UCLA

Spring 1999

Econometrics

There are twelve questions. Answer any four of them, in separate blue books.

1. Consider inference for a scalar parameter β characterized by the moment equations:

$$E(Y - X \cdot \beta) \cdot \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = 0.$$

   (a) Discuss the standard (Hansen) GMM estimator for this problem.

   (b) How would you compute the estimator?

   (c) What is the large sample distribution for the estimator?

   (d) How can you test the validity of the two moment equations?

   (e) How does the GMM estimator compare to the two-stage-least-squares estimator? Which one is more efficient?

   (f) Describe the empirical likelihood estimator for this case.

2. Consider the standard linear regression model with $Y|X \sim \mathcal{N}(X'\beta, \sigma^2)$.

   (a) What is the large sample variance of the maximum likelihood estimator?

   (b) What is the heteroskedasticity consistent variance?

   (c) How can you test for heteroskedasticity using the information matrix test?

   (d) If you reject homoskedasticity in favor of a variance that increases with X, would would you expect quantile estimates to look like for different quantiles?

   (e) Describe how you would estimate the 0.1 quantile regression function, assuming it is a linear function of X, that is, $X'\beta_{0.1}$.

   (f) How would you estimate the asymptotic variance of $\beta_{0.1}$?

3. Consider the standard censored regression model:

$$y_i = \max\{0, x_i\beta + \varepsilon_i\}$$

   Provide an example of a situation where this type of model may be useful. Discuss estimation of the model under the following assumptions (you may assume that sampling is random):

(a) $\varepsilon_i | x_i \sim N(0, \sigma^2)$

(b) $\text{Median}(\varepsilon_i | x_i) = 0$

(c) ε_i is independent of x_i.

Make sure that you justify/motivate the estimation methods in cases (b)-(c) and that you state the objective functions that define the estimators.

4. Describe kernel estimation of a univariate continuous density function from i.i.d. data. Show consistency and asymptotic normality of the estimator, making sure that you provide sufficient conditions for the results to hold. What issues arise in applying nonparametric kernel estimation in practice?

5. Consider the following panel data model:

$$y_{it}^* = \alpha_i + \beta x_{it} + \epsilon_{it},$$

where

$$\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

and

$$\alpha_i \overset{\text{ind.}}{\sim} \mathcal{N}(\mu_\alpha, \sigma_\alpha^2).$$

The ϵ's and α's are assumed to be independent of the x's and of each other. x_{it} is a scalar. For $i = 1, \ldots, n$, and $t = 1, \ldots, T$, we observe x_{it} and

$$y_{it} \equiv \begin{cases} 1 & \text{if } y_{it}^* > 0 \\ 0 & \text{if } y_{it}^* \leq 0 \end{cases}$$

Assume a diffuse prior for β, a diffuse prior for μ_α, and a $Gamma(1/2, (.01/2))$ prior for σ_α^{-2}.

(a) Outline a procedure for simulating the posterior distribution of β. In outlining your procedure, be complete, but assume that you have access to a program that can automatically generate draws from the posterior distribution in the classical linear regression model with a conjugate prior. (In other words, you give the program the appropriate y and x variables and specify the prior, and it can generate posterior draws.) Explain how the results of your procedure can be used to obtain a posterior 95% interval for β.

(b) Show how to calculate a predictive distribution for a new observation $y_{n+1,1}$, conditional on $x_{n+1,1}$ and the data on units $1, \ldots, n$ (but not conditional on β or α_{n+1}).

(c) Suggest a modification of the model and the simulation algorithm to allow α_i to be correlated with the x_{it}.

6. Suppose that Y_t is an $m \times 1$ vector of variables, measuring continuously compounded returns for m asset types. Consider the second-order VAR

$$Y_t | Y_0 = y_0, \ldots, Y_{t-1} = y_{t-1} \sim \mathcal{N}(\delta + \Pi y_{t-1} + \Gamma y_{t-2}, \Sigma), \qquad t = 2, \ldots, T.$$

Assume Σ is known, and use a constant (flat) prior for δ, Π, and Γ.

(a) Describe how to calculate the posterior distribution of δ, Π, and Γ given $y_0, y_1, \ldots, y_T$.

(b) Obtain the distribution of Y_{T+2} conditional on δ, Π, Γ, Σ, y_{T-1}, and y_T.

(c) Suppose that $Y_{t,1}$, the first element of the vector Y_t, measures continuously compounded returns on a bond index. Define W_{T+2} to be the return (in period $T+2$) on one dollar invested at time T in the bond index. Explain how to calculate a predictive distribution for W_{T+2}, based on data up to time T.

7. Consider a trivariate $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t})'$ Gaussian VAR(2):

$$\mathbf{y}_t = B_1 \mathbf{y}_{t-1} + B_2 \mathbf{y}_{t-2} + \mathbf{u}_t.$$

(a) Express this VAR in a VECM form. When is this representation valid?

(b) Suppose that $\mathbf{y}_t \sim I(1)$, and that $\beta' \mathbf{y}_t \sim I(0)$, where β is a known vector. How does this affect the VECM representation?

(c) With β known, how could you estimate this system? What is the form of the limiting distribution of the estimated coefficients in your VECM?

(d) Express B_1 in terms of β, and the estimated VECM coefficients.

(e) Attempt to construct a Wald test for the (block) exogeneity of y_1. Let b_{ij}^k denote the $i, j - th$ element of B_k. What problem do you encounter?

(f) Use the following result to complete the construction and interpretation of your test. If $w \sim \mathcal{N}(0, V)$, where w is a $p-$vector, and V is rank $q < p$, then $w'V^- w \longrightarrow \chi^2_q$, where V^- is the generalized inverse of V.

(g) Next, suppose that β is unknown. How would you estimate the system?

(h) With β unknown, how is your test for Granger Non-Causality affected?

8. You wish to conduct Bayesian inference on a $k-$variable, Gaussian VAR(p). Suppose the model is represented as

$$\mathbf{y} = (I_k \otimes X)\mathbf{c} + \mathbf{u},$$

(a) How are $\mathbf{y}$, $\mathbf{c}, \mathbf{u}$, and X constructed? What are their dimensions?

(b) What is the structure of the covariance matrix for $\mathbf{u}$? In your answer, let Σ_U denote the covariance matrix for the $k-$vector of VAR innovations.

(c) Specify a prior distribution over the parameters that will lead to a "tractable" posterior. In your prior, impose the information that (a) own lags may be more important than cross lags, and that (b) low order lags are likely to be more important than longer lags. How does this prior differ from a natural conjugate prior?

(d) Specify the two conditional posteriors $p(\mathbf{c}|\Sigma_U, \mathbf{y})$ and $p(\Sigma_U|\mathbf{c}, \mathbf{y})$ that follow from the prior you specified in part (8c).

(e) Describe the process of generating draws from the posterior distribution. What is the name of the procedure you would use to generate draws?

(f) Suppose that you believe the system is cointegrated. How would this alter your Bayesian analysis? Be as specific as possible in how you would alter your prior, and in how you would alter generating draws from your posterior. Contrast the information you impose in your prior with the full information maximum likelihood approach of Johansen.

(g) Finally, how would you test for alternative orders of cointegration in a Bayesian framework? Again be as clear as possible about the sequence of steps you would take.

9. Consider a five variable SVAR(p) $\mathbf{y}_t = (y_1, \ldots, y_5)'$, of the form

$$B_0\mathbf{y}_t + B_1\mathbf{y}_{t-1} + \cdots + B_p\mathbf{y}_{t-p} = \varepsilon_t.$$

Let $B(L) = \sum_{i=0}^{p} B_i L^i$, and let $C(L) = B(L)^{-1}$. Here L is the lag operator: $L\mathbf{y}_t = \mathbf{y}_{t-1}$.

(a) Suppose that

$$C(1) = \begin{bmatrix} X & X & X & X & X \\ X & X & X & X & X \\ 0 & 0 & X & 0 & 0 \\ X & X & X & X & X \\ X & X & X & X & X \end{bmatrix}.$$

Here, an X indicates a non-zero element, while a 0 indicates a zero restriction. Also, $C(1) = C(L)|_{L=1}$, and similarly for $B(1)$.

i. What interpretation can you give these restrictions?

ii. What restrictions are implied on $B(1)$?

iii. How does this affect identification in this model? Specifically, what instruments are generated? What role do these instruments play in identification?

(b) Next, assume that theory also implies that

$$C_0 = \begin{bmatrix} X & 0 & X & 0 & X \\ X & X & X & X & X \\ 0 & X & X & 0 & X \\ X & X & X & X & X \\ 0 & 0 & X & X & X \end{bmatrix}$$

Impose the orthogonality condition that $\Sigma = E(\varepsilon_t \varepsilon_t')$ is diagonal.

i. Construct a table indicating which variables are endogenous, and which variables are predetermined for each equation.

ii. Are any of the equations of this system identified? If so, be explicit in describing what instruments are used for identification.

(c) Finally, examine the instrumental variables implications of the orthogonality conditions above.

i. Specify how, if at all, these covariance restrictions aid identification. If these conditions do identify an equation, be explicit as to what implied instruments are used to achieve identification.

ii. For those equations that are identified, where might we expect identification to be weak? What does weak identification mean?

iii. How would the identified elements of this system be estimated? Specifically, can the identified portion of this system be estimated one equation at a time? If not, how can it be estimated?

10. Consider a nonlinear autoregressive model of the form:

$$y_t = \rho y_{t-1}^{\alpha} + u_t.$$

Assume that $u_t \sim \mathcal{N}(0, 1)$. Let $\theta = (\alpha, \rho)'$.

(a) What is the optimal Method of Moments Estimator under the above normality assumption? In what sense is this estimator optimal? Be explicit regarding the moment functions $m_t(\theta)$ that you employ.

(b) What are two features of the moment functions $m_t(\theta)$ you constructed above?

(c) How could we test the above normality assumption? Construct a vector of statistics that will be (asymptotically) mean zero under the null hypothesis of correct specification. What is the limiting distribution of this vector? How could we construct a test statistic based upon this vector? What would be its distribution?

(d) Construct a test of overidentifying restrictions based upon the restriction that $\alpha = 1$. Denote the moment functions under this restriction as $m_t^r(\rho)$. What is the limiting distribution of this test statistic?

(e) Suppose the preceding moment functions $m_t(\cdot)$ and $m_t^r(\cdot)$ are replaced by

$$\tilde{m}_t(\theta) = \frac{1}{2k+1} \sum_{j=-k}^{k} m_{t-j}(\theta),$$

and

$$\tilde{m}_t^r(\rho) = \frac{1}{2k+1} \sum_{j=-k}^{k} m_{t-j}^r(\rho).$$

What is a possible motivation for such a replacement? What properties are assumed implicitly regarding k?

(f) Consider an alternative estimator defined by

$$(\hat{\phi}, \hat{\theta}) = \arg\max_{\theta} \min_{\phi} \left(-\sum_{t=1}^{T} \ln(1 + \phi' \tilde{m}_t(\theta)) \right).$$

Interpret this estimator within the framework of Smith's 1997 *Economic Journal* paper. What is the estimator $\hat{\rho}$ here?

(g) Next, analyze this same estimation principle when applied to the restricted moment conditions $\tilde{m}_t^r(\rho)$. How does this estimator $\hat{\rho}^r$ differ from the preceding estimator, $\hat{\rho}$?

Contrast this estimator $\hat{\rho}^r$ with the GMM estimator ρ^*, which is also based on the moment conditions: $E_{t-1}(m_t^r(\rho) = 0$. What limitations does $\hat{\rho}^r$ share with ρ^*? What advantages does it have? In your discussion, make use of the simulation results of West & Newey's 1994 *Review of Economic Studies* paper.

11. Discuss differences between Bayesian and classical approaches to inference in the context of time series with possible unit roots. Discuss both testing and estimation.

12. Discuss the role of economic theory in the estimation of demand systems in markets for differentiated products such as the market for automobiles. Include in the discussion exogeneity/endogeneity of prices, nesting, and independence of irrelevant alternatives.