Exam:	20308 Econometrics (Winter term 2012/13)
Examiner:	Dr. Daniel Baumgarten
Number of questions:	The exam consists of 5 questions, all of which have to be answered.
Assessment:	A maximum of 120 points can be obtained. The number of points that can be obtained in each question is given in parentheses.
Admissible examination aids:	Calculator Dictionary
Additional information:	Selected values of the standard normal distribution (Table A1) and critical values of the chi-square distribution (Table A2) are given in the Appendix on p. 5. Should you have the impression that a question cannot be answered without an additional information or assumption, state this clearly and make a plausible assumption yourself.

Question 1 (26 points)

To analyse the determinants of students' course performance, you collect information on 908 students. All of them attended 4 different courses in the current semester (yielding $4 * 908 = 3632$ observations). The following variables are at your disposal:

performance	Course performance (=number of points obtained in the final exam)
attendance rate	Share of attended lectures and tutorials of the course (in %)
male	=1 if the student is male (and 0 otherwise)
last semester	=1 if the student is in his/her last semester (and 0 otherwise)

The regression results are given in the table below:

	OLS		Random effects		Fixed effects	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
attendance rate	0.80	(0.31)	0.75	(0.28)	0.42	(0.42)
male	−1.12	(0.95)	−0.95	(0.87)		
last semester	3.09	(1.53)	3.15	(1.40)		
constant	20.64	(12.09)	18.76	(11.55)		
Observations	3632		3632		3632	

1.1 Explain why these data are panel data. **(3 points)**

1.2 Interpret the OLS regression results economically (i.e., quantitatively) and statistically (ignoring the constant). **(6 points)**

1.3 Your colleague argues that the OLS coefficient of the attendance rate is likely to be upward biased. Do you agree? Justify your answer. **(6 points)**

1.4 To check the robustness of the OLS results, your colleague estimates a random effects and a fixed effects regression model. He also tells you that he did a Hausman test and obtained a test statistic of 7.44. **(8 points)**

i. Give the null (H_0) and the alternative hypothesis (H_A) of the Hausman test in this setting.

ii. What are the properties (in terms of consistency and efficiency) of the two estimators (random effects and fixed effects) under H_0 and H_A, respectively?

iii. Give the distribution of the test statistic and the degrees of freedom in the example at hand.

iv. Finally, make a test decision.

1.5 Why were the variables *male* and *last semester* not included in the fixed effects model? **(3 points)**

Question 2 (18 points)

2.1 Censoring vs truncation **(6 points)**

i. Explain the difference between censoring and truncation.

ii. Give one example of a truncated variable and one example of a censored variable.

2.2 Say you would like to analyse the determinants of the hours of work of married women (where many women have 0 hours of work). In what sense is the Tobit model restrictive in this application? Which alternative and less restrictive model(s) could you use? (**6 points**)

2.3 Which marginal effects could you potentially estimate in the Tobit model? Explain the different marginal effects using the previous example (hours of work of married women). (**6 points**)

Question 3 (25 points)

You use a probit regression to analyse the determinants of smoking (where the variable equals 1 for smokers and 0 for non-smokers). The explanatory variables are *educ* (years of schooling), *age* (in years), and a dummy variable which equals 1 for males (and 0 for females). The sample means of the variables and the regression results are given in the table below:

	Sample mean	Coefficient	Std. err.
educ	10	−0.08	(0.021)
age	40	0.01	(0.003)
male	0.5	0.60	(0.255)
constant	1	0.10	(0.100)
Observations		988	
Log likelihood		-1044.38	

3.1 Explain verbally (no formula!) differences between a linear model and a probit model in terms of the calculation of marginal effects. (**4 points**)

3.2 Calculate the marginal effects of the variables *educ* and *male* at the sample means, also showing how you calculate them. Note: The values displayed in Table A1 on p. 5 will be helpful in this respect. (**8 points**)

3.3 Give an economic interpretation of the marginal effect of *educ*. (Note: If you were not able to answer the previous question, assume that the marginal effect is −0.03.) (**3 points**)

3.4 What are two other ways of calculating (and reporting) marginal effects in the probit model? (**2 points**)

3.5 Your colleague argues that the variables *male* and *age* jointly do not matter for smoking. You would like to test this hypothesis. (**8 points**)

i. Give a formal notation of the null hypothesis. In this setting, what is the restricted model and what is the unrestricted model?

ii. Which tests could you use to test this hypothesis? Only state their names.

iii. State which models (restricted/unrestricted) have to be estimated for each of the suggested tests.

Question 4 (31 points)

You would like to analyse the determinants of choosing a particular night activity in a small town. People can choose among three different activities (*activity*): 1= going to the theatre (*theatre*), 2=going to the pub (*pub*), 3= staying at home (*home*). The explanatory variables are individual age (*age*, in years) and the price of the activity (*price*, in euros). The Stata regression output is given below:

```
Alternative-specific conditional logit          Number of obs      =      2190
Case variable: id                               Number of cases    =       730

Alternative variable: activity                  Alts per case: min =         3
                                                               avg =       3.0
                                                               max =         3

                                                Wald chi2(3)       =    246.59
Log likelihood = -353.06971                     Prob > chi2        =    0.0000
```

choice	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
activity						
price	-.031458	.0027041	-11.63	0.000	-.036758	-.026158
theatre	(base alternative)					
pub						
age	-.1304534	.0189233	-6.89	0.000	-.1675423	-.0933645
_cons	3.131191	.5820128	5.38	0.000	1.990466	4.271915
home						
age	.0963199	.0133602	7.21	0.000	.0701344	.1225053
_cons	-4.403239	.5474297	-8.04	0.000	-5.476181	-3.330296

4.1 What type of dependent variable do we have in this setting? (**3 points**)

4.2 In what important respect do the explanatory variables differ from each other? (**3 points**)

4.3 Interpret the coefficients of the explanatory variables (only economically, not statistically). (**8 points**)

4.4 A specification which includes individual income as additional explanatory variable yields a log likelihood value of -343.06971. Has the explanatory power of the model improved significantly? Give the name of the appropriate test, the test statistic, the degrees of freedom, and the critical value for the level of significance $\alpha = 0.05$. Calculate the value of the test statistic and make a decision. (**5 points**)

4.5 Now, a cinema opens in the small town so that you would like to re-estimate the model with four instead of three alternatives. Assume that so far, 60% of the people have stayed at home, 30% have gone to the pub, and 10% have gone to the theatre. After the cinema has opened, 50% of the people go to the cinema. (**12 points**)

i. Describe the independence of irrelevant alternatives (IIA) assumption in this specific context.

ii. If the IIA assumption were to hold, how would the choice probabilities of the other activities change after the cinema has opened?

iii. Do you think that the IIA assumption is likely to hold in this setting? Justify your answer.

iv. Assume that the IIA assumption does not hold. What are the consequences for your estimation and what can you do as a remedy?

Question 5 (20 points)

You would like to estimate a wage offer function for married women. Of the 753 women, wage offers can be observed only for those 428 women who are working. The following variables are at your disposal: *lwage* (the logarithm of the wage), *educ* (years of schooling), *exper* (labour market experience in years), *expersq* (experience squared), *age* (in years), and *inlf* (a dummy variable that equals 1 if the woman is working and 0 otherwise). Watch out: Do not overlook Question 5.4 below the regression output!

5.1 Why might it be problematic to estimate the wage offer function by OLS? (**5 points**)

5.2 A colleague suggests using Heckman's two-step estimator to estimate the wage offer function. Describe the procedure in detail (no formulae needed), also stating how many observations are used in each step. (**6 points**)

5.3 Have a look at the regression output. Does sample selection seem to be an important issue in this example? How do you know? (**5 points**)

```
Heckman selection model -- two-step estimates     Number of obs      =       753
(regression model with sample selection)          Censored obs       =       325
                                                  Uncensored obs     =       428

                                                  Wald chi2(4)       =     28.29
                                                  Prob > chi2        =    0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage						
educ	.1426316	.0416615	3.42	0.001	.0609765	.2242867
exper	.0992449	.0643027	1.54	0.123	-.0267859	.2252758
expersq	-.0017753	.0011397	-1.56	0.119	-.0040091	.0004585
age	-.0130547	.0155352	-0.84	0.401	-.0435032	.0173937
_cons	-1.388222	.9736757	-1.43	0.154	-3.296591	.5201474
inlf						
educ	.0855908	.022469	3.81	0.000	.0415524	.1296291
exper	.1287285	.0181037	7.11	0.000	.0932458	.1642111
expersq	-.0020214	.000588	-3.44	0.001	-.0031738	-.0008691
age	-.0316601	.0067258	-4.71	0.000	-.0448425	-.0184777
_cons	-.5209982	.413319	-1.26	0.207	-1.331089	.2890922
mills						
lambda	.7671333	.8209005	0.93	0.350	-.8418022	2.376069

5.4 Have again a careful look at the regression results. Do you see a possibility to improve the specification? Explain. (**4 points**)

Appendix

Table A1: Selected values of the probability density function (pdf) and the cumulative distribution function (cdf) of the **standard normal distribution**

z	-0.5	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.3	0.4	0.5
ϕ(z) (pdf)	0.352	0.368	0.381	0.391	0.397	0.399	0.397	0.391	0.381	0.368	0.352
Φ(z) (cdf)	0.309	0.345	0.382	0.421	0.460	0.500	0.540	0.579	0.618	0.655	0.691

Example: If z= 0.5, then ϕ(z)= 0.352 and Φ(z)= 0.691.

Table A2: Critical values (at the 5% level of significance) of the **chi-square distribution** for varying degrees of freedom (df)

df	1	2	3	4	5	6	7	8	9	10	11
critical value	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31	19.68

Example: The 5% critical value of the chi-square distribution with 8 degrees of freedom is 15.51.