

University of California, Davis  
Department of Economics  
Econometrics Prelim

Date: June 27, 2007  
Time: 3 hours  
Reading Time: 20 minutes

**PRELIMINARY EXAMINATION FOR THE Ph.D. DEGREE**

---

The Exam is in **THREE** parts:

- (1) You must provide answers to all 3 parts: Part 1, Part 2, and Part 3.
- (2) There is choice within each part. Please read carefully.
- (3) Each part will count for approximately 33.33% of the total score.

**PART I.**

There are two questions given. Please choose and answer ONE of the following two questions.

1. Point Allocation: (a) 20%, (b) 20%, (c) 20%, (d) 20%, (e) 20%.

Assume that  $\varepsilon_t \in \mathbb{R}^2$  is iid  $N(0, I_2)$  where  $I_2$  is the  $2 \times 2$  identity matrix. Assume that  $x_t = (x_{1t}, x_{2t})$  is given by the following structural model

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} c_{11}(L) & c_{12}(L) \\ c_{21}(L) & c_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where

$$C(L) = \begin{bmatrix} c_{11}(L) & c_{12}(L) \\ c_{21}(L) & c_{22}(L) \end{bmatrix} = \sum_{j=0}^{\infty} C_j L^j$$

- Under what condition on the coefficient matrices  $C_j$  is the process  $x_t$  weakly stationary? Is it also strictly stationary under these conditions?
- What are the conditions on  $C(L)$  such that  $x_t$  has an AR representation? Assuming that these conditions are satisfied, write down the reduced form AR representation of  $x_t$ .
- Assume that  $c_{11}(1) = 0$ . Also, assume that the conditions in b) for an AR representation are satisfied and that the order of the AR representation is known and finite. How can you identify the parameters of the model?
- Assume that the conditions of c) hold. Describe in detail how you would proceed to estimate the structural parameters of the model.
- Assume that  $C_0 = I$  and that  $x_t$  has an autoregressive representation. What are the restrictions one has to impose on  $C_j$  such that  $x_{1t}$  does not cause  $x_{2t}$  in the sense of Granger? Are the conditions different if we are interested in lack of causality in the sense of Sims?

2. Point Allocation: (a) 20%, (b) 20%, (c) 20%, (d) 20%, (e) 20%.

Let  $\varepsilon_t$  be an iid sequence with  $E\varepsilon_t = 0$  and  $E\varepsilon_t^2 = 1$ . Assume that

$$\begin{aligned} x_t &= \alpha x_{t-1} + u_t \\ u_t &= \varepsilon_t + \theta \varepsilon_{t-1} \end{aligned}$$

with  $|\alpha| < 1$ .

- a) Is  $x_t$ , as determined by the above difference equations, stationary? Prove your answer by writing down an explicit representation of  $x_t$  in terms of  $\varepsilon_t$  and by showing that this representation is stationary (or by other means).
- b) Find the spectral density of  $x_t$ .
- c) You are interested in estimating the parameter  $\alpha$ . You use OLS

$$\hat{\alpha} = \frac{\sum_{t=1}^T x_t x_{t-1}}{\sum_{t=1}^T x_{t-1}^2}.$$

Describe the asymptotic properties of the OLS estimator.

- d) Is there a better estimator than  $\hat{\alpha}$  for  $\alpha$  that does not require estimation of  $\theta$ ? If so, write down a formula for a better procedure. What are the asymptotic properties of the proposed alternative procedure.
- e) Now assume that you want to estimate both  $\alpha$  and  $\theta$ . What are the conditions for identification of these two parameters.

**PART II.**

There are two questions given. Please choose and answer ONE of the following two questions.

**Question 1.** Point allocation: (a) 25%; (b) 25%; (c) 25%; (d) 25%.

Suppose that as a part of your dissertation you are estimating a linear regression model. You are concerned about parameter stability in your model and perform the F-tests suggested by Bai and Perron (Econometrica, 1998). Your results are printed below:

\*\*\*\*\*

Output from the Bai and Perron testing procedures

\*\*\*\*\*

(i) supF tests against a fixed number of breaks

The supF test for 0 versus 1 breaks (scaled by  $q$ ) is: 7.20  
The supF test for 0 versus 2 breaks (scaled by  $q$ ) is: 19.47  
The supF test for 0 versus 3 breaks (scaled by  $q$ ) is: 12.38

The critical values at the 5% level are (for  $k=1$  to 3):  
12.89 11.60 10.46

(ii) Dmax tests against an unknown number of breaks

The UDmax test is: 19.47  
(the critical value at the 5% level is: 13.27 )

\*\*\*\*\*

The WDmax test at the 5% level is: 21.64  
(The critical value is: 14.19 )

\*\*\*\*\*

(iii) supF( $l+1|l$ ) tests using global optimizers under the null

The supF( 2 | 1 ) test is : 12.45  
The supF( 3 | 2 ) test is : 17.43  
The supF( 4 | 3 ) test is : 2.81

\*\*\*\*\*

The critical values of supF( $i+1|i$ ) at the 5% level are (for  $i=1$  to 3 ) are:  
12.89 14.50 15.42

\*\*\*\*\*

(a) Draw conclusions about the number of breaks in your sample.

(b) This question is distinct from part (a). Suppose you are estimating a linear regression model and your data set has missing observations on the  $X$  variables. Describe how you would estimate the regression coefficients using multiple imputation. How would you interpret your estimate (be specific)?

(c) This question continues from part (b). In words, compare the properties of multiple imputation estimates with those derived from the EM algorithm. In your answer, state precisely the types of missingness for which these properties apply.

(d) This question is distinct from parts (a)-(c). Suppose you are forecasting some variable  $y$  conditional on a large vector of explanatory variables  $X$  using a linear model. Three estimation methods you could use to construct your forecasts are ordinary least squares, partial least squares, and principal components regression. List the advantages and disadvantages of each method.

**Question 2.** Point allocation: (a) 25%; (b) 25%; (c) 25%; (d) 25%.

Suppose that the solution of a representative agents portfolio allocation problem is given by the following Euler equation

$$P_t = E_t \left( \beta \frac{U'(C_{t+1})}{U'(C_t)} (P_{t+1} + D_{t+1}) \right)$$

where  $C_t$  is consumption of non-durables and services,  $P_t$  is the price of a financial asset, and  $D_t$  is the dividend paid out by the financial asset. Suppose that the agents utility function is given by  $U_t = C_t^\gamma$  and you have quarterly time series data on  $P_t$  and  $D_t$  for one financial asset. In addition, you have data on the riskfree rate of return,  $R_t$  and consumption  $C_t$ .

(a) Write down the sample moment conditions that you would use to obtain the GMM estimates of the unknown parameters  $\beta$  and  $\gamma$ . Be careful to justify your choice of moments.

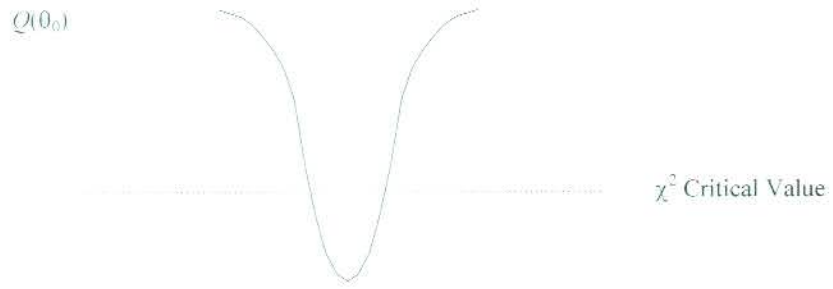
(b) Suppose that you estimate  $\beta$  and  $\gamma$  using more than two moment conditions. Write down Hansen's J-statistic and explain how you would use it to make inference about the validity of the moment conditions and the model implied by the above Euler equation.

(c) Suppose you were to estimate  $\beta$  and  $\gamma$  using a nonlinear least squares regression of  $P_t$  on  $\beta \left( \frac{C_{t+1}}{C_t} \right)^{\gamma-1} (P_{t+1} + D_{t+1})$ . Show that this estimator is inconsistent and explain intuitively why it is inconsistent.

(d) Let  $Q(\theta_0)$  denote the efficient GMM objective function evaluated at the parameter value  $\theta_0 = (\beta_0, \gamma_0)$ . Stock and Wright (Econometrica, 2000) showed that, under the null hypothesis  $H_0: \theta = \theta_0$ ,  $Q(\theta_0)$  is asymptotically distributed as Chi-square with degrees of freedom equal to the number of moment conditions. On the next page are three possible plots of  $Q(\theta_0)$  against  $\beta_0$ , labeled Example A, Example B, and Example C. For each of the three plots, draw conclusions about the properties of the model and the associated GMM estimates. Assume that the plots cover the entire parameter space and that  $\gamma_0$  is known.



Example A



$\beta_0$

Example B



$\beta_0$

Example C



$\beta_0$

PART III.

There are three questions given. Please choose and answer any TWO of the following three questions.

1. (Each sub-question is worth 20%)

You have the following 3-equation system with each equation identified, and  $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N(0, \Sigma)$ :

$$\begin{aligned} (1) \quad y_1 &= Y_1\beta_1 + Z_1\gamma_1 + \varepsilon_1 \\ (2) \quad y_2 &= Y_2\beta_2 + Z_2\gamma_2 + \varepsilon_2 \\ (3) \quad y_3 &= Y_3\beta_3 + Z_3\gamma_3 + \varepsilon_3 \end{aligned}$$

You are confident about the specification of the last two equations but uncertain about the specification of (1). To construct a test, you run 2SLS on each equation and compare it with running 3SLS on the entire system.

(a) Explain in words why you can construct a specification test by this procedure, demonstrating the effect of misspecification.

(b) To construct a formal test, you consider the estimates of each equation and consider the test statistic  $W$ .

$$W = \begin{bmatrix} \hat{\delta}_2 \\ \hat{\delta}_3 \end{bmatrix}_{2SLS} - \begin{bmatrix} \hat{\delta}_2 \\ \hat{\delta}_3 \end{bmatrix}_{3SLS}, \quad \text{where } \hat{\delta}_i = \begin{bmatrix} \hat{\beta}_i \\ \hat{\gamma}_i \end{bmatrix} \quad i = 1, 2, 3$$

Construct a large sample test under the null hypothesis of no misspecification, including the appropriate degrees of freedom.

(c) Can you think of an alternate test based on a more efficient estimation scheme?

(d) What happens to the test in (a) or (c) if equation (1) is just identified?

(e) Compare the above test to a (system) test of overidentifying restrictions.

2. (a) 40%, (b) 25%, (c) 35%.

Suppose we have data on a sample of young men's wages, along with demographic and other information and are told that the data obeys this model:

$$y_i^* = \alpha + x_i'\beta + \varepsilon_i \quad (i = 1, \dots, n)$$

where  $y^*$  is wages,  $x$  are covariates,  $\varepsilon$  is distributed  $N(0, \sigma^2)$ ,  $y^*$  is latent (that is, unobserved) and instead we observe  $y$ , where

$$y = \begin{cases} y^* & \text{if } y^* < c \\ c & \text{if } y^* \geq c. \end{cases}$$

(a). Suppose you estimate this model by ordinary least squares. Are the estimates of the parameters consistent? Analytically show your answer; please describe all your notation and

any additional assumptions you make. If you determine that the least squares estimator is not consistent for parameter estimation, propose another estimator. Show analytically.

(b). Suppose now you receive a new data set and discover that none of the realizations of your dependent variable is equal to  $c$ . How would you consistently estimate the parameters of this model? Analytically explain your choice of method.

(c). Now consider a slight variation of this model. Suppose that we have

$$y_{it}^* = \alpha_i + x_{it}'\beta + \varepsilon_{it} \quad (i = 1, \dots, n; t = 1, \dots, T),$$

where  $\varepsilon_{it}$  has the logistic distribution,  $n \rightarrow \infty$ ,  $T$  is fixed, and  $y^*$  is latent but we observe the realizations of a variable  $y_{it}$

$$\begin{aligned} y_{it} &= y_{it}^* \text{ if } y_{it}^* > c \\ &= c \text{ if } y_{it}^* \leq c. \end{aligned}$$

(d) Can you estimate the parameters of this model consistently? Outline a procedure. Please show analytically, and describe all your notation/assumptions. Are there any obvious advantages/disadvantages of your method if you are interested in estimating the effects of race on wages?

3. (a) 35%, (b) 35%, (c) 30%

Suppose you have data on a pool of unemployed individuals ( $i = 1, \dots, N$ ) each of whom has been unemployed for  $T_i$  periods. You also have information on their covariates  $x$ .

(a) You have decided that the model which best suits these data is

$$s(T|x, \theta) = \exp\{-\Lambda(T|x, \theta)\}$$

where  $s$  is the conditional survival function,  $\Lambda$  is the integrated hazard function, and  $\theta$  are parameters. You assume that the hazard function,  $\lambda$ , is Weibull. Write down a linear regression model that you will estimate, clearly showing the steps in going from the above equation to the regression model, and clearly defining any notation and assumptions you make.

(b) (You may ignore part (a)). Suppose that you want to determine whether the hazard rate is increasing, constant, or decreasing. Carefully defining all your notation/assumptions, describe an estimator which will allow the data to determine whether the hazard rate increases, decreases or remains constant over time. Describe a test for this finding.

(c). Suppose now you discover that your data consist of censored spells. In particular, right censored data. Here, you only know that the unemployment duration lasts at least  $c_i$  for individual  $i$ . Let  $d_i = 1$  if the  $i$ th spell is complete and  $d_i = 0$  otherwise. Describe an estimator for the model, be very clear in your notation and assumptions.