



**Strathmore**  
UNIVERSITY

## **STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES**

### **Bachelor of Business Science Financial Economics**

#### **BSE 4211: Panel Data**

#### **Special Examination**

---

Date: **5th** May 2022

Time: 2 Hours

---

#### **Read the Following Instructions Carefully**

1. This examination consists of FIVE questions.
2. Question One carries 30 marks.
3. All other questions carry 20 marks EACH
4. Answer Question ONE (COMPULSORY) and any other TWO questions.
5. This is a CLOSED BOOK exam.

**Question 1**

1. (a) Which of the following is a disadvantage of the random effects approach to estimating a panel model? (2 marks)
- a) The approach may not be valid if the composite error term is correlated with one or more of the explanatory variables
  - b) The number of parameters to estimate may be large, resulting in a loss of degrees of freedom
  - c) The random effects approach can only capture cross-sectional heterogeneity and not temporal variation in the dependent variable.
  - d) All of (a) to (c) are potential disadvantages of the random effects approach.
- (b) The "within transform" involves (2 marks)
- a) Taking the average values of the variables
  - b) Subtracting the mean of each entity away from each observation on that entity
  - c) Estimating a panel data model using least squares dummy variables
  - d) Using both time dummies and cross-sectional dummies in a fixed effects panel model
- (c) Treating each data point in panel data as an individual observation and not accounting for the fact that individuals are observed over many time periods is referred to as estimating a (-----) model. (2 marks)
- A)static time-series
  - B)heteroskedastic consistent
  - C)pooled cross-section
  - D)fixed-effects
- (d) Fixed-effects models are more appropriate than random-effects models when (2 marks)
- A) the error terms are autocorrelated.
  - B)the independent variables are constant over time.
  - C)the time invariant component of the error term is correlated with any of the independent variables.
  - D)the error term is not constant.
- (e) In order to determine whether to use a fixed effects or random effects model, a researcher conducts a Hausman test. Which of the following statements is false? (2 marks)
- a) For random effects models, the use of OLS would result in consistent but inefficient parameter estimation

- b) If the Hausman test is not satisfied, the random effects model is more appropriate.
- c) Random effects estimation involves the construction of "quasi-demeaned" data
- d) Random effects estimation will not be appropriate if the composite error term is correlated with one or more of the explanatory variables in the model
- (f) True or False: Give a brief discussion supporting your answer for full marks - It is much easier to obtain instruments with panel data than with cross-sectional data (2 marks)
- (g) Consider a panel regression of unemployment rates for the G7 countries (United States, Canada, France, Germany, Italy, United Kingdom, Japan) on a set of explanatory variables for the time period 1980-2000 (annual data). If you included cross-sectional effects and time effects with an intercept, you would need to specify the following number of binary variables: (2 marks)
- 21
  - 6
  - 28
  - 26
- (h) Which of the following is a difference between a fixed effects estimator and a first-difference estimator? (2 marks)
- The fixed effects estimators are always larger than the first difference estimators in a two-period panel data analysis.
  - The fixed effects estimator is more efficient than the first-difference estimator when the idiosyncratic errors are serially uncorrelated.
  - The first difference estimator is more sensitive to nonnormality and heteroskedasticity.
  - The bias in the first difference estimator depends on the time period ( $T$ ) of analysis while the bias in the fixed effect does not depend on  $T$ .
- (i) Which of the following is a property of dummy variable regression? (2 marks)
- This method is best suited for panel data sets with many cross-sectional observations.
  - The  $R$ -squared obtained from this method is lower than that obtained from regression on demeaned data.
  - The degrees of freedom cannot be computed directly with this method.
  - The major statistics obtained from this method are identical to that obtained from regression on demeaned data.

- (j) How does the random effects model capture cross-sectional heterogeneity? (6 marks)
- (k) Explain how fixed effects models are equivalent to an ordinary least squares regression with dummy variables. (6 marks)

Total for Question 1: 30 marks

## Question 2

2. (a) You use a dataset called JTRAIN to determine the effect of the job training grant on hours of job training per employee. The basic model for the three years (1987-1989) is:

$$\text{hrsemp}_{it} = \beta_0 + \delta_1 d88_t + \delta_2 d89_t + \beta_1 \text{grant}_{it} + \beta_2 \text{grant}_{i,t-1} + \beta_3 \log(\text{employ}_{it}) + a_i + u_{it}$$

The dependent variable is annual hours of training per employee, at the firm level.

grant = 1 if firm received grant for training

d89 = 1 if year = 1989

d88 = 1 if year = 1988

employ # employees at plant

The equation is estimated using fixed effects.

Fixed-effects (within) regression		Number of obs	=	390	
Group variable: fcode		Number of groups	=	135	
R-sq: within	= 0.4909	Obs per group: min	=	1	
between	= 0.0514	avg	=	2.9	
overall	= 0.2206	max	=	3	
corr(u_i, Xb) = -0.0270		F(5,250)	=	48.21	
		Prob > F	=	0.0000	
hrsemp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
d88	-1.098678	1.983157	-0.55	0.580	-5.004503 2.807148
d89	4.090049	2.481125	1.65	0.101	-.7965232 8.97662
grant	34.22818	2.858438	11.97	0.000	28.59849 39.85787
grant_l	.5040804	4.127325	0.12	0.903	-7.62468 8.632841
lemploy	-.1762661	4.287935	-0.04	0.967	-8.621347 8.268815
_cons	9.324956	14.92757	0.62	0.533	-20.07487 38.72478

- How many firms are used in the FE estimation? (1 mark)
- How many total observations would be used if each firm had data on all variables (in particular, hrsemp) for all three years? (1 mark)
- Interpret the coefficient on grant and comment on its significance. (2 marks)
- Interpret the coefficient on the year dummy variables. (2 marks)
- Is it surprising that grant<sub>-1</sub> is insignificant? Explain. (2 marks)

vi. Do larger firms provide their employees with more or less training, on average?

How big are the differences? *(4 marks)*

- (b) The Im-Pesaran-Shin Test (IPS test) is an example of an alternative to the Levin Lin Chu test above. Briefly describe the test, making sure to highlight how it compares to the LLC test *(8 marks)*

Total for Question 2: 20 marks

**Question 3**

3. (a) The file MATHPNL contains panel data on school districts in Michigan for the years 1992 through 1998. The response variable of interest in this question is  $math4$ , the percentage of fourth graders in a district receiving a passing score on a standardized math test.

$$\begin{aligned} math_{it} = & \delta_1 y94_t + \dots + \delta_5 y98_t + \gamma_1 \log(\text{rexpp}_{it}) + \gamma_2 \log(\text{rexpp}_{i,t-1}) \\ & + \psi_1 \log(\text{enrol}_{it}) + \psi_2 \text{lunch}_{it} + a_i + u_{it}, \end{aligned}$$

The key explanatory variable is  $rexpp$ , which is real expenditures per pupil in the district. The amounts are in 1997 dollars. This spending variable will appear in logarithmic form.

The  $lunch$  variable is the percent of students in the district eligible for free or reduced-price lunches, which is determined by poverty status.  $enrol$  is the school enrolment rate in the district

Additionally,

$$y94 = 1 \text{ if year } == 1994$$

$$y95 = 1 \text{ if year } == 1995$$

$$y96 = 1 \text{ if year } == 1996$$

and so on ...

You first estimate the model by pooled OLS and report the usual standard errors in parentheses. The results are reported below:

$$\begin{aligned} \hat{math4} = & -31.66 + 6.38 y94 + 18.65 y95 + 18.03 y96 + 15.34 y97 + 30.40 y98 \\ & (10.30) \quad (.74) \quad (.79) \quad (.77) \quad (.78) \quad (.78) \\ & + .534 \log(\text{rexpp}) + 9.05 \log(\text{rexpp}_{-1}) + .593 \log(\text{enrol}) - .407 \text{lunch} \\ & (2.428) \quad (2.31) \quad (.205) \quad (.014) \end{aligned}$$

$$N = 550, T = 6, R^2 = .505$$

- i. Discuss the estimated size, sign and statistical significance of the effects of the spending variables (3 marks)
- ii. Is the sign of the  $lunch_{it}$  coefficient what you expected? Interpret the magnitude and significance of the coefficient. Would you say that the district poverty rate has a big effect on test pass rates? (4 marks)

- iii. Now, you estimate the equation by fixed effects. Is the lagged spending variable still the same size and is it significant? What has changed in this model? (3 marks)

$$\begin{aligned} \text{math4} = & 6.18 \text{ y94} + 18.09 \text{ y95} + 17.94 \text{ y96} + 15.19 \text{ y97} + 29.88 \text{ y98} \\ & (.56) \quad (.69) \quad (.76) \quad (.80) \quad (.84) \\ & - .411 \log(\text{rexpp}) + 7.00 \log(\text{rexpp}_{-1}) + .245 \log(\text{enrol}) + .062 \text{ lunch} \\ & (2.458) \quad (2.37) \quad (1.100) \quad (.051) \end{aligned}$$

$$N = 550, T = 6, R^2 = .603$$

- (b) The Arrelano-Bond (1991) and Arrelano-Bover (1995)/Blundell-Bond (1998) dynamic panel estimators are increasingly popular. Both are general estimators designed for situations with “small T, large N” panels, dynamic relationships, fixed individual effects and heteroskedasticity and autocorrelation within individuals.

Provide a detailed discussion of The Arrelano & Bond Estimator and the Arrelano & Bover/Blundell & Bond Estimator making sure to highlight the key differences between the two estimators? (10 marks)

Total for Question 3: 20 marks

## Question 4

4. (a) Explain what cross-sectional dependence is (**2 marks**), 3 possible causes (**3 marks**), and how it is tested for in Panel Data (**2 marks**) (7 marks)
- (b) In the traffic dataset, you have 1982–1988 state-level data for 48 U.S. states on traffic fatality rates (deaths per 100,000). You seek to model the highway fatality rates (*fatal*) as a function of 4 independent variables: *beertax*, the tax on a case of beer, *spircons*, a measure of spirits consumption and two economic factors: the state unemployment rate (*unrate*) and state per capita personal income, \$000 (*perinc*)
- Answer the following questions after carrying out Panel Data Estimation in STATA.
- i. You estimate the above model using a Fixed Effects Estimation. Is the model jointly significant? How did you infer this? (2 marks)

```

Fixed-effects (within) regression              Number of obs   =       336
Group variable: state                        Number of groups =        48

R-sq:  within = 0.3526                       Obs per group:  min =         7
        between = 0.1146                      avg =         7.0
        overall = 0.0863                      max =         7

corr(u_i, Xb) = -0.8804                      F(4,284)        =       38.68
                                                Prob > F        =       0.0000

```

fatal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
beertax	-.4840728	.1625106	-2.98	0.003	-.8039508	-.1641948
spircons	.8169651	.0792118	10.31	0.000	.6610484	.9728818
unrate	-.0290499	.0090274	-3.22	0.001	-.0468191	-.0112808
perinc	.0001047	.0000206	5.08	0.000	.0000642	.0001453
_cons	-.3837827	.420178	-0.91	0.362	-1.210841	.4432756
sigma_u	1.1181912					
sigma_e	.15678966					
rho	.98071823	(fraction of variance due to u_i)				

```

F test that all u_i=0:      F(47, 284) =      59.77      Prob > F = 0.0000

```

- ii. Given the above results how do we interpret the coefficients on the state unemployment rate (*unrate*) and state per capita personal income, \$000 (*perinc*). Focus on the size, sign and statistical significance of the two coefficients (4 marks)
- iii. Are there significant unobserved effects from the model? How did you infer the result? (2 marks)
- iv. You generate an interaction term between: *spircons*: a measure of spirits consumption and the state unemployment rate (*unrate*). You then run the same

model as above, NOW including the interaction variable (a total of 5 independent variables now). Results are below

How would you NOW interpret the size and sign effect of *unrate*, the state unemployment rate on the fatality rates? (5 marks)

```
Fixed-effects (within) regression          Number of obs   =   336
Group variable: state                    Number of groups =    48

R-sq:  within = 0.3546                   Obs per group:  min =    7
        between = 0.1169                                     avg =   7.0
        overall = 0.0878                                     max =    7

corr(u_i, Xb) = -0.8769                   F(5,283)        =   31.10
                                           Prob > F         =   0.0000
```

fatal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
beertax	-.4553375	.1654303	-2.75	0.006	-.7809675	-.1297076
spircons	.7353677	.1179105	6.24	0.000	.5032749	.9674605
unrate	-.0413484	.0159612	-2.59	0.010	-.0727662	-.0099306
perinc	.0001073	.0000208	5.16	0.000	.0000663	.0001482
interaction	.0080411	.0086055	0.93	0.351	-.0088978	.0249801
_cons	-.3006735	.4295803	-0.70	0.485	-1.146252	.5449047
sigma_u	1.1023595					
sigma_e	.15682469					
rho	.98016279	(fraction of variance due to u_i)				

```
F test that all u_i=0:      F(47, 283) =   59.59      Prob > F = 0.0000
```

Total for Question 4: 20 marks

### Question 5

5. (a) Panel unit root testing emerged from time series unit root testing. The major difference to time series testing of unit roots is that we have to consider asymptotic behavior of the time-series dimension  $T$  and the cross-sectional dimension  $N$ . With the help of relevant formulae and hypothesis, provide a detailed discussion of the Fisher Type Test (attributed to Maddala- Wu (1991) and Choi (2001)) (6 marks)

(b) Diaz-Fernandez, et al (2017) authored a paper title “*HRM practices and Innovation Performance: A Panel-Data Approach*”. The purpose of this paper is to study the relationship between human resource management (HRM) practices and innovation performance in Spanish manufacturing firms. The paper focuses on the number of existing patents, analyzing the extent to which this variable is favored by HRM practices.

The variables and results from 2 model estimations are presented below. Answer the questions that follow:

Variable	Definition	Model 1		Model 2	
		Coefficient	SE	Coefficient	SE
<i>Control variables</i>					
Size	Logarithm of the number of employees				
<i>Intensity of Technology</i>					
OECD1	Low-tech industry (reference category)	0.004	0.019	0.004	0.025
OECD2	1 if medium-tech industry; 0 for any other category	0.008	0.079	0.018	0.081
OECD3	1 if high-tech industry; 0 for any other category	0.036	0.069	0.049	0.074
<i>Main effects</i>					
Innovation performance	Number of patents	0.047	0.013	0.048	0.139
		0.001*	0.002	0.016*	0.041
		0.002	0.001	0.002	0.004
		0.003	0.019	0.002	0.039
		0.002*	0.008	0.013*	0.041
<i>Interaction terms</i>					
				0.021**	0.003
				0.009	0.014
				0.022*	0.047
<i>Model fit statistics</i>					
		Wald $\chi^2(8) = 1.45$		Wald $\chi^2(11) = 2.04$	
		Prob $\chi^2 = 0.003$		Prob $\chi^2 = 0.008$	
		$\chi^2(20) = 17.418; p = 0.625$		$\chi^2(20) = 35.925; p = 0.105$	
<i>Hansen test</i>					
		1.31 $p = 0.189$		1.31 $p = 0.187$	
		0.415 $p = 0.678$		0.431 $p = 0.666$	
		6,887		6,887	
		1,363		1,363	
		28		31	
<b>Notes:</b> * $p < 0.01$ ; ** $p < 0.05$ and *** $p < 0.001$					

- Using the results from Model 2, interpret the effect of Employment Security on Innovation performance (2 marks)
- Using the results from Model 2, how is Investment in training in new technologies in the firm related to innovation performance? (2 marks)
- Using results from Model 1, interpret the results on “OECD2” and “OECD3”(2 marks)

- iv. What do the model fit statistics reported in the table tell us? Interpret accordingly (2 marks)
- v. Are the instruments in Model 2 valid? (2 marks)
- vi. Explain the results from the AR(1) and AR(2) tests? Why are they carried out? Do they support the model? (4 marks)

Total for Question 5: 20 marks