Econometrics I: Solutions of Homework 1

Hiroki Kato *

April 19, 2020

Contents

1 Solutions

1.1 Question 1

Let $S(\alpha, \beta)$ be the sum of squares residuals:

$$S(\alpha, \beta) = \sum_{t=1}^{T} u_t^2 = \sum_{t=1}^{T} (y_t - \alpha - \beta X_t)^2 \tag{1}$$

*e-mail: `vge008kh@student.econ.osaka-u.ac.jp`. Room 503. All materials I made are published in github: `https://github.com/KatoPachi/2020EconometricsTA.git`. If you have any errors in handouts and materials, please contact me via e-mail or make an issue in github.

The Ordinary Least Squares estimators (hereafter, OLS estimators) can be derived by minimizing (1):

$$(\hat{\alpha}, \hat{\beta}) \in \arg\min_{\alpha,\beta} S(\alpha, \beta) \tag{2}$$

The first-order conditions of this problem are

$$\frac{\partial S(\alpha,\beta)}{\partial \alpha} = -2\sum_{t=1}^{T}(y_t - \hat{\alpha} - \hat{\beta}X_t) = 0 \tag{3}$$

$$\frac{\partial S(\alpha,\beta)}{\partial \beta} = -2\sum_{t=1}^{T}X_t(y_t - \hat{\alpha} - \hat{\beta}X_t) = 0 \tag{4}$$

Note that the second-order condition is hold since the Hessian matrix is positive defenite

$$\mathbf{H} = \begin{pmatrix} \frac{\partial S(\alpha,\beta)}{\partial\alpha\partial\alpha} & \frac{\partial S(\alpha,\beta)}{\partial\alpha\partial\beta} \\ \frac{\partial S(\alpha,\beta)}{\partial\beta\partial\alpha} & \frac{\partial S(\alpha,\beta)}{\partial\beta\partial\beta} \end{pmatrix} = \begin{pmatrix} 2T & 2\sum_t X_t \\ 2\sum_t X_t & 2\sum_t X_t^2 \end{pmatrix}$$

$$\Rightarrow |\mathbf{H}| = \left(2T \cdot 2\sum_t X_t^2\right) - \left(2\sum_t X_t \cdot 2\sum_t X_t\right) = 4T\left(\frac{1}{T}\sum_t X_t^2 - (\sum_t X_t/T)^2\right)$$

$$\Rightarrow |\mathbf{H}| = 4T \cdot V(X_t) > 0$$

where $V(X_t)$ is the variance of X_t

By equation (3), we have

$$\begin{aligned}\hat{\alpha} &= \frac{1}{T}(\sum_t y_t - \hat{\beta}\sum_t X_t) \\ &= \bar{y} - \hat{\beta}\bar{X}\end{aligned} \tag{5}$$

where $\bar{y} = \sum_t y_t/T$ and $\bar{X} = \sum_t X_t/T$ (sample mean). Substituting equation(5) into equation (4) yields

$$\begin{aligned}\hat{\beta}\left(\sum_t X_t^2 - \bar{X}\sum_t X_t\right) &= \left(\sum_t y_t X_t - \bar{y}\sum_t X_t\right) \\ \hat{\beta} &= \frac{\sum_t y_t X_t - \bar{y}\sum_t X_t}{\sum_t X_t^2 - \bar{X}\sum_t X_t} \\ \hat{\beta} &= \frac{\sum_t X_t(y_t - \bar{y})}{\sum_t X_t(X_t - \bar{X})}\end{aligned} \tag{6}$$

Thus, OLS estimators are

$$(\hat{\alpha}, \hat{\beta}) = \left(\bar{y} - \hat{\beta}\bar{X}, \frac{\sum_t X_t(y_t - \bar{y})}{\sum_t X_t(X_t - \bar{X})} \right) \tag{7}$$

Note that the estimator of β, $\hat{\beta}$, can be rewritten as follows:

$$\hat{\beta} = \frac{Cov(y_t, X_t)}{V(X_t)} \tag{8}$$

where $Cov(y_t, X_t)$ is covariance (共分散) between y_t and X_t, and $V(X_t)$ is variance (分散) of X_t. To prove it, we need to recall the defenition of covariance and variance. First, the defenition of covairance is

$$Cov(y_t, X_t) = E[(y_t - E(y_t))(X_t - E(X_t))].$$

Then, sample covariance is

$$\begin{aligned}
S_{y_t, X_t} &= \frac{1}{T-1} \sum_t (y_t - \bar{y})(X_t - \bar{X}) \\
&= \frac{1}{T-1} \sum_t (y_t X_t - y_t \bar{X} - \bar{y} X_t + \bar{X}\bar{y}) \\
&= \frac{1}{T-1} \left(\sum_t y_t X_t - \bar{X} \sum_t y_t - \bar{y} \sum_t X_t + T\bar{X}\bar{y} \right) \\
&= \frac{1}{T-1} \left(\sum_t y_t X_t - \bar{y} \sum_t X_t \right) \\
&= \frac{1}{T-1} \sum_t X_t (y_t - \bar{y})
\end{aligned}$$

Second, the defenition of variance is

$$V(X_t) = E[(X_t - E(X_t))^2].$$

Then, sample variance is

$$
\begin{aligned}
S_{X_t}^2 &= \frac{1}{T-1}\sum_t (X_t - \bar{X})^2 \\
&= \frac{1}{T-1}\sum_t X_t^2 - \frac{2\bar{X}}{T-1}\sum_t X_t + \frac{1}{T-1}T\bar{X}^2 \\
&= \frac{1}{T-1}\left(\sum_t X_t^2 - \bar{X}\sum_t X_t\right) \\
&= \frac{1}{T-1}\sum_t X_t(X_t - \bar{X})
\end{aligned}
$$

Finally, we have

$$
\hat{\beta} = \frac{S_{y_t,X_t}}{S_{X_t}^2} = \frac{\sum_t X_t(y_t - \bar{y})}{\sum_t X_t(X_t - \bar{X})}.
$$

Thus, the OLS estimator of β is $\hat{\beta} = Cov(y_t, X_t)/V(X_t)$, or

$$
\hat{\beta} = \frac{\sum_t (y_t - \bar{y})(X_t - \bar{X})}{\sum_t (X_t - \bar{X})^2} \tag{9}
$$

1.2 Question 2

From (9), we have

$$
\begin{aligned}
\hat{\beta} &= \frac{\sum_t y_t(X_t - \bar{X}) - \bar{y}\sum_t (X_t - \bar{X})}{\sum_t (X_t - \bar{X})^2} \\
&= \frac{\sum_t y_t(X_t - \bar{X})}{\sum_t (X_t - \bar{X})^2}
\end{aligned}
$$

since $\sum_t (X_t - \bar{X}) = \sum_t X_t - T\bar{X} = \sum_t X_t - \sum_t X_t = 0$. Substituting $y_t = \alpha + \beta X_t + u_t$ into this equation yields

$$
\begin{aligned}
\hat{\beta} &= \frac{\sum_t (X_t - \bar{X})(\alpha + \beta X_t + u_t)}{\sum_t (X_t - \bar{X})^2} \\
&= \frac{\beta \sum_t (X_t - \bar{X})X_t + \sum_t (X_t - \bar{X})u_t}{\sum_t (X_t - \bar{X})^2} \\
&= \beta + \frac{\sum_t (X_t - \bar{X})u_t}{\sum_t (X_t - \bar{X})^2} \\
&= \beta + \sum_t \omega_t u_t
\end{aligned}
$$

where $\omega_t = (X_t - \bar{X})/\sum_t (X_t - \bar{X})^2$.

Because β and X_t are not random variables,

$$E(\hat{\beta}) = \beta + \sum_t \omega_t E(u_t) = \beta.$$

The variance of $\hat{\beta}$ is

$$\begin{aligned}
V(\hat{\beta}) &= E[(\hat{\beta} - E(\hat{\beta}))^2] = E(\hat{\beta} - \beta)^2 = E(\sum_t \omega_t u_t)^2 \\
&= E[\sum_t \omega_t^2 u_t^2 + 2\sum_t \sum_{t' \neq t} \omega_t \omega_{t'} u_t u_{t'}] \\
&= \sum_t \omega_t^2 E(u_t^2) + 2\sum_t \sum_{t' \neq t} \omega_t \omega_{t'} E(u_t u_{t'}) \\
&= \sum_t \omega_t^2 E[(u_t - E(u_t))^2] + 2\sum_t \sum_{t' \neq t} \omega_t \omega_{t'} E(u_t)E(u_{t'}) \\
&= \sigma^2 \frac{\sum_t (X_t - \bar{X})^2}{(\sum_t (X_t - \bar{X})^2)^2} = \frac{\sigma^2}{\sum_t (X_t - \bar{X})^2}.
\end{aligned}$$

To derive it, we use following properties:

- mutual independece assumption implies $E(u_t u_{t'}) = E(u_t)E(u_{t'})$.
- By $E(u_t) = 0$, $E(u_t^2) = E[(u_t - E(u_t))^2] = V(u_t)$.

1.3 Question 3

Recall that the estimator of α is

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{X}.$$

Substituting $\bar{y} = \alpha + \beta\bar{X} + \bar{u}$ (average both sides over $t \in \{1, \dots T\}$) into this equation implies

$$\hat{\alpha} = \alpha - (\hat{\beta} - \beta)\bar{X} + \bar{u}.$$

Since $E(\hat{\beta}) = \beta$ and $E(\bar{u}) = E(\sum_t u_t)/T = \sum_t E(u_t)/T = 0$, we obtain

$$E(\hat{\alpha}) = \alpha.$$

The variance of $\hat{\alpha}$ is

$$\begin{aligned} V(\hat{\alpha}) &= E[(\hat{\alpha} - E(\hat{\alpha}))^2] = E(\hat{\alpha} - \alpha)^2 = E(-(\hat{\beta} - \beta)\bar{X} + \bar{u})^2 \\ &= E[(\hat{\beta} - \beta)^2 \bar{X}^2 - 2(\hat{\beta} - \beta)\bar{X}\bar{u} + \bar{u}^2] \\ &= \bar{X}^2 E(\hat{\beta} - \beta)^2 - 2\bar{X}E(\hat{\beta} - \beta)\bar{u} + E(\bar{u}^2). \end{aligned} \tag{10}$$

The first term of equation (10) is

$$\bar{X}^2 E(\hat{\beta} - \beta)^2 = \frac{\bar{X}^2 \sigma^2}{\sum_t (X_t - \bar{X})^2}.$$

The third term of equation (10) is

$$\begin{aligned} E(\bar{u}^2) &= \frac{E(\sum_t u_t)^2}{T^2} \\ &= \frac{E[\sum_t u_t^2 + 2\sum_t \sum_{t' \neq t} u_t u_{t'}]}{T^2} \\ &= \frac{\sum_t E(u_t^2) + 2\sum_t \sum_{t'} E(u_t u_{t'})}{T^2} \\ &= \frac{\sum_t E[(u_t - E(u_t))^2] + 2\sum_t \sum_{t'} E(u_t)E(u_{t'})}{T^2} \\ &= \frac{\sigma^2}{T} \end{aligned}$$

The second term of equation (10) is rewritten as follows:

$$\begin{aligned} & 2\bar{X}E(\hat{\beta} - \beta)\bar{u} \\ =& 2\bar{X}E\left(\frac{\sum_t (X_t - \bar{X})u_t}{\sum_t (X_t - \bar{X})^2} \frac{\sum_t u_t}{T} \right) \\ =& \frac{2\bar{X}}{T\sum_t (X_t - \bar{X})^2} E\left(\sum_t (X_t - \bar{X})u_t \sum_t u_t \right) \\ =& \frac{2\bar{X}}{T\sum_t (X_t - \bar{X})^2} E\left(\sum_t (X_t - \bar{X}) \left(u_t^2 + \sum_{t' \neq t} u_t u_{t'} \right) \right) \\ =& \frac{2\bar{X}}{T\sum_t (X_t - \bar{X})^2} \sum_t (X_t - \bar{X})E(u_t^2) + \sum_t \sum_{t'} (X_t - \bar{X})E(u_t u_{t'}) \\ =& \frac{2\bar{X}}{T\sum_t (X_t - \bar{X})^2} \sum_t (X_t - \bar{X})E[(u_t - E(u_t))^2] + \sum_t \sum_{t'} (X_t - \bar{X})E(u_t)E(u_{t'}) \\ =& \frac{2\bar{X}}{T\sum_t (X_t - \bar{X})^2} \sigma^2 \sum_t (X_t - \bar{X}) = 0. \end{aligned}$$

Hence, we have the variance of $\hat{\alpha}$ as follows:

$$\begin{aligned} V(\hat{\alpha}) &= \frac{\bar{X}^2\sigma^2}{\sum_t (X_t - \bar{X})^2} + \frac{\sigma^2}{T} \\ &= \frac{\sigma^2 \sum_t X_t^2}{T\sum_t (X_t - \bar{X})^2} \end{aligned}$$

2 Review

2.1 Properties of Expectaion and Variance

In this section, we will review some properties of expctation and variance that are used to solve homework.

Mutual Independence

Let X and Y be random variables, which are mutually and independentlly distributed. Then, following properties of expectation and variance must be hold:

1. $E(XY) = E(X)E(Y)$;
2. $Cov(X, Y) = 0$

These two properties means that there is no correlation between X and Y.
Caveate: Independece is *sufficient* condition for uncorrelatedness (exceptional cases: multivariate normal distribution).

Proof. Without loss of generarity, we assume X and Y are continuous random variables. Let $f(x, y)$ be joint distribution of X and Y. By defenition, mutual independce leads to $f(x, y) = f_X(x)f_Y(y)$.

1. By defenition,

$$\begin{aligned} E(XY) &= \int\int xyf(x,y)dydx \\ &= \int\int xyf_X(x)f_Y(y)dydx \\ &= \int xf_X(x)dx \int yf_Y(y)dy \\ &= E(X)E(Y) \end{aligned}$$

2. By defenition,

$$\begin{aligned} Cov(X,Y) &= E[(X-E(X))(Y-E(Y))] \\ &= E[XY - XE(Y) - E(X)Y + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) \\ &= E(X)E(Y) - E(X)E(Y) = 0. \end{aligned}$$

Additivity of Expectaion and Variance

Let $X_1, \ldots, X_n$ be random variables, and let $a_1, \ldots, a_n$ be constants. Then, the following properties is hold:

1. $E(\sum_i a_i X_i + b) = \sum_i a_i E(X_i) + b$;
2. $V(\sum_i a_i X_i + b) = \sum_i a_i^2 V(X_i) + 2\sum_i \sum_{j \neq i} a_i a_j Cov(X_i, X_j)$

Proof. Without loss of generarity, we assume X_i are continuous random variables. Let $f(x_1, \ldots x_n)$ be joint distribution of $X_1, \ldots, X_n$.

1. By defenition,

$$\begin{aligned} &E(\sum_i a_i X_i + b) \\ =&\int \cdots \int (a_1 X_1 + \cdots + a_n X_n + b) f(x_1, \ldots x_n) dx_1 \cdots dx_n \\ =&a_1 \int \cdots \int X_1 f(x_1, \ldots x_n) dx_1 \cdots dx_n + \cdots + b \int \cdots \int f(x_1, \ldots x_n) dx_1 \cdots dx_n \\ =&a_1 \int X_1 \left(\int_{X_2} \cdots \int_{X_n} f(x_1, \ldots x_n) dx_2 \cdots dx_n \right) dx_1 + \cdots + b \\ =&a_1 \int X_1 f(x_1) dx_1 + \cdots + b = \sum_i a_i E(X_i) + b \end{aligned}$$

2. By defenition,

$$
\begin{aligned}
& V(\sum_i a_i X_i + b) \\
=& E\left[(\sum_i a_i X_i + b) - E(\sum_i a_i X_i + b)\right]^2 \\
=& E\left[(\sum_i a_i X_i + b) - (\sum_i a_i E(X_i) + b)\right]^2 \\
=& E\left[\sum_i a_i (X_i - E(X_i))\right]^2 \\
=& E\left[\sum_i a_i^2 (X_i - E(X_i))^2 + \sum_i \sum_{j \neq i} a_i a_j (X_i - E(X_i))(X_j - E(X_j))\right] \\
=& \sum_i a_i^2 E(X_i - E(X_i))^2 + \sum_i \sum_{j \neq i} a_i a_j E(X_i - E(X_i))(X_j - E(X_j)) \\
=& \sum_i a_i^2 V(X_i) + \sum_i \sum_{j \neq i} a_i a_j Cov(X_i, X_j)
\end{aligned}
$$

2.2 Optimization

Consider a case that we aim to obtain a point $x \in \mathbb{R}$ which maximizes or minimizes a function $y = f(x)$. In this case, if $x^0 \in \mathbb{R}$ attains the maximum or minimum, we have the following first order condition at the beginning:

$$\frac{df(x)}{dx}\Big|_{x=x^0} = 0. \tag{11}$$

In addition, when we consider whether the optimum is a maximum or a minimum, the sufficient condition for the optimum becomes as follows:

$$\frac{d^2 f(x)}{dx^2}\Big|_{x=x^0} < 0 \quad \text{for a maximum;} \tag{12}$$

$$\frac{d^2 f(x)}{dx^2}\Big|_{x=x^0} > 0 \quad \text{for a minimum.} \tag{13}$$

Here consider a function $y = g(\mathbf{x})(\in \mathbb{R})$ where $\mathbf{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n$, denoted as $g\colon \mathbb{R}^n \to \mathbb{R}$. If an $\mathbf{x}^0 = (x_1^0, \ldots, x_n^0)' \in \mathbb{R}^n$ maximizes or minimizes $g(\mathbf{x})$, we apply the following theorem.

If a function $g\colon \mathbb{R}^n \to \mathbb{R}$ is maximized (minimized) at the point $\mathbf{x}^0 = (x_1^0, \ldots, x_n^0)$, then the following equation holds:

$$\left.\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^0} = \begin{pmatrix} \frac{\partial g(\mathbf{x}^0)}{\partial x_1} \\ \vdots \\ \frac{\partial g(\mathbf{x}^0)}{\partial x_n} \end{pmatrix} = \mathbf{0}. \tag{14}$$

Moreover, we use the following *Hessian matrix* to discern a maximum and a minimum.

A Hessian matrix of a function $g\colon \mathbb{R}^n \to \mathbb{R}$ is defined as follows:

$$H = \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}\mathbf{x}'} = \begin{pmatrix} \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 g(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 g(\mathbf{x})}{\partial x_n \partial x_n} \end{pmatrix}$$

Assume that $g_{x_1}(\mathbf{x}^0) = g_{x_2}(\mathbf{x}^0) = \cdots = g_{x_n}(\mathbf{x}^0) = 0$ holds, where $g_{x_i}(\mathbf{x})$ for $i \in \{1, \ldots, n\}$ denotes the partial derivative of $g(\mathbf{x})$ with respect to x_i. The following theorem is a way to distinguish whether **x** attains a muximum and a minimum.

Suppose that a smooth function $g\colon \mathbb{R}^n \to \mathbb{R}$ satisfies $g_{x_1}(\mathbf{x}^0) = \cdots = g_{x_n}(\mathbf{x}^0) = 0$. Then, we can confirm that if:

1. H is a negative definite matrix, $\mathbf{x}^0$ is a maximum point.
2. H is a positive definite matrix, $\mathbf{x}^0$ is a minimum point.

As for the positiveness or negativeness of a matrix, we have the following theorem.

A necessary and sufficient condition for a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ to be positive (negative) definite is that eigenvalues λ such that $det(\mathbf{A} - \lambda \mathbf{I}) = 0$ are protive (negative), where $\mathbf{I}$ is identity matrix:

$$\mathbf{I} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & 1 & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

For example, in the case of $f(x, y) = x^2 + 4xy + 5y^2 - 2x - 8y + 5$, we have $f_x = 2x + 4y - 2$ and $f_y = 4x + 10y - 8$. By solving $f_x = f_y = 0$, we obtain an optimum point $(x, y) = (-3, 2)$. Also, the Hessian matrix is given as follows:

$$H_f = \begin{pmatrix} 2 & 4 \\ 4 & 10 \end{pmatrix}.$$

To obtain eigenvalues, we subtract the diagonal matrix with eigenvalues from the Hessian matrix:

$$H_f - \lambda \mathbf{I} = \begin{pmatrix} 2 - \lambda & 4 \\ 4 & 10 - \lambda \end{pmatrix}$$

Then, the determinant of this matrix is

$$f(\lambda) = (2 - \lambda)(10 - \lambda) - 16.$$

Since $f(\lambda)$ is convex, all eigenvalues are positives if $f(0) > 0$ (Write rough graph by yourself). Then, $f(0) = 2 \cdot 10 - 16 > 0$. Note that $f(0)$ is correspond to the determinant of Hessian matrix. Thus, $(x, y) = (-3, 2)$ is a minimum point. We can analyze an optimum of a multivariable function for more variables in the same manner.