EXTRAORDINARY EXAM

Econometrics

Universidad Carlos III de Madrid

23/06/21

Write your name and group in each answer sheet. Answer all the questions in 2:30 hours.

QUESTION 1 (30%)

Consider production data for the year 1994 on 30 US firms in the sector of primary meat industries. For each firm, values are given on production (Y, valued added in millions of dollars), and capital (K, real capital stock in millions of 1987 dollars). A log-linear production function is estimated by OLS with the following result (standard errors assuming homoskedasticity in parenthesis).

$$\ln Y_i = \underset{(0.451)}{0.701} + \underset{(0.091)}{0.756} \ln L_i + \underset{(0.110)}{0.242} \ln K_i + \hat{u}_i, \ RSS = 1.81551, \ R^2 = 0.956888, \quad (1)$$

with RSS denoting sums of squared residuals. There are also estimated by OLS two alternative specifications,

$$\ln Y_i = \underset{(0.358)}{0.010} + \underset{(0.026)}{0.524} \ln (K_i \cdot L_i) + \hat{u}_{1i}, \ RSS = 2.37214, \ R^2 = 0.94367, \quad (2)$$

$$\ln \frac{Y_i}{K_i} = \underset{(0.132)}{0.686} + \underset{(0.089)}{0.756} \ln \frac{L_i}{K_i} + \hat{u}_{2i}, \ RSS = 1.825652, \ R^2 = 0.95665, \quad (3)$$

where RSS is the sum of squares residuals. Critical values of the standard normal Z: $Z_{0.005} = 2.58$, $Z_{0.01} = 2.33$, $Z_{0.025} = 1.96$, $Z_{0.05} = 1.64$, $Z_{0.1} = 1.28$, where $\mathbb{P}(Z > Z_\alpha) = \alpha$. The critical values of the χ^2_q/q for $q = 1, \ldots, 5$ at 5% are $\chi^2_{1,0.05} = 3.84$, $\chi^2_{2,0.05}/2 = 3.00$, $\chi^2_{3,0.05}/3 = 2.60$, $\chi^2_{4,0.05}/4 = 2.37$, $\chi^2_{5,0.05}/5 = 2.21$.

a. (1/3) Test that the output elasticities with respect to capital and labor are identical using the $R^{2\prime}s$ at 5% of significance. Then, show that the test statistic can be expressed in terms of the $RSS's$.

SOLUTION: The hypothesis to be tested is

$$H_0 : \beta_{\ln L} = \beta_{\ln K} \text{ vs } H_1 : \beta_{\ln L} \neq \beta_{\ln K}.$$

The unrestricted model is (**??**) with $R^2_{unrestricted} = 0.956888$, and the restricted model is (**??**) with $R^2_{restricted} = 0.94367$ (note that the log of the product is equal to the sum of the logs). The statistic is

$$F = \frac{R^2_{unrestricted} - R^2_{restricted}}{\left(1 - R^2_{unrestricted}\right)/n} = \frac{0.956888 - 0.94367}{\left(1 - 0.956888\right)/30} = 9.1979$$

This statistic is distributed as a chi-squared with 1 degree of freedom under H_0. The critical value at 5% of significance is $\chi^2_{1,0.05} = 3.84 < 9.1979$. Therefore, we reject H_0. Now, since $1 - R^2 = RSS/TSS$

$$\begin{aligned} F &= \frac{R^2_{unrestricted} - R^2_{restricted}}{\left(1 - R^2_{unrestricted}\right)/n} = \frac{\left(1 - R^2_{restricted}\right) - \left(1 - R^2_{restricted}\right)}{\left(1 - R^2_{unrestricted}\right)/n} \\ &= \frac{\frac{RSS_{restricted}}{TSS} - \frac{RSS_{unrestricted}}{TSS}}{\frac{RSS_{unrestricted}}{TSS}/n} = \frac{RSS_{restricted} - RSS_{unrestricted}}{RSS_{unrestricted}/n} \\ &= \frac{2.37214 - 1.81551}{1.81551/30} = 9.1979. \end{aligned}$$

b. (1/3) Test that the production technology exhibits constant returns to scale. Explain whether or not the test can be carried out either using the $R^{2\prime}s$ or the $RSS's$.

SOLUTION: The hypothesis to be tested is

$$H_0 : \beta_{\ln L} + \beta_{\ln K} = 1 \text{ vs } H_1 : \beta_{\ln L} + \beta_{\ln K} \neq 1.$$

Since the error term is homoskedastic, we can use the F statistic valid only under homoskedasticity. The unrestricted model is (**??**) with $RSS_{unrestricted} = 1.81551$, and the restricted model is (**??**) with $RSS_{restricted} = 1.825652$ (notice that the log of the product is equal to the sum of the logs). The statistic is

$$F = \frac{RSS_{restricted} - RSS_{unrestricted}}{RSS_{unrestricted}/n} = \frac{1.825652 - 1.81551}{1.81551/30} = 0.16759$$

This statistic is distributed as a chi-square with 1 degree of freedom under H_0. The critical value at 5% of significance is $\chi^2_{1,0.05} = 3.84 > 0.16759$. Therefore, we don't reject H_0.

We can not use the F test statistic with R^2 because the dependent variable in the restricted model has changed with respect to the unrestricted one because the restriction is non-homogeneous.

c. (1/3) Discuss how you could obtain a 95% confidence region for $\beta_{\ln L}$ and $\beta_{\ln K}$ (confidence ellipse). What additional information do you need? Briefly comment on whether such a confidence region may assist us in testing

$$H_0 : \beta_{\ln L} = 0.9 \text{ and } \beta_{\ln K} = 0.1 \quad \text{vs} \quad H_1 : \beta_{\ln L} \neq 0.9 \text{ and/or } \beta_{\ln K} \neq 0.1.$$

Use a graph to illustrate your explanations.

SOLUTION: The confidence interval has the form,

$$CR\left(\beta_{\ln L}, \beta_{\ln K}\right) = \left\{(\beta_1, \beta_2) : (\beta_1, \beta_2) \in F(\beta_1, \beta_2) \leq \chi^2_{2,0.05}/2\right\},$$

where

$$F(\beta_1, \beta_2) = \frac{1}{2} \frac{t_1^2(\beta_1) + t_2^2(\beta_2) - 2t_1(\beta_1)\, t_2(\beta_2)\, \hat{\rho}}{1 - \hat{\rho}^2},$$

and

$$t_1(\beta_1) = \frac{\hat{\beta}_{\ln L} - \beta_1}{SE\left(\hat{\beta}_{\ln L}\right)}, \quad t_2(\beta_2) = \frac{\hat{\beta}_{\ln K} - \beta_2}{SE\left(\hat{\beta}_{\ln K}\right)}$$

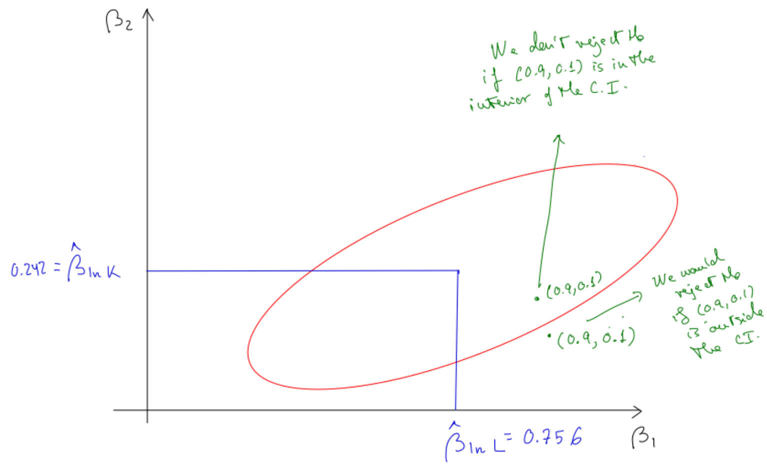
where we need the additional information on the covariance between both OLS estimates (or equivalently, t-statistics),

$$\hat{\rho} = \widehat{Corr}\left(t_1\left(\hat{\beta}_{\ln L}\right), t_2\left(\hat{\beta}_{\ln K}\right)\right) = \frac{\widehat{Cov}\left(\hat{\beta}_{\ln L}, \hat{\beta}_{\ln K}\right)}{SE\left(\hat{\beta}_{\ln L}\right) SE\left(\hat{\beta}_{\ln K}\right)}.$$

Therefore, we reject H_0 if $(0.9, 0.1) \notin CR\left(\beta_{\ln L}, \beta_{\ln K}\right)$. The representation of the confidence interval, and possible situations is represented below.

QUESTION 2. (35%)

A researcher has data for 100 workers in a large organization on hourly earnings ($earn$), skill level of the worker ($skill$), and a measure of the worker's intelligence (IQ). She hypothesizes that the relation between these variables are given by the following two



equations:

$$\begin{aligned} \ln earn &= \beta_0 + \beta_1 skills + u, \\ skills &= \alpha_0 + \alpha_1 IQ + v, \end{aligned} \tag{4}$$

where u and v are disturbance terms uncorrelated with IQ. The researcher is not sure whether u and v are correlated.A researcher has data for 100 workers in a large organization on hourly earnings ($earn$), skill level of the worker ($skill$), and a measure of the worker's intelligence (IQ). She hypothesizes that the relation between these variables are given by the following two equations:

$$\begin{aligned} \ln earn &= \beta_0 + \beta_1 skills + u, \\ skills &= \alpha_0 + \alpha_1 IQ + v, \end{aligned} \tag{5}$$

where u and v are disturbance terms uncorrelated with IQ. The researcher is not sure whether u and v are correlated.

a. (1/6) Justify whether each variable in the two equations is exogenous or endogenous and derive the reduced form equations for the endogenous variables.

SOLUTION: IQ only appears in the skills' equation. The reduced forms for $skills$ and $\ln earn$ are

$$skills = \alpha_0 + \alpha_1 IQ + v,$$

and

$$\begin{aligned}\ln earn &= \beta_0 + \beta_1 (\alpha_0 + \alpha_1 IQ + v) + u, \\ &= (\beta_0 + \beta_1 \alpha_0) + \beta_0 \alpha_1 IQ + (u + \beta_1 v),\end{aligned}$$

respectively. La variable $skills$ is exogenous in the earnings equation when $Cov(u, skills) = Cov(u, v) = 0$. That is, when u and v are correlated, $skills$ is endogenous.

b. (2/6) Demonstrate mathematically under which circumstances the OLS estimator $\hat{\beta}_1$ of β_1 is consistent and under which circumstances is inconsistent.

SOLUTION:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\widehat{Cov}(\ln earn, skills)}{\widehat{Var}(skills)} = \beta_1 + \frac{\widehat{Cov}(u, skills)}{\widehat{Var}(skills)} \\ &= \beta_1 + \alpha_1 \frac{\widehat{Cov}(u, IQ)}{\widehat{Var}(skills)} + \frac{\widehat{Cov}(u, v)}{\widehat{Var}(skills)}.\end{aligned}$$

Now, by the LLN

$$\begin{aligned}\widehat{Cov}(u, IQ) &\rightarrow Cov(u, IQ) = 0 \ w.p.1 \\ \widehat{Cov}(u, v) &\rightarrow Cov(u, v) \ \ w.p.1 \\ \widehat{Var}(skills) &\rightarrow Var(skills) > 0 \ w.p.1\end{aligned}$$

and

$$\hat{\beta}_1 \rightarrow \beta_1 + \frac{Cov(u, v)}{Var(skills)} \ w.p.1$$

That is, $\hat{\beta}_1$ is consistent when u and v are not correlated (so that $skills$ is endogenous), and is inconsistent when they are.

c. (2/6) Demonstrate mathematically how the researcher could use instrumental variables (IV) estimation to estimate consistently β_1.

SOLUTION: We could use IQ as instrument, and estimate the parameters of the reduced form of $skill$ by OLS. Then, I would test that the instrument is relevant by testing that $\alpha_1 = 0$ using a t-ratio test. If we don't reject the hypothesis, then compute the fitted values of $skill$, i.e. $\widehat{skills}_i = \hat{\alpha}_0 + \hat{\alpha}_1 IQ_i$, $i = 1, ..., n$ in a first step. In a second step, estimate the structural equation of $\ln earn$ by OLS after plugging in $\widehat{skills}_i$ in equation (4). The resulting β_1 estimate is the instrumental variable estimator. Then,

$$\widehat{\beta}_1^{IV} = \frac{\widehat{Cov}\left(\ln earn, \widehat{skills}\right)}{\widehat{Var}\left(\widehat{skills}\right)} = \frac{\widehat{Cov}\left(\ln earn, IQ\right)}{\widehat{Cov}\left(skills, IQ\right)}$$

By the LLN, and because $Cov\left(skills, IQ\right) \neq 0$,

$$\frac{\widehat{Cov}\left(\ln earn, IQ\right)}{\widehat{Cov}\left(skills, IQ\right)} \to \frac{Cov\left(\ln earn, IQ\right)}{Cov\left(skills, IQ\right)} = \beta_1 \text{ wp1},$$

since

$$\begin{aligned} Cov\left(u, IQ\right) = 0 &\implies Cov\left(\ln earn - \beta_0 - \beta_1 skills, IQ\right) = 0 \\ &\implies Cov\left(\ln earn, IQ\right) - \beta_1 Cov\left(skills, IQ\right) = 0 \\ &\implies \beta_1 = \frac{Cov\left(\ln earn, IQ\right)}{Cov\left(skills, IQ\right)}. \end{aligned}$$

d. (1/6) Explain the advantages and disadvantages of using IV rather than OLS, to estimate β_1 when there is no certainty on the consistency of $\hat{\beta}_1$.

SOLUTION:

	Advantages	Disadvantages
OLS	If $Cov\left(u, v\right) = 0$, OLS is consistent and with smaller variance than IV	If $Cov\left(u, v\right) \neq 0$, OLS is inconsistent.
IV	If $Cov\left(u, v\right) \neq 0$, OLS is inconsistent, but IV is consistent.	If $Cov\left(u, v\right) = 0$, OLS is consistent and with smaller variance than IV

QUESTION 3. (35%)

Our goal is to estimate the causal relationship between house prices and pollution. For this, we have a sample of 506 neighborhoods in the Boston area (USA). We estimate a model that relates the median dollar price of houses in each neighborhood ($price$) with the amount of nitrogen oxide in the area, measured in parts per 100 million (nox), controlling for $dist$: the weighted distance from the neighborhood to the five main employment centers, in miles, by $rooms$: the average number of rooms in the houses in the neighborhood, by $crime$: the number of crimes committed per capita (calculated as the number of crimes divided by the number of inhabitants multiplied by 100000), and by $stratio$: the average of the ratio of students per teacher in the neighborhood schools. The population model is

$$\begin{aligned}\ln(price) &= \beta_0 + \beta_1 \ln(nox) + \beta_2 \ln^2(nox) + \beta_3 dist + \beta_4 dist^2 + \beta_5 dist \cdot \ln(nox) \\ &\quad + \beta_6 rooms + \beta_7 stratio + \beta_8 crime + \beta_9 crime \cdot \ln(nox) + u,\end{aligned}$$

where the error u has zero mean, conditional to the explanatory variables considered, and the conditional variance can be a function of the explanatory variables. GRETL output with the OLS estimation of this model with the variance and covariance matrix of the estimated coefficients, as well as the estimation of a transformation, is at the end of the exam. Use the critical values in question 1.

a. **(1/3)** Provide a 95% confidence interval for the $price$ elasticity with respect to nox, for $nox = 5$, $dist = 4$, and $crime = 0.5$.

SOLUTION: The elasticity is

$$\theta = \xi_{price,nox}\Big\rfloor_{\substack{crime=0.5\\ dist=4\\ nox=5}} = \frac{d\ln(price)}{d\ln(nox)}\Bigg\rfloor_{\substack{crime=0.5\\ dist=4\\ nox=5}} = \beta_1 + 2\cdot\beta_2\cdot\ln 5 + \beta_5\cdot 4 + \beta_9\cdot 0.5.$$

Therefore,

$$\beta_1 = \theta - 2\cdot\beta_2\cdot\ln 5 - \beta_5\cdot 4 - \beta_9\cdot 0.5,$$

and substituting in the model

$$\begin{aligned}\ln(price) &= \beta_0 + (\theta - 2\cdot\beta_2\cdot\ln 5 - \beta_5\cdot 4 - \beta_9\cdot 0.5)\ln(nox) + \beta_2\ln^2(nox)\\ &\quad +\beta_3 dist + \beta_4 dist^2 + \beta_5 dist\cdot\ln(nox) + \beta_6 rooms\\ &\quad +\beta_7 stratio + \beta_8 crime + \beta_9 crime\cdot\ln(nox) + u\\ &= \beta_0 + \theta\ln(nox) + \beta_2\ln(nox)(\ln(nox) - 2\cdot\ln(5)) + \beta_3 dist + \beta_4 dist^2\\ &\quad +\beta_5\ln(nox)(dist-4) + \beta_6 rooms + \beta_7 stratio + \beta_8 crime\\ &\quad +\beta_9\ln(nox)(crime - 0.5) + u\end{aligned}$$

Therefore, looking at model 2, the elasticity θ is the coefficient of $\ln(nox)$. If we consider the output of model 2, we see that $\hat{\theta} = -0.952759$ and $SE\left(\hat{\theta}\right) = 0.120556$ and the confidence interval is $\hat{\theta} \pm 1.96\cdot SE\left(\hat{\theta}\right)$. Therefore, the confidence interval is $-0.952759 \pm 1.96\cdot 0.120556 = [-1.189, -0.716\,47]$.

The same result for $\hat{\theta}$ can be obtained with the output of Model 1 and the table of estimates of the covariances of coefficients to calculate $SE\left(\hat{\theta}\right)$.

b. (1/3) Which is the estimated *dist* value such that the relation between *price* and *dist* changes its sign when $nox = 5$?

SOLUTION:

$$\left.\frac{\partial}{\partial dist}\ln(price)\right|_{nox=5} = \beta_3 + 2\cdot\beta_4\cdot dist + \beta_5\cdot\ln(5)$$

Thus, the required value of nox satisfies

$$\beta_3 + 2\cdot\beta_4\cdot dist + \beta_5\cdot\ln(5) = 0 \Longrightarrow dist^* = -\frac{\beta_3 + \ln(5)\cdot\beta_5}{2\cdot\beta_4}.$$

Therefore, the estimated value is

$$\begin{aligned}\widehat{dist}^* &= -\frac{\hat{\beta}_3 + \ln(5)\cdot\hat{\beta}_5}{2\cdot\hat{\beta}_4}\\ &= -\frac{-0.813055 + 0.382561\cdot\ln(5)}{2\cdot 0.0168214}\\ &= 5.8661 \text{ miles.}\end{aligned}$$

c. (1/3) Obtain an estimator of the *price* elasticity with respect to *crime* for $nox = 5$ and $crime = 0.5$. Then, test at the 1% of significance whether this elasticity is different from zero.

SOLUTION: The *price* elasticity with respect to *crime* is

$$\eta = \xi_{price,nox} = \frac{d \ln (price)}{dcrime} \cdot crime = [\beta_8 + \beta_9 \ln (nox)]\, crime,$$

therefore, the estimator is,

$$\begin{aligned} \hat{\eta} &= \hat{\xi}_{price,nox}\Big]_{\substack{nox=5 \\ crime=0.5}} = \left[\hat{\beta}_8 + \hat{\beta}_9 \ln (5)\right] \cdot 0.5 \\ &= [0.202440 + (-0.113157) \ln (5)]\, 0.5 = 0.010160, \end{aligned}$$

$$\begin{aligned} SE(\hat{\eta}) &= 0.5\sqrt{\widehat{Var}\left(\hat{\beta}_8\right) + \widehat{Var}\left(\hat{\beta}_9\right) \ln^2 (5) + 2\widehat{Cov}\left(\hat{\beta}_8, \hat{\beta}_9\right) \ln (5)} \\ &= 0.5\sqrt{0.0018569 + 0.00051555 \ln^2 (5) + 2\,(-0.00096184) \ln (5)} \\ &= 0.0049053 \end{aligned}$$

Thus, the $t - ratio$ is

$$t = \frac{0.010160}{0.0049053} = 2.071\,2,$$

which is not significative at 1% comparing with the two-sided critical value from the $N(0,1)$ distribution equal to $Z_{0.005} = 2.58$.

Model 1: OLS, using observations 1–506

Dependent variable: lprice

Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	18.5445	2.37133	7.8203	0.0000
ln(nox)	−8.35034	2.37062	−3.5224	0.0005
ln^2(nox)	1.84037	0.581296	3.1660	0.0016
dist	−0.813055	0.222751	−3.6501	0.0003
dist^2	0.0168214	0.00346286	4.8577	0.0000
dist·lnox	0.382561	0.125710	3.0432	0.0025
rooms	0.242263	0.0236869	10.2277	0.0000
stratio	−0.0461280	0.00490355	−9.4070	0.0000
crime	0.202440	0.0430921	4.6978	0.0000
crime·lnox	−0.113157	0.0227057	−4.9836	0.0000

R^2	0.679808	Adjusted R^2	0.673998
$F(9, 496)$	97.99094	P-value(F)	4.0e–104

Coefficient covariance matrix

const	ln(nox)	ln^2(nox)	dist	dist^2	
5.6232	−5.5742	1.3521	−0.50367	0.0064633	const
	5.6199	−1.3739	0.49179	−0.0060575	ln(nox)
		0.33791	−0.11702	0.0014084	ln^2(nox)
			0.049618	−0.00066546	dist
				1.1991e–05	dist^2

dist·ln(nox)	rooms	stratio	crime	crime·ln(nox)	
0.28267	−2.0012e–05	−0.0033857	0.013530	−0.0061238	const
−0.27820	−0.0053622	0.0023673	−0.015960	0.0074156	ln(nox)
0.066446	0.0015378	−0.00045917	0.0045256	−0.0021542	ln^2(nox)
−0.027784	−0.00028577	0.00042420	−0.00055342	0.00018813	dist
0.00034633	−2.6614e–06	−5.6973e–06	1.1637e–05	−5.1086e–06	dist^2
0.015803	0.00023215	−0.00024123	0.00026781	−7.9769e–05	dis·ln(nox)
	0.00056107	4.1068e–05	5.1632e–05	−2.2403e–05	rooms
		2.4045e–05	1.5567e–05	−9.9102e–06	stratio
			0.0018569	−0.00097728	crime
				0.00051555	crime·ln(nox)

Model 2: OLS, using observations 1–506

Dependent variable: lprice

Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	18.5445	2.37133	7.8203	0.0000
ln(nox)	−0.952759	0.120556	−7.9030	0.0000
dist	−0.813055	0.222751	−3.6501	0.0003
dist^2	0.0168214	0.00346286	4.8577	0.0000
rooms	0.242263	0.0236869	10.2277	0.0000
stratio	−0.0461280	0.00490355	−9.4070	0.0000
crime	0.202440	0.0430921	4.6978	0.0000
$\ln(\text{nox})(\ln(\text{nox}) - 2 \cdot \ln(5))$	1.84037	0.581296	3.1660	0.0016
$\ln(\text{nox})(\text{dist} - 4)$	0.382561	0.125710	3.0432	0.0025
$\ln(\text{nox})(\text{crime} - 0.5)$	−0.113157	0.0227057	−4.9836	0.0000

R^2	0.679808	Adjusted R^2	0.673998
$F(9, 496)$	97.99094	P-value(F)	4.0e–104