PS1 - Econometrics

Romain Fernex

February 2025

Contents

1

Part II - Theory

1.1 Problem 3

To show that the OLS estimator is the same if we remove the constant and center variables, we compute $\hat{\beta}_1$ in both cases (with constant / without constant but with centered variables).

1.1.1 With a constant

We solve the following optimization problem :

$$(\hat{\beta}_0, \hat{\beta}_1) = arg \min_{\beta_1, \beta_0} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{1}$$

We derive the first order conditions with respect to β_0, β_1 :

$$\begin{aligned} -2\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ -2\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i &= 0 \end{aligned} \tag{2}$$

After replacing -2 by $\frac{1}{N}$ to substitute sample means, we get this system of normal equations :

$$\begin{aligned} \bar{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\ \bar{xy} &= \hat{\beta}_0 \bar{x} + \hat{\beta}_1 \bar{x^2} \end{aligned} \tag{3}$$

lets now find $\hat{\beta_1}$ by multiplying the first line in the system (3) by $\bar{x}$ and withdrawing it from the second line to neutralize the term in $\hat{\beta_0}$. We get the following expression of $\hat{\beta_1}$ as a function of sample means :

$$\hat{\beta_1} = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2} \tag{4}$$

Based on this new expression, we can expression $\hat{\beta_1}$ as a function of Covariance and variance terms :

$$\hat{\beta_1} = \frac{Cov(x,y)}{Var(x)} \tag{5}$$

1.1.2 Without a constant and with centered variables

We now have the following set up :

$$(\hat{\beta_1}) = arg \min_{\beta_1} \sum_{i=1}^{N} [(y_i - \bar{y}) - \beta_1(x_i - \bar{x})]^2 \tag{6}$$

As in case 1, we obtain the first order condition with respect to β_1 and replace -2 by $\frac{1}{N}$:

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}[(y_i - \bar{y}) - \hat{\beta_1}(x_i - \bar{x})](x_i - \bar{x}) = 0 \\
\Leftrightarrow \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x}) = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta_1}(x_i - \bar{x})^2 \\
\Leftrightarrow \bar{yx} - \bar{y}\bar{x} = \hat{\beta_1}(\bar{x^2} - \bar{x}^2) \\
\Leftrightarrow \hat{\beta_1} = \frac{\bar{yx} - \bar{y}\bar{x}}{\bar{x^2} - \bar{x}^2}
\end{aligned} \tag{7}$$

We find the same expression for $\hat{\beta_1}$ as in the first case (with the constant term). We have the same OLS estimator as expected.

1.2 Problem 4

1.2.1 Computing OLS estimators

To find the OLS estimator $(\hat{a}, \hat{b}, \hat{c})$ we need to solve the following minimization problem :

$$(\hat{a}, \hat{b}, \hat{c}) = \min_{a,b,c} \sum_{i=1}^{N}(y_i - a - bx_i - cz_i)^2 \tag{8}$$

We begin by obtaining the First order conditions with respect to a, b and c. This gives us :

$$\begin{aligned}
-2\sum_{i=1}^{N}(y_i - \hat{a} - \hat{b}x_i - \hat{c}z_i) &= 0 \\
-2\sum_{i=1}^{N}(y_i - \hat{a} - \hat{b}x_i - \hat{c}z_i)x_i &= 0 \\
-2\sum_{i=1}^{N}(y_i - \hat{a} - \hat{b}x_i - \hat{c}z_i)z_i &= 0
\end{aligned} \tag{9}$$

After replacing -2 by $\frac{1}{N}$ and replacing sums by sample means we get the following system of normal equations:

$$\begin{aligned}
\bar{y} &= \hat{a} + \hat{b}\bar{x} + \hat{c}\bar{z} \\
\bar{xy} &= \hat{a}\bar{x} + \hat{b}\bar{x^2} + \hat{c}\bar{xz} \\
\bar{zy} &= \hat{a}\bar{z} + \hat{b}\bar{zx} + \hat{c}\bar{z^2}
\end{aligned} \tag{10}$$

Let us rewrite this system by multiplying the first line by $\bar{x}$ ($\bar{z}$) and withdrawing it from the second line (respectively the third line) to get rid of the constant term $\hat{a}$. This yields :

$$\begin{aligned} \bar{y} &= \hat{a} + \hat{b}\bar{x} + \hat{c}\bar{z} \\ \overline{xy} - \bar{x}\bar{y} &= \hat{b}(\overline{x^2} - \bar{x}^2) + \hat{c}(\overline{xz} - \bar{x}\bar{z}) \\ \overline{zy} - \bar{y}\bar{z} &= \hat{b}(\overline{zx} - \bar{x}\bar{z}) + \hat{c}(\overline{z^2} - \bar{z}^2) \end{aligned} \tag{11}$$

To simplify this expression, we can replace sample means by their equivalents in terms of variance and covariance.
We know that $Cov(x,y) = \overline{xy} - \bar{x}\bar{y}$ (same with (x,z) and (z,y)) and that $Var(x) = \overline{x^2} - \bar{x}^2$ (same with y and z)
Now we can write :

$$\begin{aligned} \hat{a} &= \hat{b}\bar{x} + \hat{c}\bar{z} - \bar{y} \\ Cov(x,y) &= \hat{b} * Var(x) + \hat{c} * Cov(x,z) \\ Cov(z,y) &= \hat{b} * Cov(z,x) + \hat{c} * Var(z) \end{aligned} \tag{12}$$

The system of equations (12) can be written in matrix form as:

$$\begin{bmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x,z) \\ \mathrm{Cov}(z,x) & \mathrm{Var}(z) \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(x,y) \\ \mathrm{Cov}(z,y) \end{bmatrix}. \tag{13}$$

Let us denote the covariance matrix as:

$$\mathbf{C} = \begin{bmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x,z) \\ \mathrm{Cov}(z,x) & \mathrm{Var}(z) \end{bmatrix}, \tag{14}$$

and the right-hand side vector as:

$$\mathbf{d} = \begin{bmatrix} \mathrm{Cov}(x,y) \\ \mathrm{Cov}(z,y) \end{bmatrix}. \tag{15}$$

The solution for $\hat{b}$ and $\hat{c}$ is given by:

$$\begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \mathbf{C}^{-1}\mathbf{d}.$$

Now we need to compute the inverse of the covariance matrix to solve for this system (we will posit its existence by assuming that x and z are not perfectly collinear $\Leftrightarrow Var(x)Var(z) \neq Cov(x,z)^2$) :
The inverse of a 2×2 matrix is given by:

$$\mathbf{C}^{-1} = \frac{1}{\det(\mathbf{C})} \begin{bmatrix} \mathrm{Var}(z) & -\mathrm{Cov}(x,z) \\ -\mathrm{Cov}(z,x) & \mathrm{Var}(x) \end{bmatrix}, \tag{16}$$

where the determinant of $\mathbf{C}$ is:

$$\det(\mathbf{C}) = \mathrm{Var}(x)\,\mathrm{Var}(z) - \mathrm{Cov}(x,z)^2. \tag{17}$$

Substituting this, we have:

$$\mathbf{C}^{-1} = \frac{1}{\mathrm{Var}(x)\,\mathrm{Var}(z) - \mathrm{Cov}(x,z)^2} \begin{bmatrix} \mathrm{Var}(z) & -\mathrm{Cov}(x,z) \\ -\mathrm{Cov}(z,x) & \mathrm{Var}(x) \end{bmatrix}. \tag{18}$$

Using $\mathbf{C}^{-1}\mathbf{d}$, we get the following expressions for the OLS estimator $(\hat{a}, \hat{b}, \hat{c})$:

$$\begin{aligned} \hat{a} &= \hat{b}\,\bar{x} + \hat{c}\,\bar{z} - \bar{y} \\ \hat{b} &= [C^{-1}d]_{11} = \frac{Cov(x,y)\,Var(z) - Cov(z,y)\,Cov(x,z)}{Var(x)\,Var(z) - Cov(x,z)^2} \\ \hat{c} &= [C^{-1}d]_{21} = \frac{Cov(z,y)\,Var(x) - Cov(x,y)\,Cov(x,z)}{Var(x)\,Var(z) - Cov(x,z)^2} \end{aligned} \tag{19}$$

1.2.2 Constructing the new sample

We wish to construct a sample such that :

$$Cov_N(x_i, z_i) = \sum_{i=1}^{N}(x_i - \bar{x})(z_i - \bar{z}) = 0 \tag{20}$$

Given that $Corr(x, z) = \frac{Cov(x,z)}{Var(x)Var(z)}$, this implies that we need the two regressors x, z to be uncorrelated. This means that they should not be linearly related to one another.

1.2.3 Calculating the estimators given the new sample

We take the results we obtained in (19) and substitute $Cov(x, z)$ by 0. We get the following expression for OLS estimators :

$$\begin{aligned} \hat{b} &= \frac{Cov(x, y)}{Var(x)} \\ \hat{c} &= \frac{Cov(y, z)}{Var(z)} \end{aligned} \tag{21}$$

The estimators for b and c are now independent from each other, as their expression no longer needs to account for a correlation between x and z.

1.2.4 Estimating b and c using two simple regressions

If x and z are uncorrelated, as posited in (1.2.2), we may run two separate simple regressions to estimate the effect of each regressor on y. This is possible because the effect of one regressor does not influence the effect of the other on the dependent variable. The two regressions to run could be as follows :

$$\begin{aligned} &\text{- regression model 1 (to find } \hat{b}) : y_i = a + bx_i + residuals \\ &\text{- regression model 2 (to find } \hat{b}) : y_i = a + cz_i + residuals \end{aligned} \tag{22}$$

We obtain $\hat{b}$ ($\hat{c}$) by finding the slope coefficient from regression model 1 (respectively, regression model 2). Once we have $\hat{b}, \hat{c}$ we can get $\hat{a}$ by using the relation we previously determined in (19) : $\hat{a} = \hat{b}\bar{x} + \hat{c}\bar{z} - \bar{y}$