ECON4160 – ECONOMETRICS – MODELING AND SYSTEMS ESTIMATION

EXAM, SPRING 2004

Sensorveiledning (Originaltekst er satt ned til 9pt. mens veiledningen er i 12pt)

Problem 1

We want to estimate the relationship between log hourly wage and labour market experience. The relationship between log hourly wage and experience is given by:

(1) $LNWAGE_i = \alpha + \beta\ EXPERIENCE_i + u_{i,}$

where α and β are unknown parameters and u_i is a stochastic disturbance.

We have E(u| EXPERIENCE) = 0 and assume homoskedastic disturbances with no autocorrelation. For this problem we have constructed an experience variable which is considered as the sum of the true experience variable (EXPERIENCE) and a random error (ERROR). We now assume that we cannot observe the true EXPERIENCE variable, only the badly measured one. The badly measured experience variable is thus defined as:

(2) EXPERROR = EXPERIENCE + ERROR,

where ERROR is assumed to be uncorrelated with both EXPERIENCE and u. On the top of page 1 of the printout you will find the mean and variance of the error as well as of the badly measured experience variable for a sample of 681 Norwegian men between ages 18 and 45. The next part of the printout page 1 reports the variances and covariances of log wage (LNWAGE) and the badly measured experience variable. The problem is that we now only observe EXPERROR and not the true value of EXPERIENCE.

a) From the reported information on the empirical first and second order moments of the distributions of EXPERROR, LNWAGE and ERROR on top of page 1 (the means, variances and covariances) it should be possible to identify β. Explain how and calculate a consistent estimator of β based only on these statistics. Suggest which of these empirical moments that one would be less likely to have information on in real-world data samples.

This question is discussed in lecture note 6 section 4 on Identification under additional information. We have $\beta = \sigma_{xy} / \sigma_x^2 - \sigma_v^2$. The empirical moments provide us with a consistent estimator. One is less likely to have information on the variance of the error.

In terms of the observable variables we have:

(3) $LNWAGE_i = \alpha + \beta\ EXPERROR_i + w_i$

The regression results at (A) in the printout reports an OLS estimate of β in equation (3) of 0.00697.

b) Explain why this OLS estimator is a biased estimator of β. Do you have enough information on the top of this page to give an estimate of the magnitude of the bias in this case?

They should be able to show that cov(Experror, w) $\neq 0$ since w = u - βERROR and compare cov(lnwage, experror) = β var (Experror) + cov (experror,w) to the OLS estimator to establish a bias. $1/(1+\sigma_v^2/\sigma_{x*}^2)$ is about ½, since the two variances are close to equal.

c) The results at (B) in the printout are the results from in instrumental variable regression. In this model, we have used information on the person's AGE as an instrument for the experience variable. Under what crucial assumptions does this method provide us with a consistent estimator of β? Do you find these assumptions reasonable

The two crucial assumptions are that the cov(experience,age) ≠ 0 and that cov(error,age) = 0. (If the second assumption holds, the first one is equivalent to cov(experror,age) ≠ 0.) The good students should be able to show formally that these assumptions provide us with a consistent estimator.

Problem 2.

A more comprehensive model of log wages is estimated and the results reported at (C) and (D) in the printout. This model is estimated on data from all workers from 18 to 65 years of age and include years of schooling (YRSCHOOL) as well as a dummy variable for gender (WOMAN) and the true value of experience (EXPERIENCE) and its square (EXPSQUARED):

(4) $LNWAGE_i = \alpha_1 + \beta_1 \, YRSCHOOL_i + X_i \gamma_{11} + u_i,$ i=1,...,n,

where the vector X include (WOMAN, EXPERIENCE, EXPSQUARED), all of which are assumed to be exogenous. The results reported at (C) are from an OLS regression of equation (4). The coefficient for years of schooling is estimated to 0.05461. Some researchers suspect, however, that the schooling variable is endogenous in such an equation. A simple equation describing the determination of years of schooling could be given by:

(5) $YRSSCHOOL_i = \alpha_2 + \beta_2 \, LNWAGE_i + X_i \gamma_{21} + Z_i \gamma_{22} + v_i,$ i=1,...,n,

where the vector X again includes all the exogenous variables in equation (4) (WOMAN, EXPERIENCE, EXPSQUARED) and the vector Z contains a set of variables that are predetermined. These could be fathers' eduation, mothers' education, the age of the parents etc. We assume that E(u|X,Z) = E(v|X,Z) = 0 and that the disturbances u and v are homoskedastic.

a) Give the order conditions for identification of the structural parameters in these two equations. Which variables should be included in the first step, reduced form regressions of a two-stage least squares estimation of equation (4) ?

They should be able to show that the order condition is satisfied for equation (4) but not for equation (5). Exact- or over-identification depends on the number of variables in Z. All X and Z should be in the reduced form equation.

Part (D) of the printout gives the two-stage least squares estimator where we have used a set of 12 variables describing the parents' education, age and occupation as our Z-vector.

b) The two-stage least squares estimate of β_1 is 0.07154. This is larger than the OLS estimate of 0.05461. What does this result tell you about the sign of the covariance between YRSCHOOL and u in equation (4)?

From the reduced form of the model, they should be able to establish theoretically that cov(Yrschool, u)) $\neq 0$ through the disturbance term in the reduced form regression of Yrschool(=Yrschoolhat + epsilon). This disturbance is uncorrelated with all the exogenous variables in the model. They should be able to show that the OLS estimator can be decomposed into β + Bias, where the sign of the bias depends on the sign of the covariance between u and Yschool (or epsilon).

c) At Part (E) of the printout you will find results from an OLS regression of equation (4), augmented with the residual from the reduced form regression of YRSCHOOL. The residual is given the name RESIDUALYRSCHOOL.

Based on the results of this equation, which of the two estimators of β_1 would you rely on, OLS or 2SLS? Explain.

We cannot reject the null (exogeneity) and should rely on OLS not to waste information. The exogeneity test is only discussed in Greene, but a simple version, similar to what is here has been discussed in the lectures. They should be able to give an interpretation of the effect of this residual term in terms of the stochastic properties of the endogenous variable (from the reduced form).

Regarding b) and c)

The good students should be able to show formally how the covariance between Yrschool and u enters the expression for the OLS estimator and also how this relationship may be derived from the reduced form regressions. A good discussion in c) builds on these observations.

PRINTOUT PAGE 1.

The MEANS Procedure

Variable	N	**Mean**	**Variance**
error	681	-0.0263547	53.8264442
experror	681	13.5771696	106.3256153

Covariance Matrix, DF = 680

	lnwage	experror
lnwage	0.1222462	0.7412236
experror	0.7412236	106.3256153

PART (A)

OLS estimation

Dependent Variable: lnwage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.51374	3.51374	29.97	<.0001
Error	679	79.61366	0.11725		
Corrected Total	680	83.12740			

Root MSE	0.34242	R-Square	0.0423
Dependent Mean	4.59105	Adj R-Sq	0.0409
Coeff Var	7.45843		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.49640	0.02171	207.16	<.0001
experror	1	0.00697	0.00127	5.47	<.0001

PART (B)

Instrumental variable Estimation

Dependent Variable: lnwage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	12.82006	12.82006	95.63	<.0001
Error	679	91.02443	0.134057		
Corrected Total	680	83.12740			

Root MSE	0.36614	R-Square	0.12345
Dependent Mean	4.59105	Adj R-Sq	0.12216
Coeff Var	7.97503		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.325830	0.030535	141.67	<.0001
experror	1	0.019534	0.001998	9.78	<.0001

PRINTOUT PAGE 2.

PART (C)

OLS estimation

Dependent Variable: lnwage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	72.37514	18.09379	253.06	<.0001
Error	1942	138.85315	0.07150		
Corrected Total	1946	211.22829			

Root MSE	0.26739	R-Square	0.3426
Dependent Mean	4.54814	Adj R-Sq	0.3413
Coeff Var	5.87922		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.67966	0.03639	101.13	<.0001
woman	1	-0.16676	0.01220	-13.67	<.0001
yrsschool	1	0.05461	0.00258	21.13	<.0001
experience	1	0.02528	0.00190	13.32	<.0001
expsquared	1	-0.00038842	0.00004122	-9.42	<.0001

PART (D)

Two-Stage Least Squares Estimation

Dependent Variable: lnwage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	42.72578	10.68144	146.16	<.0001
Error	1942	141.9237	0.073081		
Corrected Total	1946	211.2283			

Root MSE	0.27034	R-Square	0.23139
Dependent Mean	4.54814	Adj R-Sq	0.22981
Coeff Var	5.94387		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.479323	0.152962	22.75	<.0001
yrsschool	1	0.071544	0.012820	5.58	<.0001
woman	1	-0.16103	0.013040	-12.35	<.0001
experience	1	0.023863	0.002187	10.91	<.0001
expsquared	1	-0.00034	0.000053	-6.48	<.0001

PRINTOUT PAGE 3.

PART (E)

Two-Stage Least Squares Estimation

Dependent Variable: lnwage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	72.50819	14.50164	202.91	<.0001
Error	1941	138.72010	0.07147		
Corrected Total	1946	211.22829			

Root MSE	0.26734	R-Square	0.3433
Dependent Mean	4.54814	Adj R-Sq	0.3416
Coeff Var	5.87792		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	3.47932	0.15127	23.00	<.0001
yrsschool		1	0.07154	0.01268	5.64	<.0001
woman		1	-0.16103	0.01290	-12.49	<.0001
experience		1	0.02386	0.00216	11.03	<.0001
expsquared		1	-0.00034400	0.00005252	-6.55	<.0001
residualyrschool	Residual	1	-0.01767	0.01295	-1.36	0.1726