# ECONOMETRICS RETAKE EXAM

Universidad Carlos III de Madrid

June 2016

Answer the 4 questions in 2.5 hours.

**1.** You are conducting an econometric investigation into the prices of recently sold houses in a single urban area. The sample data consists of 88 observations on the following variables.

$$
\begin{aligned}
P_i &= \text{the selling price of house } i\text{-th in thousands of dollars.} \\
HS_i &= \text{The house size of house } i\text{-th in hundred on squared feet.} \\
YS_i &= \text{The yard size of house } i\text{-th in hundred on squared feet.} \\
DC_i &= \begin{cases} 1 \text{ if house } i\text{-th is a colonial-stile house} \\ 0 \text{ otherwise} \end{cases}
\end{aligned}
$$

The linear model proposed is

$$
\begin{aligned}
\ln P_i &= \beta_0 + \beta_1 \ln HS_i + \beta_2 \ln YS_i + \beta_3 (\ln HS_i)^2 + \beta_4 (\ln YS_i)^2 + \beta_5 (\ln HS_i)(\ln YS_i) \\
&\quad + \beta_6 DC_i + U_i,
\end{aligned}
$$

where the $\beta_j$'s are regression coefficients, and $U_i$ is an error term, which is assumed to satisfy the classical assumptions. The following table shows the estimation of the model under alternative restrictions on the coefficients (standard errors in parenthesis)

| Explanatory Var. | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Intercept* | 9.321<br>(2.500) | 2.767<br>(0.593) | 8.456<br>(2.133) | 2.639<br>(0.244) |
| $\ln HS_i$ | −3.029<br>(1.398) | 0.748<br>(0.081) | −3.045<br>(1.385) | 0.750<br>(0.081) |
| $\ln YS_i$ | −0.2492<br>(0.583) | 0.112<br>(0.238) | 0.147<br>(0.037) | 0.168<br>(0.038) |
| $(\ln HS_i)^2$ | 0.5509<br>(0.254) | * * ** | 0.622<br>(0.227) | * * ** |
| $(\ln YS_i)^2$ | 0.0128<br>(0.027) | 0.006<br>(0.026) | * * ** | * * ** |
| $(\ln HS_i)(\ln YS_i)$ | 0.0920<br>(0.143) | * * ** | * * ** | * * ** |
| $DC_i$ | 0.0946<br>(0.043) | 0.067<br>(0.043) | 0.096<br>(0.043) | 0.066<br>(0.0428) |
| $RSS =$ | 2.5918 | 2.8413 | 2.6067 | 2.8432 |
| $TSS =$ | 8.0176 | 8.0176 | 8.0176 | 8.0176 |

Note: The symbol "*****" means that the corresponding explanatory variable has been omitted.

All the hypothesis tests and confidence intervals are at the 5% significance level and the 95% confidence level. You must state clearly the null and alternative hypothesis, the test statistic, the critical region (rejection rule) and the conclussion of the test.

**a.** Compare the goodness-of-fit of the four fitted equations penalizing for the number of explanatory variables included. Which of the four regression equations provides the best fit? Which provides the worst fit?

**b.** What is the estimated effect on the house price of an increase of yard size of 200 square feets in colonial-style houses with 500 square feet house size and 1000 square feet yard. Then test whether the size of the yard contributes significantly to this effect.

**c.** Test that price elasticity with respect to $HS$ is constant.

**d.** Test jointly that both, the price elasticity with respect to $HS$ and with respect to $YS$ are constant.

**2.** The government of a developing country wants to implement a program where poor families receive food stamps that can be used to purchase prepackaged foods with high nutritional value. The government decides to set up an experiment where 500 families (each with 1 child) are randomly assigned to a treatment group (eligible for food stamps, $T_i = 1$) and to a control group (ineligible for food stamps, $T_i = 0$). The government has hired a researcher to investigate the effect of food stamps on the probability that a child has poor health.

After the experiment the researcher performs a regression of $H_i$ (a binary variable that equals 1 if a child has poor health) on $F_i$ (a binary variable that equals 1 if a family received food stamps). She obtains the following OLS estimation results. The table below offers 95% conficence intervals (C.I.) for the two coefficients (both used robust standard errors), where the slope coefficient estimate and its standard error are missing.

| C.I. of OLS fit of $H$ on $F$ | | |
|---|---|---|
| | 95% C.I. | |
| *Intercept* | 0.5788 | 0.7288 |
| $F$ | -0.3007 | -0.1173 |

**a.** Obtain the slope coefficient estimate and its robust standard error from the available information. Then, interpret the estimated coefficients.

**b.** The researcher finds out that some of the families in the control group received food stamps. Explain whether or not we can interpret the estimated OLS coefficient on $F$ as the marginal effect of food stamps on child health?

When the researcher finds out that some of the families in the control group received food stamps she decides to estimate the effect of food stamps using an instrumental variable approach. She uses the assignment to the treatment group as instrument for the actual receipt of food stamps. She obtains the following first stage estimation results.

| OLS fit of $F$ on $T$ | | |
|---|---|---|
| | Robust SE | Coefficient |
| *Intercept* | 0.307 | 0.376 |
| $T$ | 0.307 | 0.624 |

**c.** Do you think that the instrument relevance condition holds? Is $T$ a weak instrument? Do you think that the instrument exogeneity condition holds?

**d.** The following table shows the averages of $H$ and $F$ for those assigned to treatment group ($T = 1$) and for those assigned to the control group ($T = 0$). Use the results in the table below to obtain the instrumental variable estimate of the effect of food stamps on the probability that a child has poor health.

| Conditional means estimates | | |
|---|---|---|
| | $T = 1$ | $T = 0$ |
| $\hat{E}(H \mid T)$ | 0.477 | 0.544 |
| $\hat{E}(F \mid T)$ | 1 | 0.376 |

where $\hat{E}(H \mid T = x)$ is the sample average of the $H_i's$ such that $T_i = x$, and $\hat{E}(F \mid T = x)$ is the sample average of the $F_i's$ such that $T_i = x$. Hint: You can exploit the fact that

$$\frac{1}{n}\sum_{i=1}^{n} H_i T_i - \bar{H}\bar{T} = \frac{n_1 n_0}{n}\left[\hat{E}(H \mid T = 1) - \hat{E}(H \mid T = 0)\right]$$

and

$$\frac{1}{n}\sum_{i=1}^{n} F_i T_i - \bar{F}\bar{T} = \frac{n_1 n_0}{n}\left[\hat{E}(F \mid T = 1) - \hat{E}(F \mid T = 0)\right],$$

where $n_1 = \sum_{i=1}^{500} T_i$ and $n_0 = 500 - n_1$.

**3.** Consider a model relating logarithm of wages per hour in euros ($\ln Wage$), $Tenure$ (number of years working in the same company), gender (using a dummy variable $Man$ taking the value 1 if man and 0 otherwise), and education summarized by the following explanatory variables

$$St0 = \begin{cases} 1 \text{ if only Elementary School} \\ 0 \text{ otherwise} \end{cases}$$
$$St1 = \begin{cases} 1 \text{ if completed High/secondary School} \\ 0 \text{ otherwise} \end{cases}$$
$$St2 = \begin{cases} 1 \text{ if University undergraduate studies.} \\ 0 \text{ otherwise} \end{cases}$$
$$St3 = \begin{cases} 1 \text{ if University postgraduate studies.} \\ 0 \text{ otherwise} \end{cases}.$$

The model

$$\ln(Wage) = \beta_0 + \beta_1 Man + \beta_2 Tenure + \beta_3 Tenure^2 + \beta_4 St1 + \beta_5 St2 + \beta_6 St3 + U$$

is estimated by ordinary least squares using the Spanish Family Expenditure Survey. The output of standard econometrics software is provided below.

Dependent Variable: $\ln(Wage)$
Method: Least squares
Included Observations: 27629

| | Coefficient | Std. Error | t-statistic | P-value |
|---|---|---|---|---|
| $Man$ | 0.280456 | 0.007082 | 39.59956 | 0.000 |
| $Tenure$ | 0.087021 | 0.001114 | 78.13693 | 0.000 |
| $Tenure^2$ | $-0.001697$ | $3.39E-05$ | $-50.07223$ | 0.000 |
| $St1$ | 0.086134 | 0.021001 | 4.101352 | 0.000 |
| $St2$ | 0.217554 | 0.019959 | 10.90003 | 0.000 |
| $St3$ | 0.752181 | 0.020249 | 37.14699 | 0.000 |
| $C$ | 1.430037 | 0.019946 | 71.69381 | 0.000 |
| R-squared | 0.419248 | Mean dependent var | | 2.387066 |
| Adjusted R-squared | 0.419122 | S.D. dependent var | | 0.759744 |
| S.E. of regression | 0.579041 | F-statistic | | 3323.418 |
| Sum squared resid | 9261.353 | Prob(F-statistic) | | 0.0000 |

We know that $\widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = 0.000121$ and $\widehat{Cov}(\hat{\beta}_5, \hat{\beta}_6) = 0.000373$. It is assumed that all classical assumptions are satisfied. All the hypothesis tests and confidence intervals are at the 5% significance level and the 95% confidence level.

**a.** Provide the OLS estimates of the coefficients $\gamma$ in the model

$$\ln(Wage) = \gamma_0 + \gamma_1 Man + \gamma_2 Tenure + \gamma_3 Tenure^2 + \gamma_4 St0 + \gamma_5 St1 + \gamma_6 St2 + U$$

**b.** Test that wages of workers with postgraduate university studies are higher than salaries than workers with undergraduate university studies.

**c.** Test that tenure has no effect on wages.

**d.** If you suspect that there are unobservable variables, like "ability", correlated with salaries and with education, explain how you would estimate the model by two step least squares. Explain how many instruments do you need and provide examples of possible instruments.

**4.** The following models have been estimated by OLS to explain the consumption of alcohol ($A$), measured as number of drinks per week, using a random sample of 500 22 years old people (heteroskedasticity robust standard errors in parenthesis).

$$\hat{A}_i = \underset{(1.1)}{2.4} - \underset{(0.4)}{0.8}\,Educ_i - \underset{(0.1)}{0.3}\,IQ_i \tag{1}$$

where $Educ_i$ and $IQ_i$ are years of education and inteligence quefficient of individual $i$-th, respectively, which are assumed to be exogenous, at least otherwise stated. The sample covariance between $Educ$ and $IQ$ is $\widehat{Cov}(Educ, IQ) = 2.7$.

**a.** Suppose you neglect $IQ$ and you estimate the model

$$A_i = \beta_0 + \beta_1 Educ_i + U_i. \tag{2}$$

Provide an expression for the bias of the OLS estimator of the marginal effect of $Educ$ in this model, and explain what is the expected sign of the bias.

**b.** Suppose that the estimate of the omitted variable model is

$$\hat{A}_i = \underset{(1.1)}{2.4} - \underset{(0.4)}{1.2}\,Educ_i$$

Provide an expression for the sample variance of $Educ$.

**c.** It happens that we know "*intelectual maturity*" ($IM$) may explain $A_i$ better than $IQ_i$, but is not observed. Then it is proposed to estimate model (2) by two stages least squares, using as instrumental variables "father education" $Fath-educ$ and "mother education" $Moth-educ$. Explain how you would test that these two instruments are jointly exogenous.

Critical Values: $\chi^2_{1,(0.05)} = 3.84$, $\chi^2_{2,(0.05)} = 5.99$, $\chi^2_{3,(0.05)} = 7.81$, $Z_{(0.025)} = 1.96$, $Z_{(0.05)} = 1.65$.