UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Postponed exam: **ECON4150 – Introductory Econometrics**

Date of exam: Thursday, June 17, 2010

Time for exam: 09:00 a.m. – 12:00 noon

The problem set covers 4 pages (incl. cover sheet)

Resources allowed:

- All written and printer resources, as well as calculator

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

Introduction to Econometrics- Econ 4150
Final Exam, Spring 2010

The exam consists of 100 points. Please allocate your time to each problem accordingly.

Problem 1 (40 points)

We want to study the association between a person's responses to a mailing from a charitable organization, and various background variables . The variable *respond* is a dummy variable equal to one if a person responded with a contribution on the most recent mailing sent by the organization. The variable *resplast* is a dummy variable equal to one if the person responded to the previous mailing, and $avggift$ is the average of past gifts.

We assume the following model

$$respond = \beta_0 + \beta_1 resplast + \beta_2 avgift + u \tag{1}$$

where u is the error term. STATA is used to estimate the model using OLS, with the following results.

Source	SS	df	MS
Model	112.749085	2	56.3745426
Residual	911.530905	4295	0 .213723542
Total	1024.27999	4297	0.240046869

Number of obs = 4268
F(2, 4265) = 263.77
Prob > F = 0.0000
R-squared = 0.1101
Adj R-squared = 0.1097

respond	Coef	Std Err	t	P>|t|	[95% Conf.	Interval]
resplast	0.3436605	0.0149953	22.92	0.000	0.314262	0.373059
$avgift$	0.0001536	0.0000899	1.71	0.088	-0.0000227	0.0003299
_cons	0.282088	0.0088379	31.92	0.000	0.0003299	0.2994148

(i) How large is the proportion of the variance in *respond* which is explained by this model?

(ii) What is the interpretation of the reported result for Adj R-squared?

(iii) Imagine that neither R-squared nor Adj R-squared were reported. How would you compute the value of Adj R-squared based on other reported results?

(iv) How do you interpret the coefficient of *resplast*?

(v) The estimated model parameters can be used to predict the expected value of *respond*, given certain values of the independent variables. Why can we interpret this expected value as a probability? Give the statistical proof.

(vi) Why is the error term u of model (1) heteroskedastic?

(vii) Because of the heteroskedastic error term, the usual t-statistics from OLS regression (such as those reported above) do not have a t distribution. Reformulate the model in such a way that the error term of the new model is homoskedastic. State all your assumptions.

(viii) How large is the variance of the error term in the new model?

Problem 2 (30 points)

Let *gprice* denote the monthly growth in the overall price level and let *gw* be the monthly growth in hourly wages. Using monthly data we obtain the following outcome for the following distributed lag model ((with standard errors in parentheses)

$$\begin{aligned}\widehat{gprice} &= \underset{(0.00057)}{-.00093} + \underset{(0.052)}{0.119}gw + \underset{(0.039)}{0.097}gw_{-1} + \underset{(0.039)}{0.04}\,gw_{-2} + \underset{(0.039)}{0.038}gw_{-3} + \underset{(0.039)}{0.081}gw_{-4} \\ &\quad + \underset{(0.039)}{0.107}gw_{-5} + \underset{(0.039)}{0.095}gw_{-6} + \underset{(0.039)}{0.104}gw_{-7} + \underset{(0.039)}{0.103}gw_{-8} + \underset{(0.039)}{0.159}gw_{-9} \\ &\quad + \underset{(0.039)}{0.110}gw_{-10} + \underset{(0.039)}{0.103}gw_{-11} + \underset{(0.052)}{0.016}gw_{-12} \\ n &= 273, R^2 = 0.317\end{aligned}$$

(i) Sketch the estimated lag distribution. At what lag is the effect of *gw* on *gprice* the largest and the smallest?

(ii) For which lags are these effects significant?

(iii) What is the estimated long-run propensity (LRP)?. Is it much different than one?. Explain what the LRP tells us for this model?

(iv) What regression would you run to obtain the standard error of the LRP?

(v) How would you test the joint significance of six more lags of *gw?* What would be the degrees of freedom in the F distribution?

Problem 3 (30 points)

Decide if you agree or disagree with the following statements and provide a brief explanation to your answer

(i) Weighted least squares is preferred to OLS when an important variable has been omitted from the model.

(ii) A trending variable cannot be used as the dependent variable in multiple regression analysis.

Please answer the following questions

(iii) When the errors in a regression have AR(1) serial correlation, why do the OLS standard errors tend to underestimate the sampling variation in the $\widehat{\beta}'s$?. Is it always true that the OLS standard errors are too small?

(iv) Suppose you want to analyze the return to education between men and women. Your objective is to test whether they are statistically equal. How would you formulate the model?. How would you test the hypothesis about the equality in returns for men and women?