Graduate IO Problem Set 1

Junya KAWAMURA

April 24, 2024

1 Question 1

1

Let

$$Y = \begin{pmatrix} log(wage_1) \\ log(wage_2) \\ \vdots \\ log(wage_N) \end{pmatrix}, X = \begin{pmatrix} 1 & educ_1 \\ 1 & educ_2 \\ \vdots & \\ 1 & educ_N \end{pmatrix}, u = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

Then, we can rewrite the linear regression model as

$$Y = X\beta + u$$

where $\beta = (\beta_0, \beta_1)^{'}$ We can define the least squared residuals as

$$\begin{aligned} S(b_0) &= \mathbf{e_0^{'}e_0} \\ &= (Y - Xb_0)'(Y - Xb_0) \\ &= Y^{'}Y - 2Y^{'}Xb_0 + b_0^{'}X^{'}Xb_0 \end{aligned}$$

We can derive the OLS estimator b by solving the following FOC.

$$\begin{aligned} \frac{\partial S(b)}{\partial b_0} &= -2X^{'}Y + 2X^{'}Xb = 0 \\ &\Longrightarrow b = (X^{'}X)^{-1}X^{'}Y \end{aligned}$$

See appendix for program code.

2

See Appendix.

3

By Proposition 2.1 of Hayashi (2000), we can derive the asymptotic variance estimator as

$$\hat{Avar}(b) = \mathbf{S_{xx}^{-1}}\mathbf{\hat{S}}\mathbf{S_{xx}^{-1}}$$

where $\mathbf{S_{xx}} = \frac{1}{n}X'X$ and $\mathbf{\hat{S}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x_i}'\hat{\beta})^2\mathbf{x_i}'\mathbf{x_i}$.

The estimates of the asymptotic standard errors are calculated as

$$SE_{\hat{\beta}} = \sqrt{\frac{\hat{b}}{N}}$$

See appendix for program code.

4

The mean independence assumption is not likely to hold if β is an unbiased estimator. One possible bias of the OLS estimator from this regression model is individual ability. I explain it by considering the following true model based on the above story. We assume the true model as

$$log(wage_i) = \alpha_0 + \alpha_1 educ_i + A_i + u_i \tag{1.1}$$

where A_i is i's ability and $E[u_i|educ_i, A_i] = 0$ holds. From the mean independence assumption, we derive $Cov(educ_i, \epsilon_i) = 0$. So, we derive

$$\begin{aligned}
\beta_1 &= \frac{Cov(educ_i, log(wage_i))}{Var(educ_i)} \\
&= \frac{Cov(educ_i, \alpha_0 + \alpha_1 educ_i + A_i + u_i)}{Var(educ_i)} \\
&= \alpha_1 + \frac{Cov(educ_i, A_i)}{Var(educ_i)} (\because E[u_i|educ_i] = E[E[u_i|educ_i, A_i]|educ_i] = 0 \Longrightarrow Cov(u_i, educ_i) = 0)
\end{aligned}$$

The second term means the omitted variable bias.

5

Exclusion restrictions are necessary for getting an IV estimator, but This IV doubts the validity of the exclusion restriction. For instance, parents with higher education levels tend to be more educational conscious. These parents may invest more in private education (e.g., tutoring) and parental nurturing at home, potentially increasing income through channels to enhance cognitive and non-cognitive abilities other than the child's years of schooling. In this case, bias issues cannot be resolved because of the exclusion restriction violation.

6

We consider the TSLS estimator. Let

$$Z = \begin{pmatrix} fatherduc_1 \\ fatherduc_2 \\ \vdots \\ fatherduc_N \end{pmatrix}$$

By regressing X on Z in the first stage, we derive the prediction of X as $\hat{X} = P_z X$ where $P_z = Z(Z^{'}Z)^{-1}Z^{'}$. By inserting $\hat{X}$ into the second stage model and regress Y on $\hat{X}$, we derive the TSLS estimator as

$$b_{TSLS} = (X^{'}P_z X)^{-1}Z^{'}P_z Y$$

Note that since the number of instrumental variables are equal to the number of endogenous variables, the Two-Stage Least Squares (TSLS) estimator and the Instrumental Variables (IV) estimator are the same. And,

$$\begin{aligned} b_{TSLS} &= \beta + (X^{'}P_z X)^{-1}Z^{'}P_z Z^{'}e \\ \Longleftrightarrow \sqrt{n}(b_{TSLS} - \beta) &= \left(\left(\frac{1}{n}X^{'}Z\right)\left(\frac{1}{n}Z^{'}Z\right)^{-1}\left(\frac{1}{n}Z^{'}X\right)\right)^{-1}\left(\frac{1}{n}X^{'}Z\right)\left(\frac{1}{n}Z^{'}Z\right)^{-1}\left(\frac{1}{\sqrt{n}}Z^{'}e\right) \end{aligned}$$

From the CLT and assumptions, we have

$$\sqrt{n}(b_{TSLS} - \beta) \underset{d}{\rightarrow} N(0, V_\beta)$$

where $V_\beta = (Q_{XZ}Q_{ZZ}^{-1}Q_{XZ})^{-1}E[e^2], Q_{XZ} = E[X^{'}Z], Q_{ZZ} = E[Z^{'}Z], Q_{ZX} = E[Z^{'X}]$
So, the asymptotic covariance matrix is estimated as

$$\hat{V}_b = (\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{XZ})^{-1}\frac{1}{n}\left(\hat{e}^2\right) \tag{1.2}$$

where $\hat{Q}_{XZ} = \frac{1}{n}X^{'}Z, \hat{Q}_{ZZ} = \frac{1}{n}Z^{'}Z, \hat{Q}_{ZX} = \frac{1}{n}Z^{'}X$. The calculation of standard error is the same as question 1-3.
See appendix for program code.

2 Question 2

1

We denote the representative utility as $V_{i,j,k}$. And, we normalize the representative utility as $\tilde{V}_{i,j,k} = V_{i,j,k} - \beta_i y$ for all $j \in \{Kinoko, Takenoko, outside\}$. Then, we have

$$\frac{log(Pr(d_{ik} = Kinoko))}{log(Pr(d_{ik} = outside)))} = \delta_{Kinoko,k} = \alpha_{Kinoko} - \beta p_{Kinoko,k} \tag{2.1}$$
$$\frac{log(Pr(d_{ik} = Takenoko))}{log(Pr(d_{ik} = outside)))} = \delta_{Takenoko,k} = \alpha_{Takenoko} - \beta p_{Takenoko,k} \tag{2.2}$$

Note that $\delta_{Kinoko,k}$, $\delta_{Takenoko,k}$ and $\delta_{outside,k}$ are the mean utilities for each j. Because we can observe $p_{j,k}$ in the data, we can estimate $(\alpha_{Kinoko}, \alpha_{Takenoko}, \beta)$ by using the Monte Carlo Simulation for Numerical Integration.

2

Because equation (2.1) and equation (2.2) do not include income information y_i, we don't need to have income information to identify all parameters. Thus, We can estimate all parameters.

3

We can estimate all parameters because all parameters do not depend on the choice occasion.

4

Let $J \equiv \{Kinoko, Takenoko, outside\}$. In this setting, we can rewrite the log-likelihood function as

$$logL_(\theta) = \sum_{i=1}^{N} \sum_{k=1}^{5} \sum_{j \in J} d_{ijk} log \left(\frac{exp(\delta_{j,k})}{\sum_{l \in J} exp(\delta_{l,k})} \right)$$

where $\delta_{Kinoko,k} = \alpha_{Kinoko} - \beta p_{Kinoko,k}$, $\delta_{Takenoko,k} = \alpha_{Takenoko} - \beta p_{Takenoko,k}$, $\delta_{outside,k} = 0$. We maximize the above log-likelihood function to estimate all parameters.
See appendix for program code.

3 Question 3

1

From the definition of variance, we have

$$E[X^2] = V[X] + (E[X])^2 = \sigma^2 + \mu^2 = 6$$

2

See appendix.