Derivation of OLS Estimator

In class we set up the minimisation problem that is the starting point for deriving the formulas for the OLS intercept and slope coefficients. That problem was:

$$\min(\beta_0, \beta_1) \ \sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \qquad (1)$$

Before going into detail, we need to ask ourselves why minimisation? In particular, why minimisation of the sum of squared residual to get the unknown parameters(β_0, β_1)? The answer comes from the observed sample data. Suppose, we plot sample data of the outcome and the predictor variables. In that case, we will find a sample relationship line that resonates with the population relationship of the respective variables (if sample is taken properly).

With regression, we can predict the outcome variable in such a way that the constructed regression line mimics the observed sample relationship line with some degree of residual error (i.e., unexplained part). This residual creates a gap when we plot both lines together. This gap, mathematically speaking, can only be reduced if we can cut down sum of squared of the residual. That is why OLS uses this minimisation approach.

As we learned in calculus, a univariate optimisation involves taking the derivative and setting equal to 0. Similarly, this minimisation problem above is solved by setting the partial derivatives equal to 0. That is, take the derivative of (1) with respect to $\beta_0$and set it equal to 0. We then do the same thing for β_1. This gives us the following:

$$\frac{\partial W}{\partial \hat{\beta}_0} = \sum_{i=1}^{N} -2(Y_i - \hat{\beta}_0 - \hat{\beta}_0 X_i) = 0 \qquad (2)$$

$$\frac{\partial W}{\partial \hat{\beta}_1} = \sum_{i=1}^{N} -2X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_0 X_i) = 0 \qquad (3)$$

Note that we have used W to denote the objective function. Now our task is to solve (2) and (3) using some algebra tricks and some properties of summations. Let's start with the first order condition for $\hat{\beta}_0$ from equation (2). We can immediately get rid of the -2 and write the following expression:

$$\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_0 X_i) = 0 \qquad (2a)$$

Now let's rearrange this expression and make use of the algebraic fact that $\sum_{i=1}^{N} Y_i = N\bar{Y}$. This leaves us with equation (4).

$$N\hat{\beta}_0 = N\bar{Y} - N\hat{\beta}_1\bar{X} \qquad (4)$$

We simply divide everything by N and amazing, we have the formula that is shown in the slides

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} \qquad (5)$$

Now let's consider solving for $\hat{\beta}_1$. This one is a bit tricky. We can first get rid of the -2 and rearrange equation (3) to equation (3a).

$$\sum_{i=1}^{N} X_i Y_i - \hat{\beta}_0 X_i - \hat{\beta}_1 {X_i}^2 = 0 \qquad (3a)$$

Now let's substitute our result for $\hat{\beta}_0$ into this expression and this gives us:

$$\sum_{i=1}^{N} X_i Y_i - \left(\bar{Y} - \hat{\beta}_1 \bar{X}\right) X_i - \hat{\beta}_1 {X_i}^2 = 0 \qquad (6)$$

With this expression, we now distribute the summation operator in equation (6) and get the following:

$$\sum_{i=1}^{N} X_i Y_i - \bar{Y} \sum_{i=1}^{N} X_i + \hat{\beta}_1 \bar{X} \sum_{i=1}^{N} X_i - \hat{\beta}_1 \sum_{i=1}^{N} {X_i}^2 = 0 \quad (7)$$

From equation (7), we can find an expression for $\hat{\beta}_1$.

Let's do it step by step:

$\Rightarrow \sum_{i=1}^{N} X_i Y_i - \bar{Y} N \bar{X} + \hat{\beta}_1 \bar{X} N \bar{X} - \hat{\beta}_1 \sum_{i=1}^{N} {X_i}^2 = 0$ $\quad [\because \sum_{i=1}^{N} X_i = N\bar{X}]$

$\Rightarrow \sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y} + \hat{\beta}_1 N \bar{X}^2 - \hat{\beta}_1 \sum_{i=1}^{N} {X_i}^2 = 0$

$\Rightarrow \hat{\beta}_1 N \bar{X}^2 - \hat{\beta}_1 \sum_{i=1}^{N} {X_i}^2 = - \sum_{i=1}^{N} X_i Y_i + N\bar{X}\bar{Y}$

$\Rightarrow -\hat{\beta}_1 N \bar{X}^2 + \hat{\beta}_1 \sum_{i=1}^{N} {X_i}^2 = \sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}$

$\Rightarrow \hat{\beta}_1 (\sum_{i=1}^{N} {X_i}^2 - N\bar{X}^2) = \sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}$

Now, solving for $\hat{\beta}_1$ gives us:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}}{\sum_{i=1}^{N} {X_i}^2 - N\bar{X}^2} \qquad (8)$$

Doesn't quite look like the formula from the class, right? Well, let's just use a couple more tricks. We can show that $\sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y} = \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})$. We can also easily derive that $\sum_{i=1}^{N} {X_i}^2 - N\bar{X}^2 = \sum_{i=1}^{N} (X_i - \bar{X})^2$.

First, let's work with $\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}$

Taking the L.H.S:

$\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})$

$\Rightarrow \sum_{i=1}^{N} (X_i Y_i - X_i \bar{Y} - \bar{X} Y_i + \bar{X}\bar{Y})$

$\Rightarrow \sum_{i=1}^{N} X_i Y_i - \bar{Y} \sum_{i=1}^{N} X_i - \bar{X} \sum_{i=1}^{N} Y_i + \sum_{i=1}^{N} \bar{X}\bar{Y}$ $\quad$ [Distributing summation operator]

$\Rightarrow \sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y} - N\bar{X}\bar{Y} + N\bar{X}\bar{Y}$ $\quad [\because \sum_{i=1}^{N} X_i = N\bar{X}; \sum_{i=1}^{N} Y_i = N\bar{Y}; \sum_{i=1}^{N} \bar{X}\bar{Y} = N\bar{X}\bar{Y}]$

$\Rightarrow \sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}$

Second, let's work with $\sum_{i=1}^{N}(X_i - \bar{X})^2 = \sum_{i=1}^{N} {X_i}^2 - N\bar{X}^2$

Taking the L.H.S:

$\sum_{i=1}^{N}(X_i - \bar{X})^2$

$\Rrightarrow \sum_{i=1}^{N}({X_i}^2 - 2X_i\bar{X} + \bar{X}^2)$ [$(X_i - \bar{X})^2$ formula]

$\Rrightarrow \sum_{i=1}^{N} {X_i}^2 - 2\bar{X}\sum_{i=1}^{N} X_i + \sum_{i=1}^{N} \bar{X}^2$ [Distributing summation operator]

$\Rrightarrow \sum_{i=1}^{N} {X_i}^2 - 2N\bar{X}^2 + N\bar{X}^2$ [$\because \sum_{i=1}^{N} X_i = N\bar{X}$]

$\Rrightarrow \sum_{i=1}^{N} {X_i}^2 - N\bar{X}^2$

Finally, we can now express $\hat{\beta}_1$ as:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \tag{9}$$

All done!

Note, other than the above, there are many ways to reach to this conclusion. One can also use matrix algebra to derive OLS parameters.