UCLA
Economics 231A

Fall Semester 1998
Guido Imbens

Core Exam

1. Lancaster (1979) models unemployment durations using exponential distributions. Specifically, for someone with characteristics x, the duration T has density

$$f_{T|X=x}(t|x;\beta) = \exp(\beta' x)\exp(-t \cdot e^{\beta' x}).$$

   (a) Describe how you would calculate the maximum likelihood estimates for β.

   (b) Suppose some of the observations are censored from the right: you only know for those observations that the duration exceeds some length T_i, where T_i may differ for different individuals. How does the likelihood function change?

   (c) Lancaster is concerned with the bias on estimates of β of omitted variables. He assumes that these omitted variables are uncorrelated with the included variables. If the model relating the duration and regressors was linear (instead of exponential), would such omitted variables lead to bias? What bias in the estimates of β does this lead to in the current model?

   (d) Describe how you would test for misspecification of the type Lancaster is concerned with using a likelihood ratio test.

   (e) Suppose you want to estimate the standard errors for the estimates in (a) assuming an exponential distribution.

   (f) Suppose you want to allow for omitted variables, and are willing to assume that the omitted variable has a gamma distribution with mean one and variance σ^2. How would you go about estimating the parameters β and σ^2?

2. Buchinsky (1994) is interested in the wage distribution and estimates quantile regression models.

   (a) Suppose you model the median of log earnings given characteristics x as $x'\beta$. What is the objective function you use to estimate β?

   (b) How do you numerically solve for the estimator?

   (c) What is the complication in deriving large sample theory for the estimator, compared to the standard gmm case?

   (d) What is the approximate variance of $\hat{\beta}$?

   (e) Given an estimate for the β, what is an estimate for the median of the level of earnings?

(f) How does the objective function change if you are interested in the .75 quantile instead of the median or 0.50 quantile?

(g) Describe the bootstrap method for estimating standard errors for the estimates of β.

(h) The variance of $\hat{\beta}$ involve the density of the residuals at zero. How would you estimate that density?

Econometrics Comprehensive Exam Questions

Fall 1998

1. Suppose you want to estimate the effect on the outcome y of the variable X in a linear regression model. You have a large cross sectional sample of mcirodata avaliable on both y and X, however the variable X only varies across well defined and mutually exclusive subgroups – e.g. X might refer to some state law which is the same for all people who live in the same state. The framework can be described as follows:

$$y_{ij} = \alpha_0 + \beta X_j + u_{ij},\ i=1, \ldots, N_j;\ \text{and } j=1, \ldots, J$$

where j refers to the various subgroups across which X various, and i refers to individuals within the subgroups, so that the total sample size $N = N_1 + \ldots + N_J$.

(a) Assuming that u_{ij} are iid errors across the sample, explain how you would estimate β and its standard error efficiently?

(b) Now suppose that you suspect $u_{ij} = \alpha_j + \varepsilon_{ij}$, and both α_j and ε_{ij} are uncorrelated with X_j. Explain intuitively what effect this would have on the properties of your estimates in (a)? How would you estimate β and its standard error in this situation?

(c) How would you test the model as assumed in (a) versus (b)?

2. Consider the panel data binary response model

$$y_{it} = 1(y_{it}^{*} > 0), \quad i=1,\ldots,N;\ t=1,\ldots,T$$

where the latent variable $y_{it}^{*} = X_{it}'\beta + \varepsilon_{it}$, $\varepsilon_{it} = \alpha_i + u_{it}$, and u_{it} is i.i.d. $N(o, \sigma_u^2)$ across i and t.

(a) Explain carefully how you would set up a fixed effects framework for this model. Be sure to state carefully any assumptions you make.

(b) Explain carefully how you might set up a random effects scheme in the situation when you suspect that α_i is correlated with x_{it}. Again, be sure to state any assumptions you make.

(c) Discuss the advantages and disadvantages of the fixed-effects and random-effects approaches that you outlined in (a) and (b). How would this differ in a linear model (e.g. suppose we observed the latent variable y_{it}^{*})?

Use a separate bluebook for this section of the examination.

I. For any $(n \times n)$ matrix A the adjoint A^* is another $(n \times n)$ matrix with elements

$$a^*_{ij} = (-1)^{i+j}|A_{ji}|,$$

where A_{ji} is the $(n-1) \times (n-1)$ matrix that results from deleting the jth row and the ith column of A, and $|\cdot|$ denotes the determinant. For instance, if

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

then the adjoint is given by

$$A^* = \begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}' = \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}.$$

When A^{-1} exists, it is given by

$$A^{-1} = |A|^{-1}A^*.$$

Similar results hold for $(n \times n)$ lag operator matrices $\mathbf{A}(L)$. Let $\mathbf{A}^*(L)$ denote the adjoint.

A. If

$$A(L) = \begin{pmatrix} 1 - \beta_1 L & 0 \\ 2 & 1 - \beta_2 L \end{pmatrix},$$

what is $A^*(L)$?

B. Under what conditions on the VAR

$$\mathbf{A}(L)\mathbf{y}_t = \mathbf{u}_t$$

will

$$\mathbf{A}(L)^{-1} = |\mathbf{A}(L)|^{-1}\mathbf{A}^*(L)?$$

What do these conditions imply about alternative representations of the system? Hint: the relevant alternative representation is sometimes stated as the Wold Decomposition Theorem.

C. Assuming that the conditions you described in (B) hold, use the adjoint matrix $\mathbf{A}^*(L)$ to construct a VARMA — Vector Autoregressive, Moving Average representation for the VAR above:

$$\mathbf{B}(L)\mathbf{y}_t = \Theta(L)\mathbf{u}_t.$$

Identify $\mathbf{B}(L)$ and $\Theta(L)$ as functions of $\mathbf{A}(L)$. Note the order of the autoregressive and moving average components. What additional feature distinguishes this VARMA representation? Hint: examine the autoregressive component.

D. For a matrix C, let $[C]_{ij}$ denote the element of C corresponding to the ith row and jth column. Denote the covariance matrix of $\mathbf{u}_t$ as Σ. Suppose that the VMA component $\Theta(L)$ you identified in (C) is first order:

$$\Theta(L) = \Theta_0 + \Theta_1 L.$$

Let $\mathbf{v}_t = \Theta(L)\mathbf{u}_t$. Derive the autocovariance function for the ith element of $\mathbf{v}_t$.

E. Given your results from (D), provide a univariate representation for the ith element of $\mathbf{v}_t$. Use this to develop a univariate representation of the ith element of $\mathbf{y}_t$. Be sure to provide expressions that identify the innovation variance, and all autoregressive, and moving average terms.

F. Consider the simple bivariate VAR

$$y_{1t} = \beta y_{2t} + u_{1t}$$
$$y_{2t} = y_{2t-1} + u_{2t}.$$

What particular features does this system exhibit? Under what conditions (if any) does this VAR satisfy the conditions you described in part (B)? What is $\mathbf{A}(L)$ for this system? What is $\mathbf{A}^*(L)$?

G. Can you represent the system in (F) in a VARMA specification, as in part (C)? To answer this question, use the adjoint you derived in part (F), and determine whether the resulting VARMA representation is correct or not. Comment on the necessity and/or sufficiency of the conditions you provided in part (B), for the existence of these VARMA representations.

H. Suppose two series have univariate ARMA(1,1) representations

$$y_{1t} = \rho_1 y_{1t-1} + e_{1t} + \theta_1 e_{1t-1}$$
$$y_{2t} = \rho_2 y_{2t-1} + e_{2t} + \theta_2 e_{2t-1}.$$

Assume as well that $\rho_1 \neq \rho_2$. Under what conditions could these variables have come from a common VAR? Explain your answer.

II. In "Generalized Method of Moments Specification Testing" Newey discusses asymptotically optimal tests. The solid lines in Figure 1 plot chi-squared densities with five and fifteen degrees of freedom respectively. Suppose that these are the asymptotic distributions of two test statistics T_5 and T_{15}. The dashed lines represent the densities of these tests under a specific alternative hypothesis. Figure 2 plots the cumulative distribution functions for these two tests, again under the null and alternative hypotheses.

A. The dashed plots in Figure 1 are simple translations of the (solid) null densities. What distributions do such translations correspond to?

B. For 5% size tests, what are the approximate critical values for each test?

C. At these critical values, what are the approximate rejection probabilities for each test, under the specified alternative? Which test has greater power against this alternative?

D. From your answer to (C), what conclusion can you draw about the effect of degrees of freedom on the power of tests?

E. Let α_5 and α_{15} represent the sizes of the two tests. For $\alpha_{15} = 0.05$, what approximate size α_5 would render the two tests equally powerful against the given alternative? What does this illustrate about the relationship between the size and power of tests?

F. Next consider distinct alternatives for the two tests. Denote these alternatives as $H_A(5)$ and $H_A(15)$. Keeping $H_A(15)$ fixed, in what direction would the alternative $H_A(5)$ have to move to equalize the power $\mathcal{P}$ of the two tests:

$$\mathcal{P}(H_A(5)) = \mathcal{P}(H_A(15)).$$

That is, would the density for $H_A(5)$ have to move to the left, or to the right of its position in Figure 1, to satisfy these equality?

G. From your answer to part (F), what conclusion can you draw about the effect of noncentrality parameters on the power of tests?

Consider the r population moment conditions

$$E(g(z,\beta_0)) = \int g(z,\beta_0)f(z,c_0)dz = 0,$$

and the analogous sample moments

$$g_T(\beta) = \frac{1}{T}\sum g(z_t,\beta).$$

Newey analyzes moment misspecification in terms of deviations from the proposed distribution $f(z,c_0)$, that is, in terms of $c \neq c_0$. Let

$$h(\beta,c) = \int g(z,\beta)f(z,c)dz.$$

Consider the sequence of alternatives

$$c_T = c_0 + \delta/\sqrt{T}.$$

H. Examine the limiting behavior of $\sqrt{T}h(\beta_0,c_T)$. Be explicit in the assumptions you make.

I. Derive a limiting distribution for $\sqrt{T}g_T(\beta_0)$. Relate the mean of this limiting distribution to your result from part (H). Again be explicit in the assumptions you make.

J. Finally, derive the limiting distribution for $\hat{\beta}$, the GMM estimator, based upon the moments $g_T(\cdot)$ and a weighting matrix W_T. Relate the mean of this limiting distribution to the mean you derived in (I). Hint: first expand $g_T(\hat{\beta})$. Then use the first order condition that the GMM estimator satisfies.

K. Consider a special case, where

$$g_T(\beta) = \begin{pmatrix} U(\beta)'Z \\ U(\beta)'X \end{pmatrix} T^{-1}.$$

Here, $U(\beta) = Y - X\beta$ are residuals from a linear model. What weighting matrices W_{ls} and W_{iv} will yield the least squares and the instrumental variables estimators $\hat{\beta}_{ls}$ and $\hat{\beta}_{iv}$ respectively?

L. Are either of these weighting matrices optimal? Explain.

M. Based upon the weighting matrices in (K), relate the limiting distributions of $\hat{\beta}_{ls}$ and $\hat{\beta}_{iv}$ to the limiting behavior of $\sqrt{T}g_T(\beta_0)$.

N. Use your results from (M) to relate the limiting variance of the difference $\hat{\beta}_{ls} - \hat{\beta}_{iv}$ to the limiting variance of $\sqrt{T}g_T(\beta_0)$.

O. Discuss the relevance of these results to Hausman tests.

Figure 1

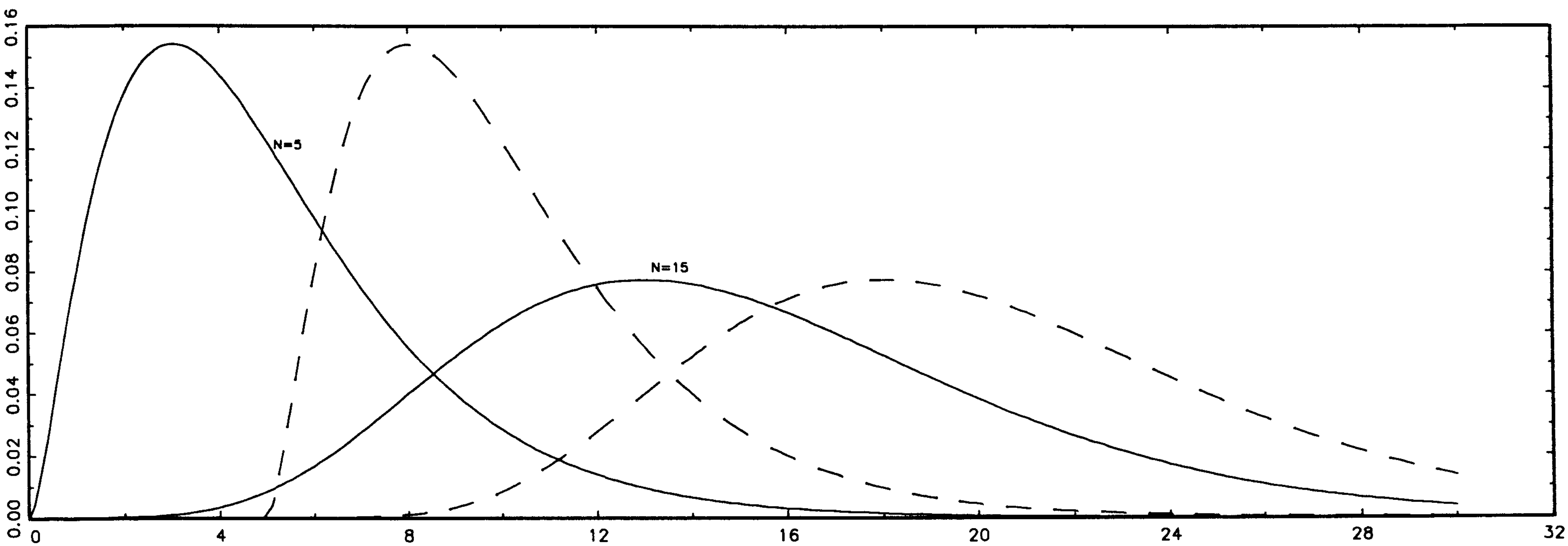

Figure 2