SC968-AS-MI

UNIVERSITY OF ESSEX

Postgraduate Examination 2011-2012

PANEL DATA METHODS

Time allowed: **TWO** hours

Candidates are permitted to bring into the examination room:
Hand-held, non-programmable calculators (containing no textual information).
Paper dictionaries (with the title in English).

Candidates must answer **THREE** questions in total: **one** question from **Section A**, **one** question from **Section B** and **Question 5** in **Section C.**

The paper consists of **five** questions in total: two questions in Sections A, two questions in Section B, and one question in Section C.

The questions are **NOT** of equal weight. Questions in Sections A and B are each worth 35%, and Question 5 in Section C is worth 30%.

The percentages shown in brackets indicate the number of marks allocated to each part of a question.

Please do not leave your seat unless you are given permission by an invigilator.
Do not communicate in any way with any other candidate in the examination room.
Do not open the question paper until told to do so.
All answers must be written in the answer book(s) provided.
All rough work must be written in the answer book(s) provided. A line should be drawn through any rough work to indicate to the examiner that it is not part of the work to be marked.
At the end of the examination, remain seated until your answer book(s) have been collected and you have been told you may leave.

SECTION A

Candidates must answer one question from Section A. Questions in Section A are each worth 35%.

1. Answer *all* parts of this question.

   (a) Empirical researchers in the social sciences make use of several different types of data. Describe the difference between aggregate-level data and individual-level data, and give an example of how both types could appear in a single data set. [4%]

   (b) Explain the difference between the following types of individual-level data sets, including a discussion of the advantages and problems associated with each type: [6%]

   (i) cross-sectional data;

   (ii) cohort studies;

   (iii)panel data.

   (c) Explain, with illustrative examples, what is meant by storing panel data in "long" and "wide" format. [5%]

   (d) Explain, using illustrative examples:

   (i) The difference between a transition matrix and a transition probability matrix, and the usefulness of each type. [4%]

   (ii) How a transition matrix may be converted to a transition probability matrix. [4%]

   (e) The following transition matrices are based on samples of young men (aged 20-29) in two countries: the Netherlands and Italy. They show the percentages who live either with their parents, or independently (ie, separately from their parents). Also, in the second year, some of the sample could not be traced; these numbers are also shown. Comment on these transition matrices, focusing in particular on what they tell you about differences between the two countries. [12%]

Question 1 continues ……

Question 1 continued ……

The Netherlands		Year 2			
		Lives with parents	**Lives independently**	**Missing from sample**	**Total**
Year 1	**Lives with parents**	350 59.8%	149 25.5%	86 14.7%	585 100.0%
	Lives independently	34 1.5%	1753 77.3%	481 21.2%	2268 100.0%
	Total	384 13.4%	1902 66.7%	567 19.9%	2853 100.0%

Italy		Year 2			
		Lives with parents	**Lives independently**	**Missing from sample**	**Total**
Year 1	**Lives with parents**	8987 86.6%	412 4.0%	983 9.5%	10382 100.0%
	Lives independently	32 1.7%	1439 77.3%	390 21.0%	1861 100.0%
	Total	9019 73.7%	1851 15.1%	1373 11.2%	12243 100.0%

2. Answer *all* parts of this question.

(a) In the context of panel data, what is meant by "within-group variation" and "between-group variation"? [6%]

(b) Describe the fixed-effects and between-group estimators, explaining how they are related to the Ordinary Least Squares estimator. Why would you expect estimates from the two models to differ, even if they are run on the same samples, using the same variables? [7%]

(c) Why can't the fixed-effects model estimate coefficients on variables such as (for example) ethnicity? [4%]

(d) What is the random effects estimator? In what sense is it "efficient" in comparison to the fixed-effects and between-group estimators? [5%]

(e) Describe the Hausman test, in the context of deciding between the fixed effects and random effects models. Why, even if the Hausman test rejects random effects, might the researcher still decide to implement and report the random effects model? [5%]

(f) Your team is working on a project analysing the determinants of psychological wellbeing. As a dependent variable, you are considering a continuous Likert scale ranging from 0 to 36, with higher scores indicating lower levels of wellbeing. Your colleague has produced the following preliminary estimates. Comment on the estimates, and suggest how he might improve the specification. [8%]

	Fixed effects		Between group		Random effects	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Female	-	-	1.412	0.00	1.496	0.00
Age	-0.110	0.00	0.124	0.00	0.118	0.00
Age squared	-0.010	0.00	-0.002	0.00	-0.001	0.00
Has a partner	-0.282	0.02	0.012	0.95	-0.193	0.05
Weekly hours of work	-0.005	0.03	-0.023	0.00	-0.009	0.00
Health limits activity	1.657	0.00	5.14	0.00	2.128.71	0.00
Constant	8.746	0.00	8.36	0.00	8.155	0.00
No. of observations	25975		25975		25975	

END OF SECTION A

SECTION B

Candidates must answer ONE question from Section B. Questions in Section B are each worth 35%.

3. Answer *all* parts of this question.

In this question we will study the relationship between financial worries and partnership breakdown. We use a sample of 5554 respondents in partnerships followed up over a 30 year period. Our aim is to study how failure of a relationship depends on financial worries, controlling for age at start of partnership and ethnicity. *Table 3.1* shows the coding and frequencies for the relevant variables and the result of the Cox proportional hazard model is shown in *Table 3.2*.

Table 3.1 Description of variables

Variable	Variable Label	Values	Value Labels	%
Ethnicity	Non-white	0	No	92
		1	Yes	8
Age at start	agegrp	0	16-25	15
		1	26-35	45
		2	36-45	30
		3	46+	10
Financial worries	worries	0	No	84
		1	Yes	16
Partnership failure	partner	0	Survived/censored	63
		1	Failed	37

Table 3.2 The Cox proportional hazard model results

_t	HR	SE	Z	p>|z|	95% CI	
Non-White	1.37	0.31	1.02	0.31	0.75	2.50
Age 26-35	0.86	0.21	0.71	0.48	0.57	1.31
Age 36-45	1.37	0.31	1.02	0.31	0.75	2.50
Age 46+	0.92	0.20	0.43	0.66	0.63	1.35
Financial worries	1.10	0.03	3.20	0.001	1.04	1.16

Question 3 continues .....

Question 3 continued

(a) Describe in words what a Cox proportional hazard model is and how it is different from a logistic regression model. [5%]

(b) Give the equation for a Cox proportional hazard model and describe all the components in the equation. Explain the relationship between the hazard ratios (HR) given in *Table 3.2* and the equation. [8%]

(c) Based on the results in *Table 3.2*, is there evidence of an association between financial worries and duration of partnerships? Interpret the magnitude of the association in terms of probabilities. [5%]

(d) Copy the following table into your answer book. Using the information in *Table 3.2*, fill in the empty cells with the hazard ratios for the majority ethnic group (white). The joint reference category (HR= 1.00) is for those with no financial worries aged 16-25 at the start of the relationship. [5%]

	Financial worries	
	No	Yes
16-25 years	1.00	
26-35 years		
36-45 years		
46+ years		

(e) How would the numbers for section (d) change if you constructed the table for non-whites instead? [2%]

(f) How could you test if you should fit a stratified Cox model which is stratified on ethnicity? Illustrate with figures how the hazard function might look if: [10%]

(i) the proportional hazards assumption for ethnicity is upheld, and
(ii) the proportional hazards assumption for ethnicity is violated.

4. Answer *all* parts of this question.

(a) Suppose that you want to analyse youth alcohol consumption with repeated measures for each individual. You want to know how drinking behaviour develops over time and how your covariates are related to drinking alcohol. Describe the type of random coefficients model you might fit. Use equations and diagrams as appropriate. [10%]

(b) You administer a questionnaire to young people aged 10-16 containing data on their attitudes to alcohol and their own and their friends' drinking behaviour. The data are described in table 4.1 below. [12%]

Table 4.1 Description of variables

Variable	Variable Label	Values
Gender	sex	Male/female
Time	time	Year 1 - 10
Alcohol makes me happy	happy	Yes/no
Alcohol gives me confidence	confidence	Yes/no
My friends drink	friends	Yes/no
Amount usually drink per occasion	quantity	0-10 units
Times per week drink	frequency	0-7 days

Below is a table of results from a series of Stata runs analysing drinking (quantity) over time (time) and the possible influence of attitudes (happy, confidence) and friends' behaviour (friends)on the amount of alcohol consumed.

The runs start with a null model (Model 0) and build up to a full random coefficients model with covariates (Model 3).

Write a summary of the findings from the analyses, including a description of the fixed and random effects, and how the variance is partitioned at each stage in the analysis.

Question 4 continues .....

Question 4 continued

Table 4.2 Table of results from linear random coefficients models

	Model 0 Coeff(se)	Model 1 Coeff(se)	Model 2 Coeff(se)	Model 3 Coeff(se)
Fixed part				
Intercept	3.49 (0.33)	1.51 (0.39)	1.10 (0.27)	-0.30 (0.37)
time	-	1.84 (0.18)	2.93 (0.29)	2.00 (0.46)
happy				1.04 (0.60)
confidence	-	-	-	1.24 (0.54)
friends	-	-	-	2.57 (0.68)
Random part				
Time	-	-	30.44 (3.49)	30.37 (3.45)
Between individuals	43.25 (4.47)	49.65 (4.65)	14.83 (2.35)	8.67 (1.82)
Within individuals	93.26 (3.56)	88.94 (3.26)	49.34 (2.03)	48.98 (1.99)
Log likelihood	-8172.5007	-8121.6531	-7869.1136	-7816.3102

(c) If you were responsible for this analysis, how would you have modelled the development of drinking among young people? What metric of time would you use? How would you measure alcohol consumption? How would you analyse whether there is any variability in the influence of friends drinking on young people's alcohol consumption? Are there any covariates and/or interactions that you feel should be included in the analysis? Explain your reasons for your decisions. Note that there are many ways to answer this section. [13%]

END OF SECTION B

SECTION C

Candidates must answer Question 5 in Section C (worth 30%).

5. You are part of a team working on a project which aims to quantify the extent of poverty (defined as living in a household with an income lower than 60% of the national median income) over the past two decades, and to identify the factors associated with poverty at the individual and household level. The head of the research team is proposing to base the research on a data set which contains ten years of repeated cross-sections, collected at two-yearly intervals. You feel that it is important to use panel data for this project to investigate longitudinal measures of poverty, and poverty dynamics. So far, your suggestion has not been taken seriously, and you have decided to make your case in writing. [30%]

   Write a memo to the head of the research team outlining the reasons why it would be a good idea to use panel data for the research, outlining in detail the insights which may be obtained from panel data research, and the reasons why these additional insights are important for an understanding of the area. You may also like to acknowledge that cross-sectional data has a number of strengths too, and to suggest ways in which cross-sectional and panel data could be combined.

   [There are no "right" or "wrong" answers to this question. It will be marked as a whole, with marks for style and persuasiveness as well as for content.]

END OF SECTION C

END OF PAPER SC968-AS-MI