Econometrics
Universidad Carlos III de Madrid
Extraordinary Exam
June 25, 2014

Instructions for the exam:

There are 2 hours and a half to answer the exam
The exam contains 10 questions referred to 2 exercises, with an overall grading of 10 points.
Use separated answer booklets for each exercise. Write your name clearly in all of them
The total grade of each exercise is indicated at the beginning of each exercise.
In case a significance level is not indicated, use a 5% one.
In case you are asked to test an hypothesis, the following items will be evaluated:

- Clear definition of the null and alternative hypotheses
- Clear definition of the test statistic
- Clear definition of the significance level and critical value. Statistical tables are attached below.
- Conclusion and interpretation of the rejection rule

CRITICAL VALUES:

$N(0,1)$
$\Pr(N(0,1) > 2,576) = 0,005$
$\Pr(N(0,1) > 2,326) = 0,01$
$\Pr(N(0,1) > 1,960) = 0,025$
$\Pr(N(0,1) > 1,645) = 0,05$
$\Pr(N(0,1) > 1,282) = 0,10$

$$c : \Pr\left(\chi^2_q > c\right) = \alpha$$

c	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$
$\alpha = 0,01$	$6,63$	$9,21$	$11,34$	$13,28$	15.09	16.81
$\alpha = 0,05$	$3,84$	$5,99$	$7,81$	$9,49$	$11,07$	$12,59$
$\alpha = 0,10$	$2,71$	$4,61$	$6,25$	$7,78$	$9,24$	$10,65$

Recall that a Student t with n degrees of freedom behaves as a $N(0,1)$ for n reasonably large $(n > 30)$. On the other hand, a Fisher F with q degrees of freedom in the numerator and n degrees of freedom in the denominator, behaves as a $\chi^2_{(q)}/q$ for n large.

1. [7 points/over 10] We are studying the factors behind the Body Mass Index (BMI), defined as the weight (in kilograms) divided by the squared of height (in meters). For a random sample we have estimated the following model (Model 1, Output 1),

$$\widehat{BMI} = \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_2 * drinks^2 + \widehat{\beta}_3 * female \qquad \text{(Model 1)}$$

where

$$\begin{aligned} drinks &= \text{number of days during the last year} \\ &\quad \text{in which the individual has drunk 5 or more glasses of alcohol.} \\ drinks^2 &= \text{squared of } drinks. \\ female &= \text{is a dummy variable that takes a value one} \\ &\quad \text{for women and zero otherwise.} \end{aligned}$$

a) Using Model 1, What is the marginal impact of *drinks* on expected BMI? Is this effect constant? For which values of *drinks* is this effect positive (negative)? Explain. **[1 point/over 10]**

Answer:

$$\text{Marginal Effect} = \frac{\partial B\hat{M}I}{\partial drinks} = \hat{\beta}_1 + 2 * \hat{\beta}_2 drinks.$$

Since the marginal effect depends on *drinks* (and $\hat{\beta}_2$ is significative at the 5% level with the corresponding t test) we can say that the impact is not constant.

The value for which the marginal impact of *drinks* changes sign is $drinks^* = -\frac{\hat{\beta}_1}{2 * \hat{\beta}_2} = 47.5$.

Since $\hat{\beta}_1 < 0$ and $\left|\hat{\beta}_1\right| > \left|\hat{\beta}_2\right|$ for values of drinks below $drinks^*$ an increase in the consumption of alcohol is associated with a reduction in the BMI. Nevertheless, an increase in the consumption of alcohol for values of drinks over $drinks^*$ is associated with an increase in BMI.

b) Using Model 1, what is the predicted difference in BMI betwen a man with $drinks = 2$ and a woman with $drinks = 6$? **[1 point/over 10]**

Answer:

$$\begin{aligned} E[BMI|female &= 0, drinks = 2] - E[BMI|female = 1, drinks = 6] \\ &= \widehat{\beta_1} * (2-6) + \widehat{\beta}_2 * (4-36) - \widehat{\beta}_3 = -\widehat{\beta_1} * 4 - \widehat{\beta}_2 * 32 - \widehat{\beta}_3 \\ &= 0.0095 * 4 - 32 * 0.0001 + 1.1418 = 1.1766 \end{aligned}$$

Alternatively we have estimated the following model (Model 2, Output 2)

$$\begin{aligned} \widehat{BMI} &= \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_2 * drinks^2 \\ &\quad + \widehat{\beta}_3 * female + \widehat{\beta}_4 * drinks * female + \widehat{\beta}_5 * drinks^2 * female \end{aligned} \qquad \text{(Model 2)}$$

c) Using Model 2, What is the predicted difference in expected BMI between a man with $drinks = 2$ and a woman with $drinks = 6$? **[1 point/over 10]**

$$\begin{aligned} & E[BMI|female = 0, drinks = 2] - E[BMI|female = 1, drinks = 6] \\ = & \widehat{\beta_1} * (2-6) + \widehat{\beta}_2 * (4-36) - \widehat{\beta}_3 - \hat{\beta}_4 * 6 - \hat{\beta}_5 * 36 = -\widehat{\beta_1} * 4 - \widehat{\beta}_2 * 32 - \widehat{\beta}_3 - \hat{\beta}_4 * 6 - \hat{\beta}_5 * 36 \\ = & -0.0424 * 4 + 32 * 0.0004 + 0.8753 + 6 * 0.1639 - 36 * 0.0016 = 1.6443 \end{aligned}$$

d) Using Model 2, what is the impact of a marginal change in *drinks* on BMI for a man? What is the impact of a marginal change in *drinks* on BMI for a woman? **[1 point/over 10]**

$$\begin{aligned}
\text{Marginal Effect} &= \frac{\partial B\hat{M}I}{\partial drinks} = (\hat{\beta}_1 + \hat{\beta}_4 * female) + 2 * drinks * (\hat{\beta}_2 + \hat{\beta}_5 * female) \\
\text{M. Effect for a man} &= \frac{\partial B\hat{M}I}{\partial drinks} = \hat{\beta}_1 + 2 * \hat{\beta}_2 drinks = 0.0424 - 2 * 0.0004 * drinks = 0.0424 - 0.0008 * \\
\text{M. Effect for a woman} &= \frac{\partial B\hat{M}I}{\partial drinks} = (\hat{\beta}_1 + \hat{\beta}_4) + 2 * drinks * (\hat{\beta}_2 + \hat{\beta}_5) = -0.1215 + 0.0024 * drinks
\end{aligned}$$

e) Using Model 2 as benchmark (unrestricted model), explain and test that the effect of a marginal change in *drinks* on BMI is linear for men. **[1 point/over 10]**

$$\begin{aligned}
H_0 &: \beta_2 = 0 \\
H_1 &: \beta_2 \neq 0
\end{aligned}$$

We construct the t-statistic

$$t = \frac{\hat{\beta}_2 - 0}{se(\hat{\beta}_2)} = \frac{-0.0004}{0.00006} = -6.6.$$

Since $|t| > 1.96$, we can reject H_0 at the 5% significance level. This means that we can reject that the marginal effect is constant and we confirm is linear for men.

f) Using Model 2 as benchmark (unrestricted model), explain and test that the effect of a marginal change in *drinks* on BMI is linear for women. **[1 point/over 10]**

Answer:

$$\begin{aligned}
H_0 &: \beta_2 + \beta_5 = 0 \\
H_1 &: \beta_2 + \beta_5 \neq 0
\end{aligned}$$

Since the variance and covariance matrix is not reported, we need to estimate the restricted and unrestricted model and calculate an F test under the assumption of homoskedasticity. The restricted model when accepting H_0 as true can expressed as

$$\begin{aligned}
\widehat{BMI} &= \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_2 * drinks^2 + \widehat{\beta}_3 * female + \widehat{\beta}_4 * drinks * female - \hat{\beta}_2 * drinks^2 * female \\
&= \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_2 * drinksh^2 * (1 - female) + \widehat{\beta}_3 * female + \widehat{\beta}_4 * drinks * female \\
&= \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_2 * drinks^2 * male + \widehat{\beta}_3 * female + \widehat{\beta}_4 * drinks * female
\end{aligned}$$

From the list of outputs that are reported, we can observe that the restricted model corresponds to output 5. Because the restricted and unrestricted (model 2, output 2) models have the same dependent variable, we can construct our F statistic using the R^2:

$$F = \frac{R^2_{unrestricted} - R^2_{restricted}}{(1 - R^2_{unrestricted})} * \frac{n}{q} = \frac{(0.0151 - 0.0143)}{(1 - 0.0151)} * \frac{233239}{1} = 189.45.$$

Since the $F-statistic$ follows a χ^2_1 asymptotic distribution under the null we get that the critical value is 3.8 when using a significance level of 5%. Therefore we can reject H_0 which means that the impact of *drinks* is not constant and we confirm is linear for women.

g) Using Model 2 as benchmark (unrestricted model), explain and test that the effect of a marginal change in *drinks* on BMI is linear for an individual whatever the gender. **[1 point/over 10]**

$$\begin{aligned} H_0 &: \beta_2 = \beta_5 = 0 \\ H_1 &: H_0 \text{ is false} \end{aligned}$$

Restricted Model:

$$\widehat{BMI} = \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_3 * female + \widehat{\beta}_4 * drinks * female,$$

which corresponds to output 4. Because the restricted and unrestricted (model 2, output 2) models have the same dependent variable, we can construct our F statistic using the R^2:

$$F = \frac{R^2_{unrestricted} - R^2_{restricted}}{(1 - R^2_{unrestricted})} * \frac{n}{q} = \frac{(0.0151 - 0.0017)}{(1 - 0.0151)} * \frac{233239}{2} = 1586.7.$$

Since the $F - statistic$ follows a $\chi^2_2/2$ asymptotic distribution under the null we get that the critical value is 5.99/2≈ 3 when using a significance level of 5%. Therefore we can reject H_0 which means that the impact of drinks is not constant in general and is linear at least for some type of individuals.

OUTPUT 1: OLS, using observations 1–233239
Dependent variable: bmi

	Coeff	S.E	t	p-value
const	26.8065	0.0182106	1472.0235	0.0000
drinks	–0.00953349	0.00484486	-1.9678	0.0491
drinks2	0.000102875	4.98842e–005	2.0623	0.0392
female	–1.14183	0.020171		

Mean Dependent Variable.	26.17626
Sum of the Squared Residuals	5196636
R^2	0.014139

OUTPUT 2: OLS, using observations 1–233239
Dependent variable: bmi

	Coeff	S.E	t	p-value
const	26.6876	0.0197415	1351.8542	0.0000
drinks	0.0424549	0.00586037	7.2444	0.0000
drinks2	−0.000419308	6.03490e–005	−6.9481	0.0000
female	−0.875380	0.0264042	–33.1531	0.0000
drinks*female	–0.163936	0.0104052	–15.7552	0.0000
drinks2*female	0.00165051	0.000107120		

Mean Dependent Variable.	26.17626
Sum of the Squared Residuals	5191109
R^2	0.015188

OUTPUT 3: OLS, using observations 1–233239
Dependent variable: bmi

	Coeff	S.E	t	p-value
const	26.0776	0.0129662	2011.2017	0.0000
drinks	0.0543016	0.00474404	11.4463	0.0000
drinks2	–0.000511294	4.90232e–005		

Mean Dependent Variable.	26.17626
Sum of the Squared Residuals	5268033
R^2	0.000594

OUTPUT 4: OLS, using observations 1–233239

Dependent variable: bmi

	Coeff	S.E.	t	p-value
const	26.1650	0.0102729	2546.9867	0.0000
drinks	0.0214939	0.00139848	15.3695	0.0000
drinks*female	–0.0445572	0.00232373	–19.1749	0.0000
female	–1.113876	0.0205		

Mean Dependent Variable.	26.17626
Sum of the Squared Residuals	5262194
R^2	0.001702

OUTPUT 5: OLS, using observations 1–233239

Dependent variable: bmi

	Coeff	S.E.	t	p-value
const	26.6876	0.0197496	1351.2966	0.0000
drink	0.0424549	0.00586279	7.2414	0.0000
female	−1.02614	0.0240875	–42.6007	0.0000
drinks*female	–0.0473893	0.00617347	–7.6763	0.0000
drinks2*male	–0.000419308	6.03739e–005		

Mean Dependent Variable.	26.17626	Mean Dependent Variable.	26.17626
Sum of the Squared Residuals	5195416	Sum of the Squared Residuals	5195416
R^2	0.014370	R^2	0.014370

NOTE: *male* is a dummy variable that takes a value one when the person is a man and zero otherwise.

2. **[3 points/over 10]** The following model is a system of simultaneous equations to study whether the openness of the economy ($open$) leads to lower inflation rates (inf),

$$\begin{aligned} inf &= \delta_{10} + \gamma_{12} open + \delta_{11} \log(pcinc) + u_1 \\ open &= \delta_{20} + \gamma_{21} inf + \delta_{21} \log(pcinc) + \delta_{22} \log(land) + u_2. \end{aligned}$$

We assume that (the logarithms of) *pcinc* (per capita income) and *land* (land for farming) are exogenous in the whole exercise.

The following estimations have been obtained by OLS and 2SLS.

Output 1: OLS estimation using the 114 observations 1–114
Dependent variable: inf

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	25,1040	15,2052	1,6510	0,1016
open	-0,215070	0,0946289	-2,2728	0,0250
lpcinc	0,0175673	1,97527	0,0089	0,9929

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	62127,5
Residual standard deviation ($\hat{\sigma}$)	23,6581
R^2	0,0452708
$\bar{R}^2$ corrected	0,0280685
$F(2,111)$	2,63167
p-value for $F()$	0,0764453

Output 2: OLS estimation using the 114 observations 1–114
Dependent variable: open

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	116,226	15,8808	7,3187	0,0000
inf	-0,0680353	0,0715556	-0,9508	0,3438
lpcinc	0,559501	1,49395	0,3745	0,7087
lland	-7,3933	0,834814	-8,8563	0,0000

Mean of dependent variable	37,0789
Std. dev. of dependent variable	23,7535
Residual sum of squares	34865,3
Residual standard deviation ($\hat{\sigma}$)	17,8033
R^2	0,453162
$\bar{R}^2$ corrected	0,438249
$F(3,110)$	30,3855
p-value for $F()$	$< 0,00001$

Output 3: OLS estimation using the 114 observations 1–114
Dependent variable: inf

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	-12,615	21,0313	-0,5998	0,5498
lpcinc	0,191394	1,98158	0,0966	0,9232
lland	2,55380	1,08049	2,3635	0,0198

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	61903,2
Residual standard deviation ($\hat{\sigma}$)	23,6154
R^2	0,0487174
$\bar{R}^2$ corrected	0,0315772
$F(2,111)$	2,84229
p-value for $F()$	0,0625432

Output 4: OLS estimation using the 114 observations 1–114
Dependent variable: open

Variable	Coefficient	Standard dev.	t statistic	p-value
const	117,085	15,8483	7,3878	0,0000
lpcinc	0,546479	1,49324	0,3660	0,7151
lland	-7,5671	0,814216	-9,2937	0,0000

Mean of dependent variable	37,0789
Std. dev. of dependent variable	23,7535
Residual sum of squares	35151,8
Residual standard deviation ($\hat{\sigma}$)	17,7956
R^2	0,448668
$\bar{R}^2$ corrected	0,438734
$F(2,111)$	45,1654
p-value for $F()$	<0,00001

Output 5: 2SLS estimation using the 114 observations 1–114
Dependent variable: inf
Instruments: lland

Variable	Coefficient	Standard dev.	t statistic	p-value
const	26,8993	15,4012	1,7466	0,0807
open	-0,337487	0,144121	-2,3417	0,0192
lpcinc	0,375823	2,01508	0,1865	0,8520

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	63064,2
Residual standard deviation ($\hat{\sigma}$)	23,8358
$F(2,111)$	2,62498
p-value for $F()$	0,0769352

Hausman Test –
Null hypothesis: OLS estimates are consistent
Asymptotic test statistic: $\chi^2_1 = 1{,}35333$
with p-value = 0,244697

a) Discuss the possible identification of each equation of the system, the weakness of the available instruments and perform the correspondent hypothesis tests whenever is possible. **[1 point/over 10]**

Answer: the second equation is not identified because there are not available instruments for the endogenous regressor *inf*. The first equation can be (just) identified using $\log(land)$ as an instrument for *open*. Looking at output 4 (reduced form for *open*), it can be checked that the instrument $\log(land)$ is significant with $t = -9,2937$ and the corresponding F statistic is $F = t^2 = -9,2937 \approx 86.4$ which shows that the instrument is not weak.

b) Explain how you would perform a test of the exogeneity of the instruments used in the two-stage estimation for a equation and whether it is posible to apply it for the equations of the given system. **[1 point/over 10]**

Answer: the exogeneity test would be performed using the residuals of each structural equation after parameter estimation by 2SLS, fitting a regression with them as dependent variable over all exogenous variables available, and then checking the joint significance of only the instruments omitted in the structural equation. This procedure can only be used when there is overidentification: more available instruments that endogenous regressors in the equation.

In the system given this not the case, and it is not possible to check the exogeneity of $\log(land)$ in the first equation because it is exactly identified nor in the second one because it is not identified since there are not available instruments.

c) Test whether the effect of *open* over inf is lower than -0.2. If *open* were not a determinant of inf, (but inf is a determinant of *open*), explain the properties of the estimates of Output 1. **[1 point/over 10]**

Answer: the test required is

$$\begin{aligned} H_0 &: \gamma_{12} = -0.2 \\ H_1 &: \gamma_{12} < -0.2 \end{aligned}$$

and we use a t test

$$t = \frac{\hat{\gamma}_{12} - (-0.2)}{se(\hat{\gamma}_{12})} = \frac{-0.3375 + 0.2}{0.1441} = -0.9542$$

using Output 5 (2SLS), which is not significant with respect to the 5% one-sided critival value, -1.65, so we can not confirm that the effect is lower than -0.2.

In this case, $\gamma_{12} = 0$ and $\gamma_{21} \neq 0$, *open* would still be endogenous in the first equation, since substituting the first equation in the second one we obtain

$$\begin{aligned} open &= \delta_{20} + \gamma_{21}\{\delta_{10} + \gamma_{12} open + \delta_{11}\log(pcinc) + u_1\} + \delta_{21}\log(pcinc) + \delta_{22}\log(land) + u_2 \\ &= \delta_{20} + \gamma_{21}\{\delta_{10} + \delta_{11}\log(pcinc) + u_1\} + \delta_{21}\log(pcinc) + \delta_{22}\log(land) + u_2 \\ &= \delta_{20} + \gamma_{21}\delta_{10} + \{\gamma_{21}\delta_{11} + \delta_{21}\}\log(pcinc) + \delta_{22}\log(land) + u_2 + \gamma_{21}u_1 \end{aligned}$$

and we can check that *open* is still correlated with u_1 because $\gamma_{21} \neq 0$ and therefore is endogenous in the first equation and the corresponding OLS estimates of Output 1 are inconsistent.