Exam ECONOMETRICS (20308)
Summer Term 2011

REMARKS:

- The exam comprises 5 tasks containing subtasks. All of them have to be answered.
- You have 120 minutes to solve the exam and you can get 120 points in total (1 point $\hat{=}$ 1 minute).
- Please indicate the meaning of the variables you use.
- Permitted examination aid is a dictionary.

Good luck!

1. **General Aspects**

   (a) One widely applied estimation method in econometrics is the method of ordinary least squares. Explain why this approach is difficult to reconcile with microeconometric data. ***(6 minutes)***

   (b) Define the term "time series data". What is the difference to panel data? ***(3 minutes)***

   (c) Describe the idea and functioning of Maximum Likelihood estimation and name three possible algorithms to maximize the Log-Likelihood function. ***(6 minutes)***

2. **Binary Qualitative Dependent Variables**

   (a) Explain why the standard multiple regression model is not an appropriate approach to deal with binary dependent variables. ***(4 minutes)***

   (b) Derive the latent variable framework for a binary variable ($y = 0, 1$) assuming a threshold value of zero. State the marginal effect of a change in an explanatory variable on the probability to observe an outcome $y = 1$. As a contrast state the standard multiple regression model and its marginal effect of an explanatory variable. What is the important difference? ***(12 minutes)***

   (c) State the sample likelihood and log-likelihood function for a binary variable and derive the first-order-condition for optimizing the function with respect to the Maximum-Likelihood-estimator (simplification of the derivative is not necessary). ***(6 minutes)***

   (d) Define the common parametrization for the probability of an outcome $y = 0, 1$ in the Probit Model. Derive the marginal effect of a change in an explanatory variable on the change in the probability and explain how it is interpreted. ***(8 minutes)***

3. **Multinomial Models**

   (a) Define the probability of one multinomial model of your own choice (formula and explanation of the parameters). Indicate the name of the model and provide an example. ***(6 minutes)***

(b) For the Conditional Logit model, the marginal effect of a change in the exogenous parameter $x_k, k = 1,...,J$ on the probability to choose alternative $j = 1,...,J$ is

$$\frac{\partial p_j}{\partial x_k} = p_j(\mathbf{1}(j = k) - p_k)\beta \tag{1}$$

How does the marginal change affect the $j = 1,...,J$ probabilities in the model? Explain the terms own-effect and cross-effect. How are they related? ***(5 minutes)***

(c) Consider the estimation of Multinomial models for the choice between four different modes of fishing beach, pier, private, charter. The following table (from Cameron and Trivedi, 2005, p.493) displays estimated coefficients for a choice between the alternative modes, estimated as a Conditional Logit (CL), Multinomial Logit (MNL) and a Mixed Logit Model (Mixed).

Table 15.2. *Fishing Mode Multinomial Choice: Logit Estimates*[a]

			Model type		
Regressor	**Type**	**Coefficient**	**CL**	**MNL**	**Mixed**
Price (P)	Specific	β_P	−0.021	–	−0.025
Catch rate (C)	Specific	β_{CR}	0.953	–	0.358
Intercept	Invariant	α_1 : *Beach*	–	0.0	0.0
		α_2 : *Pier*	–	0.814	0.778
		α_3 : *Private*	–	0.739	0.527
		α_4 : *Charter*	–	1.341	1.694
Income (I)	Invariant	β_{I1} : *Beach*	–	0.0	0.0
		β_{I2} : *Pier*	–	−0.143	−0.128
		β_{I3} : *Private*	–	0.092	0.089
		β_{I4} : *Charter*	–	−0.032	−0.033
− ln L			−1311	−1477	−1215
Pseudo-R^2			0.162	0.099	0.258

[a] Type of regresssor is alternative-specific (price and catch rate) or alternative-invariant (income). Outcomes are (1) beach, (2) pier, (3) private, and (4) charter. MLE estimates are for conditional logit (CL), multinomial logit (MNL), and mixed logit (Mixed) models. MNL and Mixed models are normalized to base category beach. All estimates except that for β_{I4} are statistically significant at 5%.

What information from the table can you use to conclude about the goodness of fit of the models? Verbally explain how the measure is computed. What model fits the data best? ***(6 minutes)***

(d) Explain with the help of an example what the assumption of independence of irrelevant alternatives (IIA) implies and where it is assumed. How can it be tested? ***(8 minutes)***

(e) When is it suitable to use a Nested Logit Model? State the main assumptions and explain how the probability to choose a certain alternative is modeled with the help of an example. ***(5 minutes)***

4. **Selection Models**

(a) What possible sources of biased sampling are there? Use examples to describe how the different possibilities lead to biased samples. ***(9 minutes)***

(b) State the sample log-likelihood function for the Standard Normal Regression Model (Tobit Model). What do the two parts of the log-likelihood function stand for? ***(8 minutes)***

(c) The conditional expected value in the Tobit Model is

$$E[y|x] = \Phi\left(\frac{x'\beta}{\sigma}\right) \times (x'\beta) + \sigma\phi\left(\frac{x'\beta}{\sigma}\right).$$

State the marginal effect and contrast it with the marginal effect when using ordinary least squares. How can both effects be compared? ***(5 minutes)***

(d) The Heckit Model is an approach to account for biased parameter estimates caused by selected samples. State the model setup and shortly explain how it is estimated. What is the role of the so-called inverse Mill's ratio? ***(8 minutes)***

5. **Panel Models**

(a) When data is available for two points in time, we can evaluate policy measures. One possible approach is the Difference-in-Difference (DID) estimator. State the model and the formula of the DID-estimator and verbally explain what it measures. ***(7 minutes)***

(b) What does the Chow test do? Provide an example. State the model to apply the test and verbally explain how the test functions. ***(8 minutes)***