Examination:	20386 – Financial Econometrics	**Summer Term 2011**
Examiner:	Prof. Dr. Dr. Bodo Vogt	**29.07.2011**
Aids permitted:	pocket calculators, English dictionaries, two sheets of paper (double sided) typed or hand written with formulas, definitions, solved problems etc.	

The examination is comprised of **six** problems. All of them are to be solved. You can reach a maximum of **120 points. Good luck!**

Problem 1: (20 points)

Please give a short answer of the following questions.

a) Explain why a stochastic error term is added to a regression equation. (4 points)
b) Show how the residual e_i is defined. Is there any difference between the residual and the stochastic error term? If yes, what is the distinction between them? (4 points)
c) Consider the following linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Please define the meaning of the regression coefficient β_1 in this equation. (4 points)

d) Multicollinearity, serial correlation and heteroscedasticity are considered to be violations of some of the classical assumptions. Please write down which classical assumptions are violated in the presence of multicollinearity, serial correlation and heteroscedasticity. Make sure that you explain what these assumptions state and not simply write their numbers. (4 points)
e) Define what is meant under unbiased estimator. (4 points)

Problem 2: (20 points)

Please answer the following multiple choice questions. Each question has only **one** correct answer. For each correct answer you receive 2 points.

1) Ordinary Least Squares is a regression estimation technique that calculates the $\hat{\beta}s$ so as to minimize:

   a) The sum of the squared residuals
   b) The sum of the squared stochastic error terms
   c) The sum of the squared coefficients of determination

2) Omitted variables are known to cause a bias. The omitted variable bias is equal to:

a) $Bias = \beta_{in} f(r_{in,om})$

b) $Bias = \beta_{om} f(r_{in,om})$

c) $Bias = \beta_{om} \overline{R}^2_{in,om}$

3) Which of the following statements is **true**:

a) Imperfect multicollinearity is caused by a specification error
b) Imperfect multicollinearity is a deterministic linear relationship existing between two or more independent variables
c) Imperfect multicollinearity is a strong but not deterministic linear relationship between two or more independent variables

4) There are several methods for the detection of severe multicollinearity. Which of the following method is **incorrect**?

a) High simple correlation coefficients between the explanatory variables hint at the possibility of severe multicollinearity
b) High variance inflation factors hint at the possibility of severe multicollinearity
c) High total sum of squares hints at the possibility of severe multicollinearity

5) Which of the following statements is **incorrect**:

a) Pure serial correlation causes bias in the coefficient estimates
b) Pure serial correlation causes OLS to no longer be the minimum variance estimator
c) Pure serial correlation causes the OLS estimates of the $SE(\hat{\beta})s$ to be biased, leading to unreliable hypothesis testing

6) Which of the following methods can be used as a remedy for pure heteroscedasticity:

a) Generalized Least Squares
b) Weighted least Squares
c) Two-Stage Least Squares

7) Which of the following is a Distributed Lag Model:

a) $Y_t = \alpha_0 + \beta_0 X_t + \lambda Y_{t-1} + \varepsilon_t$

b) $Y_t = \beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t} + \varepsilon_t$

c) $Y_t = \alpha Y_{t-1} + \varepsilon_t$

a) Granger causality or precedence is a circumstance in which one time series variable consistently and predictably changes before another variable
b) Spurious correlation is a strong relationship between two or more variables that is caused by a real casual relationship
c) Spurious correlation might be caused by a nonstationary time series

Problem 3: (30 points)

In university U a lot of the lectures are offered as elective courses. The number of students who are going to attend a given course is, therefore, unknown before the beginning of the semester. This lack of knowledge about the number of students willing to attend a given lecture has turned out to be really problematic over the last semesters – it has often happened that lectures with very high attendance were scheduled to be held in relatively small lecture halls where there are not enough sitting places for all the interested students. The department responsible for the scheduling of the lecture halls, therefore, decided to run a regression model in an attempt to be able to better forecast the number of students who are going to attend a given elective course. Cross-sectional data for 30 elective courses were collected and the following regression equation was estimated (standard errors in parentheses):

$$\hat{S}_i = 3 + 2E_i + 10AG_i + 3PR_i - 2LS_i + 10R_i$$
$$(1.9)\quad (5)\qquad (3.1)\quad (1)\qquad (2)$$

$$\bar{R}^2 = 0.80 \qquad N=30$$

S_i – Number of students attending the ith elective course;

E_i – Number of years of experience of the lecturer teaching the ith elective course;

AG_i – Last semester's average grade in the ith elective course;

PR_i – Last semester's passing rate in the ith elective course;

LS_i – Number of students who attended the ith elective course last semester;

R_i – A dummy variable equal to 1 if the ith elective course is very relevant for one or several study programs and equal to 0 otherwise.

a) Create and test appropriate hypotheses (one-sided) about the slope coefficients of the equation at the 5 percent level. (11 points)
b) What economic problems (out of omitted variables, irrelevant variables and multicollinearity), if any, appear to be in this equation? Explain. (8 points)
c) Do you think that the exclusion of some of the independent variables in the above equation can improve the regression equation? If yes, make specific suggestions. Do you think that the inclusion of additional explanatory variables might improve the regression equation? If yes, in what way should that (these) variables be related to the dependent and independent variables of the estimated equation. Explain your answer. (11 points)

Problem 4: (15 points)

How does two-stage least squares (2SLS) work? What is the identification problem and when can 2SLS be applied? Why is 2SLS a good method? (5 points for each sub-question)

Problem 5: (15 points)

A liner probability model is a linear in the coefficients equation used to explain a dummy dependent variable:

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

a) The above model can be estimated with OLS:

$$\hat{D}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

What is the meaning of $\hat{D}_i$, β_1 and β_2 ? (5 points)

b) What are the problems of the Linear Probability Model?(5 points)
c) Suggest two other techniques (simply write them down, a formal definition is not required), alternative to the Linear Probability Model, which can be used to estimate equations with dummy dependent variables. By what methods are the models you have suggested usually estimated? (5 points)

Problem 6: (20 points)

a) What is meant under univariate time series modeling? (4 points)
b) Define the Lag operator L. What is the inverse of $1-\alpha L$? (4 points)
c) Please write down the equation of an ARMA (2, 3) process? (4 points)
d) What does ARIMA mean? (2 points)
e) Assume that the following AR(1) process is stationary and ε_t is white noise:

$$y_t = \alpha y_{t-1} + \varepsilon_t$$

Please derive the expectation of the above AR (1) process. (6 points)

Critical Values of the t-Distribution

	Level of Significance					
d.f.	**One Sided:**	**10.0%**	**5.0%**	**2.5%**	**1.0%**	**0.5%**
	Two Sided:	**20.0%**	**10.0%**	**5.0%**	**2.0%**	**1.0%**
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.604
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.106
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807
24		1.318	1.711	2.064	2.492	2.797
25		1.316	1.708	2.060	2.485	2.787
26		1.315	1.706	2.056	2.479	2.779
27		1.314	1.703	2.052	2.473	2.771
28		1.313	1.701	2.048	2.467	2.763
29		1.311	1.699	2.045	2.462	2.756
30		1.310	1.697	2.042	2.457	2.750
40		1.303	1.684	2.021	2.423	2.704
60		1.296	1.671	2.000	2.390	2.660
120		1.289	1.658	1.980	2.358	2.617
∞		1.282	1.645	1.960	2.326	2.576