Graduate IO Problem Set 1

Junya KAWAMURA

April 26, 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.

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

The estimation result is as follows.

Table 1.1: Analytical OLS estimation Result

	OLS
Constant	-0.1852
educ	0.1086
Num.Obs.	428

See appendix for program code.

2

The estimation result is as follows.

Table 1.2: OLS estimation Result

	Analycal OLS	Numerical OLS
Constant	-0.1852	-0.1849
educ	0.1086	0.1086
Num.Obs.	428	428

See appendix for program code.

3

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

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

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

The estimates of the asymptotic standard errors are calculated as

$$SE_{\hat{\beta}} = \sqrt{\frac{A\hat{var}(b)}{N}}$$

The estimation result is as follows.

Table 1.3: Analytical OLS estimation Result

	Analytical OLS
Constant	-0.1852
	(0.1703)
educ	0.1086
	(0.01338)
Num.Obs.	428

Note: Standard errors are in parentheses.

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. Based on the above story, I explain it by considering the following true model. 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 equals the number of endogenous variables, the Two-Stage Least Squares (TSLS) estimator and the Instrumental Variables (IV) estimator are the same. And,

$$b_{TSLS} = \beta + (X^{'}P_zX)^{-1}Z^{'}P_zZ^{'}e$$

$$\iff \sqrt{n}(b_{TSLS} - \beta) = \left(\left(\frac{1}{n}X^{'}Z\right)\left(\frac{1}{n}Z^{'}Z\right)^{-1}\left(\frac{1}{n}Z^{'}X\right)\right)^{-1}\left(\frac{1}{n}X^{'}Z\right)\left(\frac{1}{n}Z^{'}Z\right)^{-1}\left(\frac{1}{\sqrt{n}}Z^{'}e\right)$$

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.
The estimation result is as follows.

Table 1.4: Estimation Result

	OLS	IV
Constant	-0.1852	0.4411
	(0.1703)	(0.4451)
educ	0.1086	0.05917
	(0.01338)	(0.03506)
Num.Obs.	428	428

Note: Standard errors are in parentheses.

See appendix for program code.

2 Question 2

1

We consider the case where no outside option exists in the choice set. This means that consumer i chooses alternative j from the choice set $\{Kinoko, Takenoko\}$. Then, we normalize the utility as $\tilde{U}_{i,j,k} = U_{i,j,k} - \beta_i y$ for all $j \in \{Kinoko, Takenoko\}$. And, we denote the representative utility as $\delta_{j,k} \equiv \alpha_j - \beta p_{jk}$ for all $j \in \{Kinoko, Takenoko\}$. Then, we have,

$$\begin{aligned}\frac{log(Pr(d_{ik} = Kinoko))}{log(Pr(d_{ik} = Takenoko)))} &= \delta_{Kinoko,k} - \delta_{Takenoko,k} \\ &= (\alpha_{Kinoko} - \alpha_{Takenoko}) + \beta(p_{Takenoko,k} - p_{Kinoko,k}). \end{aligned} \tag{2.1}$$

Since we can observe choice probabilities for $j \in \{Kinoko, Takenoko\}$ and there is enough variation about $(p_{Takenoko,k} - p_{Kinoko,k})$, we can estimate the parameter $(\alpha_{Kinoko} - \alpha_{Takenoko})$ and β in equation (2.1). However, we cannot estimate α_{Kinoko} and $\alpha_{Takenoko}$ separately in equation (2.1), we cannot estimate all parameters.

2

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

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

$$\frac{log(Pr(d_{ik} = Takenoko))}{log(Pr(d_{ik} = outside)))} = \delta_{Takenoko,k} = \alpha_{Takenoko} - \beta p_{Takenoko,k} \tag{2.3}$$

Note that $\delta_{Kinoko,k}$, $\delta_{Takenoko,k}$ and $\delta_{outside,k}$ are the representative utilities for each j. Because equation (2.2) and equation (2.3) do not include income information y_i, we don't need to have income information to identify all parameters. And, 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.

3

We cannot estimate all parameters because there is no variation at $p_{Kinoko}, p_{Takenoko}$. If there is no variation at $p_{Kinoko}, p_{Takenoko}$, Multicollinearity issues arise because $p_{j,k}$ is equal to the scalar multiple of 1.

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.
The estimation result is as follows.

Table 2.1: Estimation Result

	estimates
α_{Kinoko}	7.258
$\alpha_{Takenoko}$	7.841
β	0.03834
Num.Obs.	1110

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

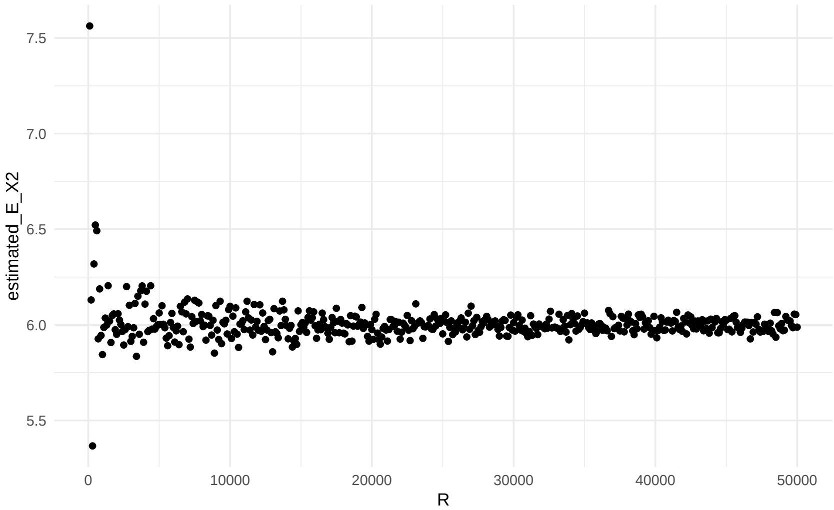
We set $R = 1000$. Then, we calculate the value of $E[X^2]$ by a Monte Carlo simulation. The estimation result is as follows.

Table 3.1: Estimation Result when R = 1000

	estimates
$E[X^2]$	5.90099
R	1000

The numerical value is approximately 5.90, revealing a considerably close approximation to the analytical value of 6. However, due to the law of large numbers, the value of R goes infinity, the numerical value converges in probability to $E[X^2]$. Hence, increasing R decreases the error in the numerical value. This is demonstrated in the graph below, where the horizontal axis represents the number of R and the vertical axis corresponds to the numerical value associated with R. It can be observed that as R increases, the discrepancy with the analytical value decreases.

Figure 3.1: Estimation result depends on the value of R



See appendix for program code.