Econometrics-2022-2023
Home assignment 5-6. Stochastic Regressors. Simultaneous Equations. Binary model
To be submitted by December 28, 23:55

1. [30 marks] The following simultaneous equations model is considered:

$$Y = \beta_1 + \beta_2 X + u \qquad \textbf{(1)}$$

$$X = \alpha_1 + \alpha_2 Y + v \qquad \textbf{(2)}$$

where X and Y are endogenous variables, and u and v are identically and independently distributed disturbance terms with zero means. The sample consists of n observations (X_i, Y_i).

1.1. [10 marks] □ Derive reduced form system of equations for the system above.
□ Using reduced form system show that in equations (1)-(2) Gauss-Markov conditions (GMC) are violated

1.2. [10 marks] □ Show that OLS estimator $\hat{\alpha}_2^{OLS}$ of α_2 is inconsistent.

1.3. [10 marks] □ What can be said on the identification of the second equation?
An additional instrument is introduced into the equation system - an exogenous variable Z that correlates with a variable Y but does not correlate with a random term v of the second equation.

$$Y = \beta_1 + \beta_2 X + \beta_3 Z + u \qquad \textbf{(1*)}$$

$$X = \alpha_1 + \alpha_2 Y + v \qquad \textbf{(2)}$$

□ Show that the instrumental variable estimator $\hat{\alpha}_2^{IV}$ based on the instrument Z is consistent.

□ The researcher decides to use two-stage least squares (TSLS) hoping to obtain a more efficient estimator of α_2. First he fits OLS regression

$$\hat{Y} = h_1 + h_2 Z \qquad \textbf{(3)}$$

saves the fitted values, and uses them as an instrument for *Y* in equation (2). Demonstrate that obtained TSLS estimator $\hat{\alpha}_2^{TSLS}$ is the same as $\hat{\alpha}_2^{IV}$.

2. [20 marks] During the pandemic, traditional cinemas suffer losses and companies providing films for viewing on the Internet (like Netflix) thrive. A researcher is interested in investigating the expenditures on internet films using cross-section data for 43 countries assuming that expenditure on films, *q*, is related to total consumer expenditure, *z*, by the relationship

$$q = \alpha + \beta z + v$$

where *v* is a disturbance term which satisfies the Gauss-Markov conditions. Both variables *q* and *z* are measured with error, and the researcher believe that any error in the estimation of *q* affects the estimate of *z* by the same amount: $y_i = q_i + w_i$ and $x_i = z_i + w_i$ where y_i is the estimated value of q_i, x_i is the estimated value of z_i, and w_i is the measurement error affecting both variables in observation *i*. It is assumed that the expected value of *w* is zero and that *v* and *w* are distributed independently of *z* and of each other. Note since expenditure on films is a component of total consumer expenditure, β will lie between 0 and 1.

2.1. [10 marks] □ Derive an expression for the large-sample bias in the estimate of β when Ordinary Least Squares is used to regress y_i on x_i, and determine its sign if this is possible.

2.2. [10 marks] □ The researcher is worried of the fact that the analysis could be affected by positive correlation of *w* with *z*, as observations with large *z* tend to have larger measurement errors *w*. Comment.
□ Trying to overcome consequences of bias caused by measurement errors the researcher decided to use disposable personal income, *I* as an instrument for total consumer expenditure, *z*, assuming that *I* correlates with *z* but not correlates with *v* and *w*. Comment providing necessary proofs, taking into account that consumer expenditures on films, *q*, still are under measurement errors *w* .

3. [30 marks] Consider a linear probability model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i;\ i = 1, 2, ..., n,$$

where $Y_i = 1$ if the event takes place, $Y_i = 0$ otherwise and $E(u_i) = 0$.

3.1. [10 marks] □ What is the interpretation of this model and its coefficients?

□ What problems are connected with linear probability model? (Answer exactly to the question, no mathematical details are expected here)

3.2. [10 marks] The alternative approach to estimation of linear regression is based on the using of so called logit model

□ Outline the idea and general structure of the logit model (*no explanation of ML estimation is expected here and no mark will be given for this*).

□ What is the difference in the interpretation of its coefficients and evaluation of the marginal effects of the factors?

3.3. [10 marks] How to evaluate statistical quality of the binary choice (logit and probit) models?

□ How to evaluate whether the coefficients of the logit model are significant?

□ What are McFadden R^2 and *LR* statistic? How they can be used?

4. [20 marks] Consider simple linear regression

$$Y_i = \beta_1 + \beta_2 X_i + u_i;\ i = 1, 2, ..., n, \quad (1)$$

where $E(u_i) = 0$. (*Answer exactly to the question, no mathematical details connected with normal distribution of disturbance term are expected here*)

4.1 [10 marks] □ Explain the concept of the likelihood function and the maximum likelihood estimators of the regression coefficients

□ Explain why the likelihood function is usually used in the form of loglikelihood function.

4.2. [10 marks] □ Describe briefly the properties of MLE estimators.

□ Under what additional conditions does estimating regression (1) by the maximum likelihood method become preferable to estimating by the OLS method? (*Answer exactly to this question, no details are expected here*)