Econ. 240B, Fall 2001 Daniel McFadden

OLD FINAL EXAM QUESTIONS

Question 1 (30 minutes) You have data on the log of food consumption, c_{it}, and the log of family income, y_{it}, for a large number of households i = 1,...,N over two years t = 1,2. You are interested in estimating the income elasticity of food consumption. You suspect that income may be measured with error, but you are convinced that there is no correlation over time in either the disturbances in the food consumption equation or in the measurement errors in income if they exist. Propose a consistent estimator for the income elasticity, and explain why it is consistent. Describe how you would test whether there is measurement error in income.

Question 2 (30 minutes) As in Question 1, you have data on the log of food consumption, c_{it}, and the log of family income, y_{it}, for a large number of households i = 1,...,N over two years t = 1,2. You are convinced there is no measurement error in income, but you suspect that there is a "household" effect that makes the food consumption of one household persistently higher than that of another. What is the "fixed effects" estimator of the income elasticity of food consumption? What is the "random effects" estimator? Under what conditions will the first be consistent, but not the second?

Question 3 (30 minutes) The San Francisco Symphony has contract talks every second year. Let $d_t = 1$ if the talks in year t result in a strike, $d_t = 0$ otherwise, for years t = 0,2,...,2T. You postulate a latent variable model $y_t^* = \beta_1 + \beta_2 d_{t-2} + \varepsilon_t$, with $d_t = 1$ if and only if $y_t^* \geq 0$. You believe the disturbances ε_t are independent over time, and within each period have a standard normal distribution. How you would test the hypothesis that $\beta_2 = 0$? Be as specific as you can. The actual data can be summarized as follows:

Observations	$d_{t-2} = 1$	$d_{t-2} = 0$	Row sum
$d_t = 1$	5	7	12
$d_t = 0$	6	2	8
Column sum	11	9	T = 20

Question 4 (30 minutes) In the regression $y = \alpha + x\beta + z\gamma + z^2\delta + \varepsilon$, you are interested only in β. You have a random sample i = 1,...,N, and the disturbances satisfy Gauss-Markov conditions. Before collecting more data, you would like to know whether estimation of β would be biased by dropping the variables z and z^2 from the regression. How would test for this? Be precise about the conditions under which dropping the variables in question would <u>not</u> bias the estimate of β, and be sure that the test you propose is <u>consistent</u>; i.e., as $N \to +\infty$, the power of the test goes to one when the null hypothesis is false, and is bounded above by the significance level of the test if the null hypothesis is true. Describe the test statistic you would use, and describe its computation. Under the null hypothesis, state the asymptotic distribution of your test statistic, including degrees of freedom.

Question 5 (1 hour) You have time-series observations t = 1,...,T on variables x_t, y_{1t}, y_{2t}. You believe these data satisfy a model

$$y_{1t} = x_t\beta_1 + \varepsilon_{1t}$$
$$y_{2t} = x_t\beta_2 + \varepsilon_{2t}$$

in which x_t is exogenous and common to the two equations, and the coefficients are different in the two equations. The disturbances ε_{1t} and ε_{2t} have zero means given the x's, and the covariance structure:

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \rho \cdot \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} + \eta_t \cdot \begin{bmatrix} \lambda \\ \lambda \end{bmatrix} + \begin{bmatrix} \sigma_1 \nu_{1t} \\ \sigma_2 \nu_{2t} \end{bmatrix},$$

where η, ν_1, and ν_2 have zero means and unit variances, and are uncorrelated with each other and uncorrelated across time. Parameters of the covariance structure are ρ, λ, σ_1, and σ_2. Econometrician Z claims that to get efficient estimates, it is sufficient to do FGLS separately for each equation. Do you agree? Why?

Question 6 (30 minutes) You believe that the probability that the price of a firm's stock rises when it announces that it is making a takeover bid for another company is given by a function $F(\alpha + \beta \cdot DER)$, where F is a CDF and DER is the firm's debt to equity ratio. You have a sample of takeover announcements, and observations on the direction of change in the firm's stock price following the announcement: D = 1 if the price rises, D = 0 otherwise. You want to test the null hypothesis that $\beta = 0$. (a) Derive a Lagrange Multiplier test for this hypothesis, assuming F is a logistic function, $F(c) = 1/(1 + e^{-c})$. (b) How would you change your answer if F is an unknown, continuously differentiable CDF?

Question 7 (30 minutes) You are considering running a regression $y_t = x_t\beta + u_t$, where t = 1,...,T indexes households in a survey, y is a dummy variable that is one if the household has purchased a car in the past year and is zero otherwise, and x is a vector of variables: a constant, household income, household size, and the average price of cars in the state in which the household lives. (a) Discuss the appropriateness of each of the Gauss-Markov assumptions for this application. (b) You would like to test whether household size has any impact on car purchases. Discuss the applicability of the standard theory of hypothesis testing in regression models. In calculating a t-statistic for your test, is it best to use the standard error for this coefficient that is printed out by the least squares program, or to do something else? If so, what?

Question 8 (1 hour) You have observations (x_i,s_i,k_i) for a sample $i = 1,...,N$, where i indexes zip codes, s_i is the area of the zip code in square kilometers, k_i is the number of permits to construct a single-family residence issued within the zip code area in a one month period, and x_i is a vector of explanatory variables, such as residential density and distance from city center.

(a) As a base model, you postulate that $\mathbf{E}k_i = \lambda \cdot s_i$, where λ is a positive constant interpreted as the rate of construction per unit of time per square kilometer. Find an estimator for λ, and specify conditions on the population under which your estimator will be consistent asymptotically normal. (You do not need to prove a theorem, and the conditions you state should be as specific to this problem as possible, rather than generic conditions for asymptotic theory.)

(b) One possible data generation process consistent with the postulate $\mathbf{E}k_i = \lambda \cdot s_i$ is that k_i has a Poisson distribution, with $\text{Prob}(k_i = m) = \exp(-\lambda \cdot s_i)\cdot(\lambda \cdot s_i)^m/m!$ One implication of this Poisson model is that $\text{Var}(k_i) = \lambda \cdot s_i$, so that the variance equals the expectation. Derive an overidentifying restrictions test for the Poisson assumption, and simplify it as much as you can.

(c) Starting from the base model, find a test for dependence on x_i of the conditional expectation of k_i given x_i. Assume as an alternative to no dependence that $\mathbf{E}(k_i|x_i) = \lambda \cdot s_i \cdot \exp(-x_i\beta)$

QUESTION 9 (45 minutes) Assume you have constructed the following aggregate econometric model of the U.S. economy:

$$c = \gamma_{11}y + \beta_{11} + \beta_{12}c_{-1} + \varepsilon_1$$

$$i = \gamma_{21}y + \beta_{21} + \beta_{22}x + \beta_{23}i_{-1} + \varepsilon_2$$

$$y = c + i + g$$

where c = private consumption expenditure, i = private investment expenditure, y = gross national product, g = government expenditure, and x = exports. The variables c, i, and y are endogenous, and g and x are exogenous. Assume that you have yearly data on consumption and investment from 1950 through 1998, and on the remaining variables from 1951 through 1998.

a. How would you go about estimating the equations in this model if the disturbances are not serially correlated? Describe the assumptions that must hold for the estimation technique you propose to be consistent. Be explicit.

b. Describe how you would test the hypothesis that $\beta_{12} = 0$. Be specific in terms of the estimates you would use, the test statistic, and the distribution this statistic would have under the null hypothesis, including degrees of freedom.

c. In the investment equation, is it legitimate to test the hypothesis that $\gamma_{21} = 0$ by using the t-statistic from an OLS estimation of this equation?

d. How would you test for serial correlation in the investment equation? How would your answer to part (a) change if you find that ε_2 is serially correlated?

QUESTION 10 (30 minutes) Demand functions for labor and capital services per unit of output in a sample of competitive firms are postulated to satisfy

$$L = \beta_{11} + \beta_{12}\ \sqrt{R/W}\quad + \varepsilon_1 \text{ and } K = \beta_{21}\ \sqrt{W/R}\quad + \beta_{22} + \ \ \varepsilon_2$$

where R = service price of capital, W = wage rate, L = labor demand, and K = capital services demand. The theory of the firm suggests that $\beta_{12} = \beta_{21}$. The disturbances ε_1 and ε_2 have different variances, and are correlated, but are independent across observations on different firms.

a How would you estimate the parameters in these equations? Be explicit about how you would set the problem up, and what regressions you would run .

b How would you test the hypothesis that $\beta_{12} = \beta_{21}$? Be explicit.

QUESTION 11 (30 minutes) You are interested in the durability of commercial buildings, as a function of real construction cost per square foot . Your model is $\log(\text{Life}) = \alpha + \beta(\text{Cost}) + \varepsilon$, and you assume that ε is normally distributed. You draw a sample of building permits for buildings constructed in 1920, and from these permits get construction cost per square foot .

a. Suppose that you can determine whether each building is surviving or not in 2001, but can get no other information on Life. How would you estimate α and β?

b. Suppose you find that the building permit records include information on remodeling, so that for each building that is not surviving in 2001, but was remodeled after 1960, you have a date at which the last remodel, if any, was done, so that the building survived at least until that date. How would you estimate α and β in this case?

QUESTION 12 (30 minutes) Ms. Hicks regresses the daily excess return of General Motors stock on the total market daily excess return and a constant, using data from Jan. 1, 1981 through Dec. 31, 1989, and obtains $R^2 = 0.3$. Mr. Keynes runs the same regression, but uses monthly excess returns over the same period, and obtains $R^2 = 0.5$. What conclusions can you draw about which regression is better, and which will yield the more precise estimates of the market beta of General Motors? Justify your answer.

Hints or Solution Outlines:

1. Use y_{2t} as an instrument for y_{1t}, and vice versa.
2. The RE regression is inconsistent when the individual effect is correlated with y.
3. Calculate a LM test, which can be computed without doing any estimation.
4. This is a conventional omitted variables test, and any member of the trinity will work. State the conditions for consistency of these tests in linear models.
5. ρ-difference to put this system of equations in seemingly unrelated regression format. Then, if ρ were known, this system would have the property that OLS applied to each equation is efficient

for the β's. You know that for FGLS in regression with normal disturbances, any consistent estimates of the covariance parameters yield efficient estimates in the second stage. Therefore, you can do OLS on this problem untransformed, estimate ρ, and then do OLS on the ρ-differenced data to get estimators that are equivalent to FGLS.

6. The LM test is just a test for independence in a 2×2 contingency table of d versus DER, and is the same whether the CDF is logistic or any continuously differentiable alternative.

7. The regression must be heteroscedastic. You can do t-tests, but you should use a robust heteroscedasticity-consistent standard error, or do FGLS.

8. The sample average of k_i/s_i is CAN for λ by a law of large numbers and a central limit theorem, provided the observations are independent and k_i/s_i has a finite variance. Apply the GMM theory for a test of over-identifying restrictions, and use an LM test for dependence on x_i.

9. 2SLS works for the first two parts. OLS is legitimate to test the null hypothesis $\gamma_{21} = 0$.

10. Set this SUR problem up as a stacked regression with the parameter transformation $\beta_{12} = \beta_{21} + \alpha$, and use a T-test for $\alpha = 0$ in FGLS estimates of this model.

11. The probability a building survives in 1992 is $\Phi(\alpha+\beta(\text{Cost})-\log(72))$, a probit model. The probability that it survives in 1992 given that it survived to year S for $1960 \le S < T$ is $\Phi(\alpha+\beta(\text{Cost})-\log(72))/\Phi(\alpha+\beta(\text{Cost})-\log(S-1920))$.

12. The aggregation of daily to monthly data is a linear operation on the variables. If the daily data satisfy Gauss-Markov assumptions, then OLS applied to the daily data is BLUE, and OLS applied to the monthly data is another estimator linear in the dependent variables which is therefore less efficient. The values of R^2 are irrelevant.