ECONOMETRICS FINAL EXAM 2022-23: PART B

Universidad Carlos III de Madrid

Answer all questions in 90 minutes

1. (30%) Consider the following model to estimate the effects of smoking on the annual income (*income*) in US,

$$\ln(income) = \beta_0 + \beta_1 \cdot cigs + \beta_2 \cdot educ + u, \tag{1}$$

where *cigs* is the average of the smoked cigarettes per day and *educ* are the years of education. To reflect that tobacco use may be jointly determined with *income*, the following equation is postulated for the demand for cigarettes::

$$cigs = \gamma_0 + \gamma_1 \cdot \ln(income) + \gamma_2 \cdot \ln(cigpric) + \gamma_3 \cdot restaurn + v, \tag{2}$$

where *cigpric* is the price of a pack of cigarettes in cents, and *restaurn* is a dummy variable that takes the value 1 if the person lives in a state where smoking in restaurants is prohibited. In this demand model, ln(*income*) and *cigs* are endogenous variables and *educ*, ln(*cigpric*) and *restaurn* are exogenous variables. We have a random sample of the variables in the model. The errors (u, v) are homoskedastic and their covariance is unknown.

   (a) (3%) We are interested in estimating the parameters of the two previous structural equations. What is the interpretation of the coefficients β_1 and γ_1 in terms of an exogenous increase in cigarette consumption and income, respectively?

   (b) (7%) Prove that *cigs* is an endogenous variable in (1) from its reduced form.

   (c) (10%) Explain how you could get consistent estimators of β_1 and γ_1. (5%) Carefully explain the assumptions necessary for the proposed estimators to be consistent (5%). In your answer you must indicate if the parameters in each of the equations (1) and (2) are (i) unidentified, (ii) over-identified or (iii) exactly identified.

   (d) (10%) Can you check if at least one of the variables *educ*, $\ln(cigpric)$ or *restaurn* is exogenous? (5%) Explain how you would make that contrast (5%).

ANSWERS:

a. β_1 : *one additional cigarette smoked, leaving the rest of the explanatory variables fixed, supposes an average change in income of* $100 \cdot \beta_1\%$. γ_1: *a 1% increase in income, leaving the rest of the explanatory variables fixed, implies an average change of* $\gamma_1/100$ *cigarettes smoked.*

b. *We obtain the reduced form of cigs:*

$$\begin{aligned} cigs &= \gamma_0 + \gamma_1 \cdot [\beta_0 + \beta_1 \cdot cigs + \beta_2 \cdot educ + u] + \gamma_2 \cdot \ln(cigpric) \\ &\quad + \gamma_3 \cdot restaurn + v, \\ cigs \cdot (1 - \gamma_1\beta_1) &= (\gamma_0 + \gamma_1\beta_0) + \gamma_1\beta_2 educ + \gamma_2 \cdot \ln(cigpric) + \gamma_3 \cdot restaurn + (\gamma_1 u + v). \end{aligned}$$

If $\gamma_1\beta_1 \neq 1$,

$$\begin{aligned} cigs &= \frac{(\gamma_0 + \gamma_1\beta_0)}{(1 - \gamma_1\beta_1)} + \frac{\gamma_1\beta_2}{(1 - \gamma_1\beta_1)} educ + \frac{\gamma_2}{(1 - \gamma_1\beta_1)} \cdot \ln(cigpric) \qquad (3) \\ &\quad + \frac{\gamma_3}{(1 - \gamma_1\beta_1)} \cdot restaurn + \frac{1}{(1 - \gamma_1\beta_1)} (\gamma_1 u + v). \end{aligned}$$

Therefore,

$$\begin{aligned} Cov(cigs, u) &= Cov\left(\frac{\gamma_1}{(1 - \gamma_1\beta_1)} u, u\right) + Cov\left(\frac{1}{(1 - \gamma_1\beta_1)} v, u\right) \\ &= \frac{1}{(1 - \gamma_1\beta_1)} [\gamma_1 Var(u) + Cov(v, u)], \end{aligned}$$

which is different from zero except in the trivial case in which $\gamma_1 = 0$ *and, additionally,* $Cov(v, u) = 0$.

c. *By MC2E: In a first stage we would obtain the reduced form of cigs,*

$$cigs = \pi_0 + \pi_1 educ + \pi_2 \cdot \ln(cigpric) + \pi_3 \cdot restaurn + \varepsilon,$$

which is in fact (3) in terms of the reduced form parameters. Once the parameters of this equation have been estimated by OLS, we obtain the predicted values

$$\widehat{cigs}_i = \hat{\pi}_0 + \hat{\pi}_1 educ_i + \hat{\pi}_2 \cdot \ln(cigpric_i) + \hat{\pi}_3 \cdot restaurn_i.$$

In a second stage we estimate the parameters of the equation

$$\ln(income_i) = \beta_0 + \beta_1 \cdot \widehat{cigs}_i + \beta_2 \cdot educ_i + error, \qquad (4)$$

the OLS estimator in this equation is the 2SLS estimator. This estimator will be consistent if a) the variables educ, $\ln(cigpric)$ *and restaurn are exogenous (they are not correlated with u) and b) The explanatory variables in (4),* $\widehat{cigs}$ *and educ, do not observe perfect multicollinearity. The parameters* *in (a) are overidentified (*$\ln(cigpric)$ *and restaurn are the instrumental variables and there is only one endogenous explanatory variable in the model, cigs) and the parameters in (b) are exactly identified (the variable educ is the instrumental variable and there is also a single endogenous explanatory variable,* $\ln(income)$*).*

d. *We can only check the exogeneity of* $\ln(cigpric_i)$ *and restaurn in the equation (1) because it is the only one that is over-identified. To do this, we estimate the parameters of (a) by 2SLS and calculate the residuals of the structural equation*

$$\begin{aligned}\hat{u}_i &= \ln(income_i) - \ln(\widehat{income_i}), \\ \ln(\widehat{income_i}) &= \hat{\beta}_0^{MC2E} + \hat{\beta}_1^{MC2E} \cdot cigs_i + \hat{\beta}_2^{MC2E} \cdot educ_i.\end{aligned}$$

Then we consider the model

$$\hat{u}_i = \alpha_0 + \alpha_1 \cdot educ_i + \alpha_2 \cdot \ln(cigpric_i) + \alpha_3 \cdot restaurn_i + error,$$

and test

$$H_0 : \alpha_2 = \alpha_3 = 0 \text{ vs } H_1 : \alpha_2 \neq 0 \text{ and/or } \alpha_3 \neq 0,$$

compute the F statistic and consider the *statistic* $J = 2 \cdot F$. *We have* $m = 2$ *instrumental variables and* $k = 1$ *endogenous variable (cigs) in the model. Under* H_0, J *is approximately distributed as a* χ_1^2 $(m - k = 1)$*. We reject* H_0 *if* $J > \chi_{1,\alpha}^2$, *where* $\Pr\left(\chi_1^2 > \chi_{1,\alpha}^2\right) = \alpha$.

2. (10%) Suppose you are interested in explaining the probability that an individual smokes by ignoring the possible endogeneity between cigarette consumption and income. To this end, we create a binary variable, *smoke*, which indicates whether $cigs > 0$ ($smoke = 1$) or not ($smoke = 0$). Discuss the advantages and disadvantages of using the linear probability model to explain the decision to smoke or not smoke. The answer must clearly indicate what the linear probability model consists of. Briefly indicate an alternative to remedy the deficiencies indicated.

<u>ANSWER:</u>

Slides 6 and 10 of Topic 7 (macrogroup).

3. (10%) The following simple linear regression model has been estimated with data (Y_i, X_i), $i = 1, ..., 70$, and the fit is the following:

$$\hat{Y}_i = \underset{(1.22)}{3.25} - \underset{(0.6)}{2.5} X_i,$$

we know that both Y_i and X_i are positive for all $i = 1, ..., n$, and that the sample variance of X is $s_X^2 = 1.23$. What will be the sign of the slope estimator in a regression model without a constant using the same data? Justify the answer. (An answer without a correct justification will not have any credit).

ANSWER:

The estimate of the slope, δ, in the model without intercept

$$Y_i = \delta \cdot X_i + u_i$$

is

$$\hat{\delta} = \frac{\sum_{i=1}^{n} Y_i \cdot X_i}{\sum_{i=1}^{n} X_i^2} > 0 \text{ because } X_i > 0 \text{ and } Y_i > 0 \text{ for all } i = 1, ..., n.$$

4. (10%) Consider

$$\ln Y_i = \beta_0 + \beta_1 \ln X_i + \beta_2 \ln^2 X_i + \beta_3 D_i + \beta_4 (D_i \cdot \ln X_i) + \beta_5 \left(D_i \cdot \ln^2 X_i\right) + u_i,$$

where $E(u_i | X_i, D_i) = 0$ and $D_i \in \{0, 1\}$ is a binary variable. (a) (5%) Explain how you would test the expected elasticities of Y with respect to X in the two models for each value of D_i. (b) (5%) Explain how you would test that the expected elasticity of Y with respect to X is constant in the two models for any value of X.

ANSWER:

(a) $H_0 : \beta_4 = \beta_5 = 0$ *vs* $H_1 : \beta_4 \neq 0$ *and/or* $\beta_5 \neq 0$. *This is tested with an F* statistic.

(b) $H_0 : \beta_2 = \beta_5 = 0$ *vs* $H_1 : \beta_2 \neq 0$ *and/or* $\beta_5 \neq 0$. *This is tested with an F* statistic.