assumptions about this data you would need to maintain, and what small-sample or asymptotic properties for the parameters you would expect to result from your estimation methodology.

(b) In contrast to the characterization in part (a), suppose that ε_i the weather that prevailed during the season in which the commodity y was raised. Furthermore, assume that weather conditions are difficult, if not impossible, to forecast at the time that planting decisions for y were made by farmers.

(i) With this new characterization of ε_i, again discuss the properties of estimating the $\log\alpha$, β_1, and β_2 using a least squares estimator. Be sure to explain your answer and the way in which this second characterization of ε_t affects your conclusion.

(ii) If you think that the least squares estimator in (i) will be biased (or inconsistent), propose at least one alternative estimator which would yield desirable statistical properties for $\log\alpha$, β_1, and β_2. Otherwise, do not answer this part of the question.

3. Consider the binary logit model for which the probability of choosing $y_t = 1$ is:

$$P(x_t) \equiv \Pr(y_t = 1 | x_t) = \frac{1}{1 + e^{-\gamma - \delta x_t}}. \tag{5}$$

where x_t is a $K \times 1$ vector and γ and δ are parameters to be estimated.

(a) Construct the Lagrangian Multipler (LM) test statistic for testing the hypothesis that all of the slope coefficients (but not the intercept) are equal to zero for the logit model. What will be the distribution of this statistic under the null hypothesis?

(b) Prove that the LM statistic in part (a) is equal to NR^2, where N is the sample size and the R^2 is for the regression of $[y_t - P]$ on x_t and $P = \frac{\sum_{t=1}^{N} y_t}{N}$. [HINT: In working your way through this part, you will find it easier to do all of your analysis conditional on x_t, for all $t = 1, \dots, N$.]