Exam:	20308 Econometrics (Summer term 2013)
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 critical values of the chi-square distribution (Table A1) 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 (27 points)

To analyse whether there is a wage penalty for overweight workers, you have collected information on 1108 workers. The following variables are at your disposal:

ln wage	Logarithm of the hourly wage
overweight	=1 if the individual is overweight and 0 otherwise
age	Age (in years)
ln firm size	Logarithm of firm size

The regression results of the linear regression model (OLS) are given in the table below:

	Coef.	Std. err.
overweight	−0.06	(0.03)
age	0.10	(0.04)
age squared	−0.001	(0.001)
ln firm size	0.81	(0.55)
constant	1.64	(1.09)
Observations	1108	

1.1 Interpret the regression results economically (i.e., quantitatively), ignoring the constant. (**8 points**)

1.2 Is ln *firm size* a statistically significant predictor of wages? (**2 points**)

1.3 How would you test whether wages depend on workers' age? Give the name of the appropriate test, the null (H_0) and the alternative hypothesis (H_A), the test statistic, the distribution of the test statistic, and the degrees of freedom. (**5 points**)

1.4 Your colleague argues that the coefficient of the variable *overweight* is upward biased (i.e., less negative than the true coefficient). Do you agree? Justify your answer. (**6 points**)

1.5 Your colleague suggests using an instrumental variables estimator. (**6 points**)

i. Which assumptions must be fulfilled by an instrumental variable?

ii. Would *number of times per week eating at a fast food restaurant* be a valid instrument for the variable *overweight*? Justify your answer.

Question 2 (22 points)

You have panel data and would like to estimate a linear regression model.

2.1 What are panel data? (**4 points**)

2.2 Explain in what way the use of panel data and panel estimation techniques can help to solve certain problems of the simple linear regression model estimated by ordinary least squares (OLS). (**6 points**)

2.3 Discuss the advantages and the disadvantages of the random effects estimator and the fixed effects estimator. (**6 points**)

2.4 Describe two alternative ways of estimating the fixed effects regression model. (**6 points**)

Question 3 (28 points)

You are interested in the determinants of bonus payments. 299 of the 912 workers in your sample have not received any bonus (*bonus*=0). The explanatory variables are *educ* (years of schooling) and *male* (a dummy variable which equals 1 for males and 0 for females). The regression results are given below:

```
Tobit regression                                  Number of obs   =        912
                                                  LR chi2(2)      =     351.16
                                                  Prob > chi2     =     0.0000
Log likelihood = -4729.8711                       Pseudo R2       =     0.0358
```

bonus	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	130.2342	26.05619	5.00	0.000	79.097	181.3714
educ	64.02143	3.382981	18.92	0.000	57.38207	70.66078
_cons	-684.4828	53.73152	-12.74	0.000	-789.9349	-579.0307
/sigma	369.9142	11.3669			347.6058	392.2225

```
Obs. summary:        299  left-censored observations at bonus<=0
                     613     uncensored observations
                       0 right-censored observations
```

3.1 Explain the difference between censoring and truncation. Which of the two cases do we have in the present application? (**6 points**)

3.2 Why would it be problematic to estimate the model by ordinary least squares (OLS)? (**5 points**)

3.3 Interpret the coefficient of the variable *educ* (only economically, not statistically). (**3 points**)

3.4 Which (other) marginal effects could you potentially estimate in the Tobit model? Explain them for the example at hand. (**4 points**)

3.5 Your colleague suggests estimating a hurdle model (also known as two-part model) instead of a Tobit model. Describe the procedure (no formulae needed), also stating how many observations are used in each step. (**6 points**)

3.6 In what sense would a hurdle model be more flexible than a Tobit model? (**4 points**)

Question 4 (24 points)

You would like to analyse the determinants of the choice of the field of study at a small university. Students can choose among three different fields of study (*field*): 1= medicine, 2= economics, 3= business. The explanatory variables are the individual number of points obtained in the (general) university entry exam (*points*) and the current unemployment rate (in %) among graduates in the particular field of study (e.g., the current unemployment rate among economists, etc.).

4.1 Why is the mixed logit model appropriate in this setting? (**4 points**)

4.2 Interpret the marginal effects of the mixed logit model displayed below (only economically, not statistically). (**6 points**)

Pr(choice = business|1 selected) = .49623539

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
unemployment							
medicine	.004945	.000592	8.35	0.000	.003785	.006105	11.634
economics	.048493	.00485	10.00	0.000	.038987	.057999	5.0439
business	-.053438	.004603	-11.61	0.000	-.062459	-.044417	7.9879
casevars							
points	-.029664	.006973	-4.25	0.000	-.043331	-.015997	4.2256

4.3 Describe the independence of irrelevant alternatives (IIA) assumption. (**3 points**)

4.4 Do you think that the IIA assumption is likely to hold in this setting? Justify your answer. (**3 points**)

4.5 How can you test whether the IIA assumption holds? Describe the test in detail, stating (**8 points**)

i. the name of the test,

ii. the null (H_0) and the alternative hypothesis (H_A) of the test,

iii. the test procedure,

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

Question 5 (19 points)

You would like to analyse the determinants of individual life satisfaction, which is measured on a 3-point scale: 1= not at all satisfied; 2= partly satisfied; 3= very satisfied. The following variables are at your disposal: *income* (monthly income measured in 1000 EUR), *age* (age in years), *kids* (number of kids). The regression results are given below:

```
Ordered logistic regression                     Number of obs   =        742
                                                LR chi2(3)      =     591.34
                                                Prob > chi2     =     0.0000
Log likelihood = -519.49673                     Pseudo R2       =     0.3627
```

satisfaction	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
income	1.144245	.087108	13.14	0.000	.9735165	1.314973
age	.1173498	.0074545	15.74	0.000	.1027392	.1319603
kids	.759337	.060476	12.56	0.000	.6408063	.8778678
/cut1	8.54528	.5260901			7.514163	9.576398
/cut2	11.20015	.6096874			10.00519	12.39512

5.1 Why is the ordered logit model an appropriate regression model in this setting? (**4 points**)

5.2 Given the regression results, what can you say about the marginal effects of *income* on the three categories of the dependent variable? (**5 points**)

5.3 A specification which includes the yearly number of sunshine hours (*sunshine*) and the crime rate (*crime*) in the home town of the individual as additional explanatory variable yields a log likelihood value of -512. 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**)

5.4 Your colleague suggests estimating a multinomial logit (MNL) instead of an ordered logit. Comment on this suggestion. (**5 points**)

Appendix

Table A1: 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.