Spring 2022

Quantitative Analysis
Chung-Ming Kuan
Jui-Chung Yang

Problem Set 5

Due: 03/28

Part One: Hand-Written Exercise

1. Verify the statement on slide 39, Lecture 4. That is, write down the 3×3 matrix $\mathbf{R}\tilde{\mathbf{D}}\mathbf{R}'$ using the notation d_{ij}, where $\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$ and
$$\tilde{\mathbf{D}} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1(k+1)} \\ d_{21} & d_{22} & \cdots & d_{2(k+1)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{(k+1)1} & d_{(k+1)2} & \cdots & d_{(k+1)(k+1)} \end{bmatrix}.$$

2. Verify the statement on slide 51, Lecture 4. That is, $\hat{\beta}_{1,IV} = \dfrac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}$.

3. Verify the statement on slide 56, Lecture 4. That is, $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\text{2SLS}} - \boldsymbol{\beta_o}\right) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{D_o})$.

4. Answer the following question with "True" or "False" and briefly explain it. All notations are defined as in our lecture slides.

   (a) Using the OLS method, $\mathbf{X}'\hat{\mathbf{u}}$ is no longer equal to $\mathbf{0}$ when $\mathbf{X}$ is endogenous.

5. Consider the model
$$y_i = \beta_0 + \beta_1 x_i + u_i$$
that includes an endogenous regressor $\mathbf{x}$. To solve the endogeneity, we find an instrumental variable $\mathbf{z}$. Let $\tilde{\beta}_1$ denote the IV estimation of β_1 and $\hat{\beta}_1$ denote the OLS estimation of β_1 with the endogenous regressor $\mathbf{x}$.

   (a) Show that the condition $\rho_{zu} = 0$ ensures that $\tilde{\beta}_1$ is consistent for β_1.

   (b) When will the bias of $\hat{\beta}_1$ be less than $\tilde{\beta}_1$? Express your answer with ρ_{zu}, ρ_{zx} and ρ_{xu}.

Part Two: Part Two: Computer Exercise

1. Please load the dataset `SchoolingReturns` in `R`, which is a cross-section data from the U.S. National Longitudinal Survey of Young Men (NLSYM) in 1976, containing 3,010 observations on 22 variables. The variable we are interested in modelling is "`wage`". However, using the variable "`education`", the years of education, to explain "`wage`" is problematic because it can be argued that schooling is endogenous (and thus "`experience`"

is also endogenous since it equals to `age - education - 6`). Thus, we conduct 2SLS estimations with the outcome log(`wage`), endogenous regressors "`education`", "`experience`" and the square of "`experience`" with their IV "`nearcollege`", "`age`" and the square of "`age`". Other exogenous regressors are "`ethnicity`", "`smsa`" and "`south`".

(a) Perform the first stage of 2SLS.

(b) Perform the second stage of 2SLS. Show the estimated coefficient for "`education`".

(c) Perform 2SLS with the function "`ivreg`" and show the estimated coefficient for "`education`". Verify that it's identical to (b).