ECO 82800
Panel Econometrics

Final Exam
21 May 2015, 9:30am – 11:30pm

This exam is a closed-book, closed-notes exam. Calculators without matrix functions are allowed. The exam consists of seven questions, most of them with parts. Points per question are as indicated; parts are weighted equally. The total is 100 points. Budget your time. You may answer the questions in any order, but *label them clearly and keep parts of a question together*.

1. (24 points)

a. Define the model that is the basis for the Hausman-Taylor estimator.

b. Describe the instrument matrix that Amemiya and MaCurdy proposed as a substitute for the Hausman-Taylor instrument matrix.

c. Now, let the panel dataset be unbalanced: $i = 1, \dots, N$ and $t = 1, \dots, T_i$.

c.1 Describe a complicating issue with regard to the Amemiya-MaCurdy instrument matrix.

c.2 What solution do you suggest?

2. (12 points)

Consider the following panel data model:

$$y = Y\gamma + X\beta + u$$

where y is an $(NT \times 1)$ dependent variable, Y is an $(NT \times g)$ matrix of endogenous variables correlated with μ and ν, X is an $(NT \times k)$ matrix of exogenous explanatory variables that includes a column of ones, $u = Z_\mu \mu + \nu$, $Z_\mu = I_N \otimes \iota_T$, $\mu_i \sim iid(0, \sigma_\mu^2)$, $\nu_{it} \sim iid(0, \sigma_\nu^2)$. Let A be a suitable matrix of instrumental variables. Formulate a type of 2SLS estimator that exploits the error components structure of the disturbance term.

3. (24 points)

Consider the model $y_{it} = \delta y_{i,t-1} + \mu_i + \nu_{it}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$. Presume that y_{i0} is observed for all i. Assume that $\mu_i \sim iid(0, \sigma_\mu^2)$ and $\nu_{it} \sim iid(0, \sigma_\nu^2)$. Define $\Delta y_{it} = y_{it} - y_{i,t-1}$. Let A_i be the instrument matrix for observation i (stacking over all applicable t) that is constructed in preparation of GMM estimation. Let the matrix A stack these A_i matrices for all i.

a. Write down the Arellano-Bond instrument matrix A_i. What dimensions does it have?

b. Motivate why this instrument matrix is appropriate.

c. Is this instrument matrix appropriate under all circumstances?

d. An initial Arellano-Bond estimator is given by

$$\hat{\delta}_{(1)} = \left(\Delta y'_{(-1)} A (A'(I_N \otimes G)A)^{-1} A' \Delta y_{-1}\right)^{-1} \Delta y'_{(-1)} A (A'(I_N \otimes G)A)^{-1} A' \Delta \mathrm{y}$$

Derive this matrix G. What dimensions does it have?

e. Suppose $N = 20$ and $T = 100$.

e.1 What problem do you run into when computing $\hat{\delta}_{(1)}$?

e.2 What solution do you suggest for this problem?

f. Arellano and Bond have suggested a two-step estimator $\hat{\delta}_{(2)}$ that builds on information gained with the aid of the first-step estimator $\hat{\delta}_{(1)}$. Explain why this two-step estimator might be preferred in some circumstances.

4. (24 points)

In the table below, the dependent variable is UNION, which equals 1 if the individual is covered by collective bargaining agreements and 0 if not. The explanatory variables include years of schooling, log of experience (=age–school–6), various demographic and residential variables, a lagged union status dummy, a trend variable, and industry dummies. The dataset describes 545 individuals over the years 1980-1987. In the table below, the slopes of the industry dummies are omitted.

a. Between the three probit columns, the effect of race (BLACK and HISP) seems to differ substantially. Explain which estimate, if any, is most trustworthy.

b. Comparing the random effects probit and random effects logit columns, the t-statistics are nearly identical but the parameter estimates differ substantially. Explain.

c. The fixed effects logit model appears to use only 1512 observations (person-years), as opposed to 3815 for random effects logit. Explain.

	Standard probit		Standard probit		Random effects probit		Random effects logit		Fixed effects logit	
	Est.	t	Est.	t	Est.	t	Est.	t	Est.	t
Lag(UNION)			1.872	32.98	1.087	10.27	1.877	10.13	0.286	2.06
lnEXPER	0.202	2.43	-0.076	-0.60	-0.025	-0.11	0.010	0.02	0.142	0.21
SCHOOL	0.007	0.49	-0.012	-0.64	-0.019	-0.49	-0.024	-0.33	n.a.	
MARRIED	0.118	2.50	0.125	2.15	0.184	2.02	0.331	2.00	0.171	0.82
BLACK	0.495	7.43	0.379	4.58	0.716	3.83	1.290	3.81	n.a.	
HISPANIC	0.257	3.82	0.152	1.84	0.351	1.98	0.644	1.99	n.a.	
Rural	-0.003	-0.06	-0.025	-0.35	0.000	0.00	0.023	0.10	-0.007	-0.02
DISABILITY	-0.414	-2.15	-0.286	-1.20	-0.313	-1.08	-0.675	-1.22	-0.536	-0.97
North East	0.200	2.74	0.074	0.81	0.240	1.30	0.439	1.31	2.144	2.15
South	-0.014	-0.20	-0.003	-0.04	-0.026	-0.16	-0.074	-0.24	-1.061	-0.98
North Central	0.240	3.41	0.145	1.63	0.234	1.33	0.454	1.42	0.275	0.38
t = (Year-1979)	-0.061	-3.54	0.002	0.10	-0.022	-0.57	-0.050	-0.70	-0.068	-0.62
Intercept	-0.500	-1.91	-0.754	-2.11	-0.861	-1.29	-1.663	-1.37		
σ_μ					1.050		1.880			
σ_ν					1.000		$\pi/\sqrt{3}$			
$\sigma_\mu^2/(\sigma_\mu^2+\sigma_\nu^2)$					0.525		0.518			
log-likelihood	-2236.92		-1346.87		-1308.84		-1306.78		-523.836	
N of observations	4358		3815		3815		3815		1512	
N of groups					545		545		216	

5. (16 points)

For $i = 1, \dots, N$ and $t = 1, \dots, T$, let y_{it} and x_{it} represent time series data for country i over time. x_{it} is a k-dimensional vector. Let us assume that these time series have unit roots. We are interested in estimating the relationship $y = f(x;u) = \alpha + x'\beta + u$ where u represents a disturbance. Comment on the following statements:

a. Estimating the equation $y_{it} = \alpha_i + x_{it}'\beta_i + u_{it}$ for each country separately with OLS yields an estimate $\hat{\beta}_i$ that converges to β_i as $T \to \infty$.

b. Estimating the equation $y_{it} = \alpha_i + x_{it}'\beta_i + u_{it}$ on the pooled dataset of all countries with OLS yields an estimate $\hat{\beta}$ that converges to $\beta = \frac{1}{N}\sum_{i=1}^{N}\beta_i$ as $T \to \infty$.

c. Regardless of whether x and y have unit roots, the equation $y_{it} = \alpha_i + x_{it}'\beta_i + u_{it}$ should never be estimated with panel econometric methods because both intercept and slopes vary with i.

d. If u_{it} contains a unit root, the variables y and x are not cointegrated. However, cointegration is a meaningless concept in the context of panel econometrics since estimators $\hat{\beta}$ are not impacted by it.