Derivation of OLS and the Method of Moments Estimators

In lecture and in section we set up the minimization problem that is the starting point for deriving the formulas for the OLS intercept and slope coefficient. We decided to minimize the sum squared of the vertical distance between our observed $y_i$ and the predicted $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} W = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (1)$$

To solve the minimization problem, we need to take the partial derivatives with respect to our parameters and set them equal to 0. This gives us:

$$\frac{\partial W}{\partial \hat{\beta}_0} = \sum_{i=1}^{N} -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (2)$$

$$\frac{\partial W}{\partial \hat{\beta}_1} = \sum_{i=1}^{N} -2x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (3)$$

Now we have a system of two equations and two unknowns ($\hat{\beta}_0$ and $\hat{\beta}_1$), so our task is to solve (2) and (3) using some algebra tricks and properties of summation. Let’s start with the first order condition for $\hat{\beta}_0$ (which is equation (2)). We can immediately divide both sides by -2 and write:

$$\sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (4)$$

This is the formula for the intercept that we were given in lecture! Now let’s consider solving for $\hat{\beta}_1$, which is a bit more tricky. First we cancel the -2 as before and rearrange to get:

$$\sum_{i=1}^{N} (x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) x_i - \hat{\beta}_1 x_i^2) = 0 \quad (5)$$

Note that the summation applies to everything in the above equation. We can distribute this sum to each term in the expression to get:

$$\sum_{i=1}^{N} (x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) x_i - \hat{\beta}_1 x_i^2) = 0 \quad (6)$$

Anything that is not indexed by $i$ can be pulled out in front of the sum, as we did above. We can again use the fact that $\sum_{i=1}^{N} x_i = N \bar{x}$, and then solve (7) for $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} x_i y_i - N \bar{x} \bar{y}}{\sum_{i=1}^{N} x_i^2 - N \bar{x}^2} \quad (8)$$
Doesn’t quite look like the formula from class, right? We have to use two more tricks to get it into the recognizable form. See the appendix to this derivation for proofs of the following facts:

\[
\sum_{i=1}^{N} (x_i y_i) - N \bar{x} \bar{y} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \tag{9}
\]

\[
\sum_{i=1}^{N} (x_i^2) - N \bar{x}^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{10}
\]

If we plug these into equation (8), we get something much more familiar:

\[
\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{11}
\]

All done!

Now, we can also derive the same formulas using the sample moment conditions discussed in lecture. A sample moment condition is the sample counterpart of a population moment; for example, \( \mathbb{E}[u] = 0 \) is a population moment, and its sample counterpart is that \( \sum_{i=1}^{N} \hat{u}_i = 0 \). The two conditions we had were:

\[
\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{12}
\]

\[
\frac{1}{N} \sum_{i=1}^{N} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{13}
\]

Take a close look at equations (12) and (13), and then compare them to equations (2) and (3). The only differences between (2) and (12) are the factors \( \frac{1}{N} \) and \(-2\). If we divide both sides of (2) by \(-2\) (like we did above), and multiply both sides of (12) by \( N \), the equations would be exactly the same! We can do this to equations (3) and (13), to see that they are exactly the same as well. Thus, these two systems of two equations are equivalent, so their solutions are also equivalent. If we solve (12) and (13) for \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \), we would get the same formula as if we solved (2) and (3).

**Appendix**

We’d like to show that \( \sum_{i=1}^{N} (x_i y_i) - N \bar{x} \bar{y} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \). First notice that we can expand the right hand side, \( \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{N} (x_i y_i - \bar{xy}_i + \bar{yx}_i - \bar{xy}) \). Now let’s start with the left hand side of this equation:

\[
\sum_{i=1}^{N} (x_i y_i) - N \bar{x} \bar{y} = \sum_{i=1}^{N} (x_i y_i - \bar{xy}_i + \bar{yx}_i + \bar{yx}) - N \bar{xy} \tag{14}
\]

\[
= \sum_{i=1}^{N} (x_i y_i - \bar{xy}_i + \bar{yx}_i) + \sum_{i=1}^{N} (\bar{xy}_i + \bar{yx}_i) - N \bar{xy} \tag{15}
\]
Let’s focus on the second summation, and once again use the fact that \( \sum_{i=1}^{N} x_i = N \bar{x} \):

\[
\begin{align*}
\sum_{i=1}^{N} (x_i y_i - \bar{x} y_i) &= N \bar{x} \bar{y} + N \bar{y} \bar{x} - N \bar{y} \\
\sum_{i=1}^{N} (x_i y_i - \bar{x} y_i) &= N \bar{x} \bar{y} \\
\sum_{i=1}^{N} (x_i y_i - \bar{x} y_i + \bar{y}) &= N \bar{x} y + \bar{y} y \\
\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{N} (x_i - \bar{x})^2 \bar{y}
\end{align*}
\] (16) (17) (18) (19)

Finally, we’d like to show that \( \sum_{i=1}^{N} (x_i^2) - N \bar{x}^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 \). Again, notice that we can expand the right hand side:

\[
\sum_{i=1}^{N} (x_i - \bar{x})^2 = \sum_{i=1}^{N} (x_i^2 - 2\bar{x}x_i + \bar{x}^2)
\] (20)

Now if we distribute the summation, pull out all constant terms, and use (once and for all) the fact that \( \sum_{i=1}^{N} x_i = N \bar{x} \), we get:

\[
\begin{align*}
\sum_{i=1}^{N} x_i^2 - 2\bar{x} \sum_{i=1}^{N} x_i &+ \sum_{i=1}^{N} \bar{x}^2 = \sum_{i=1}^{N} (N \bar{x}^2) + N \bar{x}^2 \\
\sum_{i=1}^{N} x_i^2 - 2\bar{x} (N \bar{x}) &+ \bar{x}^2 = \sum_{i=1}^{N} (2N \bar{x}^2 + N \bar{x}^2) \\
\sum_{i=1}^{N} x_i^2 - 2N \bar{x}^2 &+ \bar{x}^2 = \sum_{i=1}^{N} (x_i^2 - N \bar{x}^2)
\end{align*}
\] (21) (22) (23) (24)

And we’re done!