Stockholm University
Department of Statistics
Ellinor Fackle Fornius, Jessica Franzén

HT 2008

EXAM IN ECONOMETRICS, PART I
2009-01-08

Exam hours: 9.00-14.00
Calculator and one book of your choice are allowed.
The exam consists of five problems which are worth 10 points each. For the maximum of 10 points, detailed and clear solutions are required.

Problem 1. (10 points)

Let Y and X denote the labor force participation rate of women in 1972 and 1968, respectively, in each of 19 cities. The regression output for this data set is given below. It was also found that the explained sum of squares $ESS = 0.0358$ and the residual sum of squares $RSS = 0.0544$. Suppose that the model $Y_i = \beta_1 + \beta_2 X_i + u_i$ satisfies the usual regression assumptions.

$Variable$	$Coefficient$	$se\left(\widehat{\beta}\right)$	$t-test$	$p-value$
Constant	0.203311	0.0976	2.08	0.0526
X	0.656040	0.1961	3.35	0.0036
$n = 19$	$R^2 = 0.397$	$\overline{R}^2 = 0.362$	$\widehat{\sigma} = 0.0566$	$df = 17$

a) Compute $Var(Y)$ and $Corr(Y, X)$.

b) Suppose that the participation rate of women in 1968 in a given city is 45%, that is $X = 0.45$. What is the estimated participation rate of women in 1972 for the same city?

c) Suppose further that the mean and variance of the participation rate of women in 1968 are 0.5 and 0.005, respectively. Construct a 95 % prediction interval for the estimate in b).

d) Test the hypothesis: $H_0: \ \beta_2 = 1$ versus $H_1: \ \beta_2 > 1$ at the $\alpha = 5\%$ significance level.

e) If Y and X were reversed in the above regression, what would you expect R^2 to be?

Problem 2. (10 points)

a) Define perfect multicollinearity.

b) State with reasons whether the following statement is true or false:

Despite perfect multicollinearity, OLS estimators are BLUE.

c) Consider the model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

Assume that there are low pair-wise correlations between the regressors X_2, X_3 and X_4. Can you with certainty state that there is or is not any problems with multicollinearity in this case? Motivate your answer.

Problem 3. (10 points)

a) Explain the term heteroscedasiticity. Why should we be cautious in the presence of heteroscedasiticity?

b) We have 55 data points for the dependent variable Y and the explanatory variable X. Use the regression results below to perform White's general heteroscedasiticity test. State null and alternative hypothesis, test statistic and consclusion.

Dependent variable: Y

	Variable Coefficient	Std. error	t	Prob.
Constant	94.20878	50.85635	1.852449	0.0695
X	0.436809	0.078323	5.577047	0.0000

$R^2 = 0.369824$

Dependent variable: $\widehat{u}_i^2$

	Variable Coefficient	Std. error	t	Prob.
Constant	13044.00	21156.58	0.616545	0.5402
X	-53.12260	71.48347	-0.743145	0.4607
X^2	0.059795	0.058860	1.015887	0.3144

$R^2 = 0.134082$

c) Explain the method of generalized least squares (GLS) and how it can be applied in the presence of heteroscedasiticity.

Problem 4. (10 points)

Consider a data material where X is temperature in Fahrenheit at takeoff for a space shuttle and Y is a binary variable equal to 1 if takeoff failed and 0 if takeoff was succesful.

a) The probability of takeoff failure will be estimated with the *Logit* model. Write the model expression and explain how the parameters in the model can be estimated.

b) The model is fitted and the result is given in the SAS output below. Interpret the regression coefficient for temperature, $\widehat{\beta}_2$, and the odds-ratio $e^{\widehat{\beta}_2}$.

Model Information

Distribution	Binomial
Link Function	Logit
Dependent Variable	y

Number of Observations Read	22
Number of Observations Used	22
Number of Events	6
Number of Trials	22

Response Profile

Ordered Value	y	Total Frequency
1	1	6
2	0	16

PROC GENMOD is modeling the probability that y='1'.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	20	14.3770	0.7188
Scaled Deviance	20	14.3770	0.7188
Pearson Chi-Square	20	16.7049	0.8352
Scaled Pearson X2	20	16.7049	0.8352
Log Likelihood		-7.1885	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	23.4033	11.8317	0.2137	46.5929	3.91	0.0479
x	1	-0.3610	0.1755	-0.7050	-0.0170	4.23	0.0397

c) From the fitted model, find the probability of takeoff failure when the temperature is 31 degrees Fahrenheit.

d) Would you advise launching the space shuttle at a day when temperature is 31 degrees Fahrenheit?

Problem 5. (10 points)

Consider the regression model

$$Y_i = \alpha + \beta X_i + u_i \tag{1}$$

with corresponding parameter estimators $\widehat{\alpha}$ and $\widehat{\beta}$.

There are errors of measurement in the dependent variable Y such that we can only observe Y_i^* according to

$$Y_i^* = Y_i + \varepsilon_i$$
$$\text{where } E\left(\varepsilon_i\right) = 0$$

Therefore, instead of estimating model (1) the following model is estimated

$$Y_i^* = \alpha + \beta X_i + u_i + \varepsilon_i$$

with corresponding parameter estimators $\widehat{\alpha}^*$ and $\widehat{\beta}^*$. Show that

$$E\left(\widehat{\beta}^*\right) = \beta$$

Hint:

$$\widehat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}$$

$$\widehat{\beta}^* = \frac{\sum x_i Y_i^*}{\sum x_i^2}$$