



**STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES**  
**BSE 4211: PANEL DATA ANALYSIS**  
**END OF SEMESTER EXAMINATION FOR BACHELOR OF BUSINESS**  
**SCIENCE IN FINANCIAL ECONOMICS**

Date: 15<sup>th</sup> December, 2022

Time: 2 hours

---

**Instructions**

1. This examination consists of **Five** questions.
2. Answer **Question One**(Compulsory) and **any other two** questions.

**Question 1**

Consider the model  $y_{it} = \alpha_{it} + X'_{it}\beta_{it} + \epsilon_{it}$  and a minimum working example with three individuals,  $i = 1, 2, 3$ , three years,  $t = 2019, 2020, 2021$ , and variable weight with hypothetical observations.

- (a) If empirical data is collected to estimate the above model, under what circumstances would the data set be considered balanced or unbalanced {2 marks}
- (b) Using the minimum working example, state and explain the two ways of representing panel data sets {4 marks}
- (c) Using both the model and the minimum working example, state the various sources of variations in panel data modelling {2 marks}
- (d) While considering the model, state and explain the five flavours of panel data models that can be estimated {8 marks}
- (e) Using the popular flavour in 1(d) above and clearly stated equations, distinguish between a one and a two way error component model {4 marks}
- (f) Using the model, state the various circumstances that dictate the use of panel data models {6 marks}
- (g) Using the model and the minimum working example, state the advantages that a researcher using panel data models has over and above those using time series or cross-sectional data sets {4 marks}

**[30 marks]**

The information below relates to Vella and Verbeek (1998) estimation of an individual wage equation using the Youth Sample of the National Longitudinal Survey held in the USA. The sample comprises of 545 full-time working males who completed their schooling by 1980 and then followed over the period 1980-1987. The data set contains observations of the young males wages, (in logs), years of schooling, experience, union membership, marriage status and ethnicity. The results arising from this estimation are detailed in Table 1

Table 1: Individual wage equation for young males

Variables	Pooled OLS	Fixed Effects	Random Effects	Between Estimator
Years of schooling	0.0994*** (0.00468)		0.101*** (0.00891)	0.0946*** (0.0109)
Experience	0.0892*** (0.0101)	0.117*** (0.00842)	0.112*** (0.00826)	-0.0504 (0.0503)
Experience Squared	-0.00285*** (0.000707)	-0.00430*** (0.000605)	-0.00407*** (0.000592)	0.00512 (0.00321)
Unionised (yes)	0.180*** (0.0171)	0.0821*** (0.0193)	0.107*** (0.0178)	0.271*** (0.0466)
Married (yes)	0.108*** (0.0157)	0.0453** (0.0183)	0.0628*** (0.0168)	0.144*** (0.0412)
Black	-0.144*** (0.0236)		-0.144*** (0.0476)	-0.139*** (0.0489)
Hispanic	0.0157 (0.0208)		0.0202 (0.0426)	0.00478 (0.0427)
Constant	-0.0347 (0.0646)	1.065*** (0.0267)	-0.107 (0.111)	0.492** (0.221)
Observations	4,360	4,360	4,360	4,360
R-squared	0.187	0.178		0.219
Number of nr		545	545	545

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Use this information to answer question two and three

## Question 2

- With the help of clearly stated equations, state the various models that were used to estimate the relationships in Table 1 {3 marks}
- State and explain the assumptions that were made about the (i) individual uniqueness and (ii) individual uniqueness and the predictor variables in estimating the first three columns in Table 1 {9 marks}
- State the transformation used to purge out individual uniqueness and the formula used to obtain  $\hat{\beta}_{FE}$  in column two of Table 1 {3 marks}

- (d) Explain why the variables years of schooling and ethnicity are dropped in the specification in column 2 of Table 1 {3 marks}
- (e) Using an elaborate equation show that the findings in column 3 of Table 1 are a weighted average of column 2 and column 4 results {2 marks}

[20 marks]

### Question 3

- (a) Figure 1 shows a section of the stata output from the results of the fixed effects model. Using these results and a clearly stated test, explain how researchers discriminate between  $\hat{\beta}_{FE}$  and  $\hat{\beta}_{POLS}$  {4 marks}

Figure 1: Fixed effects model detailed results

```

Group variable: nr                Number of groups =      545
R-squared:                        Obs per group:
  Within = 0.1780                  min =      8
  Between = 0.0005                 avg =     8.0
  Overall = 0.0638                 max =      8
corr(u_i, Xb) = -0.1139           F(4, 3811) =     206.38
                                   Prob > F =     0.0000

```

wage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
school	0 (omitted)					
exper	.1168467	.0084197	13.88	0.000	.1003392	.1333542
exper2	-.0043009	.0006053	-7.11	0.000	-.0054876	-.0031142
union						
yes	.0820871	.0192907	4.26	0.000	.044266	.1199083
married						
yes	.0453033	.0183097	2.47	0.013	.0094056	.081201
ethn						
black	0 (omitted)					
hisp	0 (omitted)					
_cons	1.06488	.0266607	39.94	0.000	1.012609	1.11715
sigma_u	.4000539					
sigma_e	.35125535					
rho	.5646785	(fraction of variance due to u_i)				

```

F test that all u_i=0: F(544, 3811) = 9.71                Prob > F = 0.0000

```

- (b) What alternative procedures and estimations would yield detailed results of the test in 3(a) above {2 marks}
- (c) Figure 2 shows the output from a test used to discriminate between the  $\hat{\beta}_{POLS}$  and  $\hat{\beta}_{RE}$ . Using this output and a clearly stated test, explain how researchers discriminate between  $\hat{\beta}_{POLS}$  and  $\hat{\beta}_{RE}$  {4 marks}

**Figure 2:** A test discriminating between random effects and pooled OLS

```

Estimated results:

```

	Var	SD = sqrt(Var)
wage	.2836728	.5326094
e	.1233803	.3512553
u	.1053439	.3245673

```

Test: Var(u) = 0
      chibar2(01) = 3216.73
      Prob > chibar2 = 0.0000

```

- (d) Figure 3 shows the output from a test used to discriminate between the  $\hat{\beta}_{FE}$  and  $\hat{\beta}_{RE}$ . Using this output and a clearly stated test, explain how researchers discriminate between  $\hat{\beta}_{FE}$  and  $\hat{\beta}_{RE}$  {4 marks}

**Figure 3:** A test discriminating between random effects and pooled OLS

```

Estimated results:

```

	Var	SD = sqrt(Var)
wage	.2836728	.5326094
e	.1233803	.3512553
u	.1053439	.3245673

```

Test: Var(u) = 0
      chibar2(01) = 3216.73
      Prob > chibar2 = 0.0000

```

- (e) Figure 4 shows the output from a test used to test for heteroscedasticity. State the model that must precede this test and interpret the tests findings. {4 marks}

**Figure 4:** A test discriminating between random effects and pooled OLS

```

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (545) = 2.2e+05
Prob>chi2 = 0.0000

```

- (f) Given the tests in 3(a) to 3(d) and your conclusions use the appropriate model to interpret the effect of experience on the wages earned by young males {2 marks}

[20 marks]

#### Question 4

Roodman 2006 demonstrated how to do `xtabond2` using the Arellano and Bond (1991) firm level employment data set. The data set contains observations on the level of employment,  $n$ , the firms wage,  $w$ , gross capital,  $k$ , and the aggregate output of the sector that the firm belongs to,  $ys$ , based on a sample of 140 UK firms surveyed annually between 1976 and 1984. Variable names ending in L1 and L2 in the data set indicate lagged copies of the various variables. the findings from this demonstration are as shown in Table 2 Use this information to answer the following questions:

- (a) Using clear logical statements, state the kind of panel model connecting firm employment and the firm characteristics {2 marks}
- (b) Using elaborate equations, explain the pitfalls of estimating the model stated in 4(a) using static panel data models {3 marks}
- (c) With the help of equations, show that the pitfall in 4(b) is a micro panels problem {3 marks}
- (d) Using firm employment data set , state any four instruments that can be used to address the problem in 4(b) {2 marks}
- (e) With the help of clear explanation and the results in Table 2, state the bounds of the coefficients of lagged employment {4 marks}
- (f) Considering your answers in 4(e) above, state whether the coefficients of the lagged employment meet the expected theoretical expectations {2 marks}
- (g) In view of your answers in 4(a) through 4(f), what is the effect of previous level of employment on current employment? {4 marks}

[20 marks]

#### Question 5

- (a) Panel data sets are either Macro or Micro, while clearly stating the criteria used distinguish between macro and micro panels {10 marks}
- (b) Davidson et al.(1978) estimated a panel ARDL model of standard consumption function for a sample of OECD countries. The estimation involved regressing consumption,  $c_{it}$  against income  $y_{it}$  and inflation  $\pi_{it}$ . The results from this estimation are as shown in Figure 5.

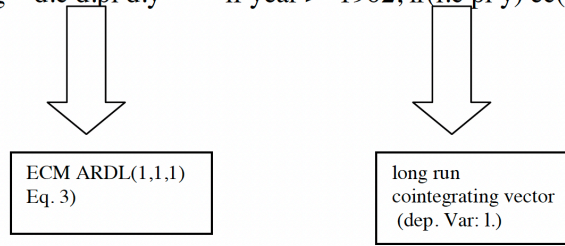
Table 2: The dynamics of firm level employment

Variables	Fixed Effects	Two Step	POLS
	Effects	GMM	
nL1	0.733*** (0.0393)	0.818*** (0.0860)	1.045*** (0.0337)
nL2	-0.139*** (0.0400)	-0.112** (0.0502)	-0.0765** (0.0328)
w	-0.560*** (0.0570)	-0.682*** (0.143)	-0.524*** (0.0488)
wL1	0.315*** (0.0610)	0.656*** (0.202)	0.477*** (0.0487)
k	0.388*** (0.0310)	0.353*** (0.122)	0.343*** (0.0255)
kL1	-0.0805** (0.0385)	-0.154 (0.0863)	-0.202*** (0.0401)
kL2	-0.0278 (0.0328)	-0.0305 (0.0321)	-0.116*** (0.0285)
ys	0.469*** (0.123)	0.651*** (0.190)	0.433*** (0.123)
ysL1	-0.629*** (0.158)	-0.916*** (0.264)	-0.768*** (0.166)
ysL2	0.0580 (0.135)	0.279 (0.186)	0.312*** (0.111)
yr1978	0.0119 (0.0262)	-0.0113 (0.00920)	
yr1979	0.0166 (0.0247)		0.0159 (0.0144)
yr1980	0.0231 (0.0250)	0.0150 (0.0140)	0.0220 (0.0167)
yr1981	-0.0135 (0.0260)	-0.0250 (0.0261)	-0.0222 (0.0204)
yr1982	-0.0225 (0.0230)	-0.0464* (0.0279)	-0.0150 (0.0207)
yr1983	-0.0161 (0.0209)	-0.0422 (0.0332)	0.00739 (0.0204)
yr1984		-0.0352 (0.0352)	0.0154 (0.0230)
Constant	1.780*** (0.501)		0.275 (0.351)
Observations	751	611	751
R-squared	0.797		0.994
Number of id	140	140	

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Figure 5:** Davidson et al (1978) standard consumption function

- xtpmg d.c d.pi d.y if year >=1962, lr(l.c pi y) ec(ec) replace pmg



```
Iteration 0: log likelihood = 2270.3017 (not concave)
Iteration 1: log likelihood = 2319.1636
Iteration 2: log likelihood = 2322.938
Iteration 3: log likelihood = 2326.7589
Iteration 4: log likelihood = 2327.0742
Iteration 5: log likelihood = 2327.0749
Iteration 6: log likelihood = 2327.0749
```

Pooled Mean Group Regression  
(Estimate results saved as pmg)

				Log Likelihood = 2327.075			
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ec	pi	-.4658438	.0567332	-8.21	0.000	-.5770388	-.3546488
	y	.9044336	.0086815	104.18	0.000	.8874182	.9214491
	-----						
SR	ec	-.1998761	.0321683	-6.21	0.000	-.2629247	-.1368275
	pi						
	D1.	-.0182588	.0277523	-0.66	0.511	-.0726522	.0361347
	y						
	D1.	.3269355	.0574236	5.69	0.000	.2143873	.4394838
	_cons	.1544507	.0216942	7.12	0.000	.1119307	.1969706

Use these results to answer the following questions:

- Identify the key parameters that are important for checking the correctness of the specification of this model and interpret them {4 marks}
- State any other two alternative estimators that Davidson et al (1978) could have used to estimate this relationship {3 marks}
- State the conditions under which each of the models would have been appropriate to estimate the consumption function {3 marks}

[20 marks]

Be More be Strathmore! 😊