UNIVERSITY OF OSLO
DEPARTMENT OF ECONOMICS

Postponed exam:ECON4160 – Econometrics – Modeling and systems estimation, spring 2004

Date of exam: Wednesday, August 18, 2004

Time for exam: 9:00 a.m. – 12:00 noon

The problem set covers 7 pages

Resources allowed:
- All written and printed resources and calculator

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

Problem 1

Page I of the enclosed data sheets reports two regression analyses of the probability of having done unregistered work during the last year. The variables included in the models are:

Unreg = a dummy variable taking the value of 1 if the individual has reported to do unregistered work during the last 12 months and the value of 0 otherwise. This is the dependent variable.
Y80 = a dummy variable taking the value of 1 if the observation is from 1980 and 0 if the observation is from 2003 (only observations from these two years are included in the analyses).
Woman = a dummy variable taking the value of 1 if the individual is a woman. (0 otherwise)
College = a dummy variable taking the value of 1 if the individual has higher education than high-school. (0 otherwise)
Age = the age of the individual
Risk = the perceived probability of being exposed by the tax authorities if the individual undertakes unregistered work. Takes the value from 0 to 1.

The first model reports results from a Linear probability model. The second model reports results from a Logit model.

a) Explain why the logit-model is preferable to the linear probability model in this case. Based on the results in the logit-model, is it possible to conclude that the level of unregistered work has changed over the twenty years between the two surveys?
b) The mean value of unreg= 0.15. Calculate the marginal effect of age for the average individual in both the linear probability model and the logit model. Explain how you calculated this effect. Consider an individual with a very small probability of doing unregistered work, for example 0.01. Calculate the marginal effect of age for this individual in both models.
c) Give an interpretation of the coefficient for the variable risk in the Logit model, using a random utility interpretation of the discrete choice model.

Problem 2

Lnwdisp is a measure of wage dispersion in a country in a given year (measured by the log of the 9th decile over the 1st decile of hourly wages). Theories of wage determination suggest that wage dispersion may be lower where there is coordination across firms in wage negotiations. Page II and III of the data sheets report some regression analyses of the relationship between wage dispersion and coordination in wage negotiations, using a panel of data from several years from 16 countries. The unit of observation is country*X*year. Coordination is measured by an index ranging from 1 to 4, with 1 being completely decentralised wage setting at the level of the establishment, and 4 being completely centralised wage setting at the national level.

The first model on page II of the data sheets shows that there is a negative correlation between wage dispersion and coordination in wage negotiations. Furthermore, the trend variable t suggests that there has been a negative trend in wage dispersion over the years of observation. The second model confirms that this negative relationship remains significant even when we control for the share of higher education in the population (edushare), the openness of the economy to international trade (open) and the log of union density (lnudens).

a) The coefficient for t changed very little, while the coefficient for coordination became smaller in size after inclusion of the additional variables. Discuss briefly why this happened.

The model may be written like this, for country i in year t:

1) $\text{Lnwdens}_{it} = a + X_{it}\, b + d\ \text{Coordination}_{it} + u_{it}$

X is the vector of explanatory variables in addition to coordination. An objection to the OLS regression model may be that coordination could be endogenous: Perhaps lower wage dispersion promotes coordination in bargaining. Such a relationship could be formalized as follows:

2) $\text{Coordination}_{it} = A + Z_{it}\, B + D\ \text{Lnwdens}_{it} + v_{it}$

Z is a vector of explanatory variables that may influence the level of coordination. We now have a model with two equations and two endogenous variables.

b) Explain why the OLS regression model of equation (1) may produce biased results in this case. What is required of the vector Z in this model, for equation 1 to be exactly identified?

On Page III) of the data sheets, the results from a 2SLS regression analysis are presented. The effect of coordination is still significantly negative, albeit somewhat smaller in size. In this specification, two political variables (the share of left and right parties in the cabinet) as well as the total size of the population (L, R and lnP), are used as instruments for coordination in wage negotiations.

c) Specify the assumptions necessary for the three variables L, R and lnP to be valid instruments. Explain why these assumptions are necessary to for the 2SLS estimators to be consistent. Based on the information given on the data sheets, what would be your estimate of the bias in the OLS specification? Suggest a way of testing if endogeneity bias is a problem for the OLS specification of equation 1.

PAGE I)

OLS estimates

```
      Source |       SS       df       MS              Number of obs =
1317
-------------+------------------------------           F(  5,  1311) =
32.16
       Model |  18.6848826     5  3.73697652           Prob > F      =
0.0000
    Residual |  152.332581  1311  .116195714           R-squared     =
0.1093
-------------+------------------------------           Adj R-squared =
0.1059
       Total |  171.017464  1316   .12995248           Root MSE      =
.34087

------------------------------------------------------------------------------
       unreg |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
         y80 |   .0564154   .0204257     2.76   0.006     .0163447
.0964862
       woman |  -.1185244    .019515    -6.07   0.000    -.1568084   -
.0802404
         age |  -.0035995   .0007839    -4.59   0.000    -.0051373   -
.0020617
     college |  -.0903182     .02055    -4.40   0.000    -.1306327   -
.0500037
        risk |  -.2064121   .0411888    -5.01   0.000    -.2872153   -
.1256089
       _cons |   .4414538   .0395904    11.15   0.000     .3637863
.5191212
------------------------------------------------------------------------------
```

Logit estimates

```
                                                  Number of obs   =
1317
                                                  LR chi2(5)      =
160.96
                                                  Prob > chi2     =
0.0000
Log likelihood = -483.88787                       Pseudo R2       =
0.1426

------------------------------------------------------------------------------
       unreg |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
         y80 |   .4633312   .1695316     2.73   0.006     .1310553
.7956071
       woman |    -1.1454   .1918774    -5.97   0.000    -1.521473   -
.7693271
         age |  -.0304774   .0069908    -4.36   0.000    -.0441791   -
.0167757
```

```
     college |  -.8514831   .1983471    -4.29   0.000    -1.240236   -
.4627299
        risk |  -2.227739   .4235508    -5.26   0.000    -3.057884   -
1.397595
       _cons |   .6159416   .3311691     1.86   0.063     -.033138
1.265021
------------------------------------------------------------------------------
---
```

Page II)

The REG Procedure
Model: MODEL1
Dependent Variable: lnwdisp

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6.67309	3.33655	238.19	<.0001
Error	191	2.67556	0.01401		
Corrected Total	193	9.34865			

Root MSE	0.11836	R-Square	0.7138
Dependent Mean	1.01167	Adj R-Sq	0.7108
Coeff Var	11.69905		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.71849	0.03655	47.02	<.0001
t	1	-0.00786	0.00188	-4.18	<.0001
coordination	1	-0.24782	0.01136	-21.82	<.0001

The REG Procedure
Model: MODEL2
Dependent Variable: lnwdisp

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	7.54979	1.50996	157.81	<.0001
Error	188	1.79886	0.00957		
Corrected Total	193	9.34865			

Root MSE	0.09782	R-Square	0.8076
Dependent Mean	1.01167	Adj R-Sq	0.8025
Coeff Var	9.66896		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.71810	0.05860	29.32	<.0001
t	1	-0.00775	0.00158	-4.91	<.0001
edushare	1	0.64566	0.09303	6.94	<.0001
OPEN	1	-0.11734	0.02592	-4.53	<.0001
lnudens	1	-0.05057	0.01717	-2.94	0.0036
coordination	1	-0.19524	0.01237	-15.78	<.0001

Page III)

The SYSLIN Procedure

Two-Stage Least Squares Estimation

Model lnwdisp

Dependent Variable lnwdisp

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	5.550313	1.110063	105.73	<.0001
Error	188	1.973725	0.010499		
Corrected Total	193	9.348649			

Root MSE	0.10246	R-Square	0.73768
Dependent Mean	1.01167	Adj R-Sq	0.73070
Coeff Var	10.12802		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.731382	0.061582	28.12	<.0001
t	1	-0.00650	0.001720	-3.78	0.0002
edushare	1	0.702031	0.099671	7.04	<.0001
lnudens	1	-0.09321	0.023972	-3.89	0.0001
coordination	1	-0.14236	0.023540	-6.05	<.0001
OPEN	1	-0.13919	0.028337	-4.91	<.0001

Test for Overidentifying Restrictions

Num DF	Den DF	F Value	Pr > F
2	186	1.65	0.1943