## 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::

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

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  $\beta_1$  and  $\gamma_1$  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  $\beta_1$  and  $\gamma_1$ . (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%).
- 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.

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:

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

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).

4. (10%) Consider

$$\ln Y_{i} = \beta_{0} + \beta_{1} \ln X_{i} + \beta_{2} \ln^{2} X_{i} + \beta_{3} D_{i} + \beta_{4} \left( D_{i} \cdot \ln X_{i} \right) + \beta_{5} \left( D_{i} \cdot \ln^{2} X_{i} \right) + u_{i},$$

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.