FINAL EXAM - ECONOMETRICS

Answer each question on a different booklet. Duration: 2 hours

1. Consider four random variables (Y, X, Z, W) such that,

$$\mathbb{E}(Y|X, Z, W) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W. \quad (1)$$

We are only interested in estimating β_1, we are not interested in the other parameters β_0, β_2 and β_3. Suppose we have available a random sample $\{Y_i, X_i, Z_i\}_{i=1}^n$ of (Y, X, Z), but we do not observe W. Yet, we know that,

$$\mathbb{E}(W|X, Z) = \gamma_0 + \gamma_1 Z. \quad (2)$$

a. Express (1) using an error term U with conditional mean, given X, Z and W, equal to zero. Do the same with (2) using an error term V with conditional mean, given X and Z, equal to zero. Combine these two equations to express Y as a function of X, Z and an error term $\epsilon = (U + \beta_3 V)$, in a way that the coefficients depend on β_0, β_1, β_2, β_3, γ_0 and γ_1. ¿How much do you expect Y to vary when X increases by one unit and only Z is kept constant?

b. Consider the Ordinary Least Squares (OLS) estimator of the coefficients in a model where Y is the outcome variable while (X, Z) are the explanatory variables. In light of your answer to the previous point a., carefully explain why the OLS coefficient of X would be a biased or unbiased estimator of β_1. Do the same for the OLS coefficient of Z with respect to β_2.

c. Suppose now that W was independent of (X, Z). How would your answer to the previous point b. change?

2. Consider estimating a production function that relates the quantity produced (output), Y, to the amount of labour used, L, as well as to the amount of capital used (machinery and installations), K. A flexible econometric model is described by the translog function,

$$\log Y = \beta_0 + \beta_1 \log L + \beta_2 \log K + \beta_3 \left(\frac{1}{2}\log^2 L\right) + \beta_4 \left(\frac{1}{2}\log^2 K\right) + \beta_5 (\log K \cdot \log L) + U. \quad (3)$$

Such function has been estimated with ordinary least squares using data from 27 states in the US about the primary metal industry (SC33) for the period 1957-1958. Suppose, for the rest of the exercise, that the classical assumptions hold.

A more restrictive production function, which imposes constant elasticity of substitution, is the Cobb-Douglas one,

$$\log Y = \beta_0 + \beta_1 \log L + \beta_2 \log K + U. \quad (4)$$

The following GRETL output tables (models 1-5) show estimates of model (3), (4) and some variations of these specifications. Reported standard errors are robust to heteroskedasticity.

a. Using the translog model (3), provide an expression for the elasticity of Y with respect to K. Calculate the estimated value of such elasticity when the following combination of production factors is used $(K, L) = (318, 1712)$.

b. The Cobb-Douglas specification (4) is nested in the translog one (3) for particular values of the parameters. Test whether the Cobb-Douglas specification in (4) is correct. Write down the null hypothesis of correct specification as well as the alternative one, propose a statistic to run the test, the decision rule, and the outcome of such test.

c. Test that the elasticity of Y with respect to L is equal one in model (4) using two statistics. On one hand, a t statistic and, on the other hand, a test statistic based on the difference between the sum of squares residuals in the restricted and unrestricted models. Discuss the validity of the two tests when the errors are heteroscedastic.

d. Test the hypothesis of constant returns to scale in the Cobb-Douglas specification (4).

Model 1: OLS, using observations 1–27
Dependent variable: $\log Y$

	Coefficient	Std. Error	t-ratio	p-value
const	0.944197	2.91075	0.3244	0.7489
$\log L$	3.61364	1.54807	2.3343	0.0296
$\log K$	-1.89311	1.01626	-1.8628	0.0765
$\left(\frac{1}{2}\log^2 L\right)$	-0.964052	0.707385	-1.3628	0.1874
$\left(\frac{1}{2}\log^2 K\right)$	0.0852947	0.292609	0.2915	0.7735
$(\log K \cdot \log L)$	0.312387	0.438927	0.7117	0.4845

Mean dependent var	7.443631	S.D. dependent var	0.761153
Sum squared resid	0.679927	S.E. of regression	0.179937
R^2	0.954862	Adjusted R^2	0.944114
$F(5, 21)$	88.84734	P-value(F)	2.12e–13

Model 2: OLS, using observations 1–27
Dependent variable: $\log Y$

	Coefficient	Std. Error	t-ratio	p-value
const	1.17064	0.326782	3.5823	0.0015
$\log L$	0.602999	0.125954	4.7875	0.0001
$\log K$	0.375710	0.0853459	4.4022	0.0002

Mean dependent var	7.443631	S.D. dependent var	0.761153
Sum squared resid	0.851634	S.E. of regression	0.188374
R^2	0.943463	Adjusted R^2	0.938751
$F(2, 24)$	200.2489	P-value(F)	1.07e–15

Model 3: OLS, using observations 1–27
Dependent variable: $\log(Y/K)$

	Coefficient	Std. Error	t-ratio	p-value
const	1.17064	0.326782	3.5823	0.0015
$\log L$	0.602999	0.125954	4.7875	0.0001
$\log K$	-0.624290	0.0853459	-7.3148	0.0000

Mean dependent var	-0.002291	S.D. dependent var	0.356159
Sum squared resid	0.851634	S.E. of regression	0.188374
R^2	0.741779	Adjusted R^2	0.720260
$F(2, 24)$	34.47179	P-value(F)	8.79e–08

Model 4: OLS, using observations 1–27
Dependent variable: $\log(Y/K)$

	Coefficient	Std. Error	t-ratio	p-value
const	1.17064	0.326782	3.5823	0.0015
$\log(L/K)$	0.602999	0.125954	4.7875	0.0001
$\log K$	−0.000902242	0.00392627	−0.2298	0.8202

Mean dependent var	−0.002291	S.D. dependent var	0.356159
Sum squared resid	0.851634	S.E. of regression	0.188374
R^2	0.741779	Adjusted R^2	0.720260
$F(2, 24)$	34.47179	P-value(F)	8.79e–08

Model 5: OLS, using observations 1–27
Dependent variable: $\log(Y/L)$

	Coefficient	Std. Error	t-ratio	p-value
const	0.674270	0.333598	2.0212	0..0541
$\log K$	0.135068	0.044442	3.0392	0..0055

Mean dependent var	1.679979	S.D. dependent var	0.251845
Sum squared resid	1.204167	S.E. of regression	0.219469
R^2	0.269790	Adjusted R^2	0.240581
$F(1, 25)$	9.236706	P-value(F)	0.005494

3. In a study of the determinants of women's wages, the following model is considered:

$$\log wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 kidslt6 + \beta_5 kidsge6 + \beta_6 nwifeinc + \beta_7 nkidslt6 + \beta_8 nkidsge6 + \beta_9 (nwifeinc \cdot kidslt6) + \beta_{10} (nwifeinc \cdot kidsge6) + U,$$

where $wage$ is the hourly wage estimated from total income, $educ$ is the number of years of education, $exper$ is the number of years of professional experience. $kidslt6$ is a binary variable equal to one if the worker has children aged less than 6, zero otherwise. $kidsge6$ is a binary variable equal to one if the worker has children aged between 6 and 18, $nkidslt6$ is the number of children aged less than 6, $nkidsge6$ is the number of children aged between 6 and 18. $nwifeinc$ is the total income of the household in dollars, excluding the income of the woman.

Because U is expected to be correlated to $educ$, the number of years of education of the mother ($meduc$) is used as an instrument. Using data from Mroz (1987, *Econometrica* 55, 765-799), we estimate the model by two stages least squares, using $meduc$ as an instrument. We also report the results of the OLS estimates of the reduced form for $educ$. Hereafter, we assume that $Cov(meduc, U) = 0$.

a. Compute the estimated percentage variation in wages ($wage$) when professional experience ($exper$) increases by 1% for a woman with 10 years of experience keeping constant the rest of explanatory variables. Compute the estimated elasticity of wages to $nwifeinc$ for a woman with one child aged 5 and one child aged 10, a value of $nwifeinc$ of 200 dollars per month, 6 years of experience ($exper$) and 15 years of education ($educ$).

b. Test whether the instrument is relevant. Suppose that we want to let the return to education depend on the number of children by including the interaction terms $educ \cdot nkidslt6$ and $educ \cdot nkidsge6$; what instruments can you use to estimate the model with these two new explanatory variables?

c. Suppose that we have information on the years of education of the father ($feduc$), and we believe that $Cov(feduc, U) = 0$. How to test that instruments $meduc$ and $feduc$ are exogenous? Describe the null and alternative hypotheses, the test statistic, the distribution under the null hypothesis, the decision rule and how you will run the test.

Model 6: TSLS, using observations 1–753
Dependent variable: *lwage*
Instrumented: educ. Instruments: *meduc*

	Coefficient	Std. Error	t-ratio	p-value
const	0.0105551	0.0047843	2.2062	0.0136
educ	0.0440428	0.0220666	1.9959	0.0229
exper	0.0437386	0.0139363	3.1385	0.0017
$exper^2$	−0.000845089	0.000411675	−2.0528	0.0401
*kidslt*6	0.0160184	0.283771	0.0564	0.9550
*kidsge*6	−0.0333846	0.0406833	−0.8206	0.4119
nwifeinc	0.0143522	0.00605226	2.3714	0.0177
*nkidslt*6	−0.0128727	0.386999	−0.0333	0.9735
*nkidsge*6	0.193369	0.173542	1.1143	0.2652
$nwifeinc \times kidslt6$	0.000744505	0.00872943	0.0853	0.9320
$nwifeinc \times kidsge6$	−0.00835355	0.00693235	−1.2050	0.2282

Model 7: OLS, using observations 1–753
Dependent variable: *educ*

	Coefficient	Std. Error	t-ratio	p-value
const	7.92157	0.354448	22.3490	0.0000
exper	0.0993556	0.0258172	3.8484	0.0001
$exper^2$	−0.00201766	0.000838280	−2.4069	0.0163
*nkidslt*6	0.582198	0.354473	1.6424	0.1009
*nkidsge*6	−0.0508087	0.0812421	−0.6254	0.5319
nwifeinc	0.0593655	0.0115765	5.1281	0.0000
*kidslt*6	−0.0626910	0.540865	−0.1159	0.9078
*kidsge*6	0.114852	0.369433	0.3109	0.7560
$nwifeinc \times kidslt6$	−0.0106848	0.0146770	−0.7280	0.4668
$nwifeinc \times kidsge6$	−0.0111834	0.0142899	−0.7826	0.4341
meduc	0.274200	0.0216502	12.6650	0.0000

Mean dependent var	12.28685	S.D. dependent var	2.280246
Sum squared resid	2827.628	S.E. of regression	1.952132
R^2	0.276829	Adjusted R^2	0.267083
$F(10, 742)$	28.40364	P-value(F)	2.94e–46

SOME CRITICAL VALUES: $Z_{0.90} = 1.282$, $Z_{0.95} = 1.645$, $Z_{0.975} = 1.96$, $\chi^2_{1,95} = 3.84$, $\chi^2_{1,975} = 5.02$, $\chi^2_{2,95} = 5.99$, $\chi^2_{2,975} = 7.378$, $\chi^2_{3,95} = 7.81$, $\chi^2_{3,975} = 9.3484$, $\chi^2_{4,95} = 9.49$, $\chi^2_{4,975} = 11.1433$, $\chi^2_{5,95} = 11.07$, $\chi^2_{5,975} = 12.8325$, where $\mathbb{P}(Z \leq Z_\alpha) = \alpha$ y $\mathbb{P}\left(\chi^2_m \leq \chi^2_{m,\alpha}\right) = \alpha$, Z is distributed as a normal with mean zero y variance one, and χ^2_m as a chi-square with m degrees of freedom.