ECON 2007 - Term 1
Additional Notes and Proofs on OLS

Alexandros Theloudis*

November 7, 2015

We are interested in estimating the relationship between x and y in the population. We assume that there is a linear association between the two (Assumption *SLR.1*):

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

where β_0 and β_1 are the parameters of interest and u is an error term. The population is not directly observable as it is assumed to be infinite; however we can still learn a lot about the aforementioned relation by hinging on random samples from the population. A random sample can be denoted as $\{(x_i, y_i) : 1, \ldots, n\}$; this notation implies that we have n observations (x_i, y_i) where $\{x_i, y_i\}$ are independently drawn from the same aforementioned population (we say that $\{x_i, y_i\}$ are *i.i.d. - independent and identically distributed*). This is Assumption *SLR.2*.

The error terms are assumed to have zero conditional mean: conditional on x the expected value of u in the population is 0. This can be written as $\mathbb{E}(u|x) = 0$; it implies that no x conveys any information about u (Assumption *SLR.4*).

The question is: how can we get β_0 and β_1? Well, we can't unless we have infinite amount of information. However, given a random sample of observations, the OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are BLUE: Best Linear Unbiased Estimates of the items of interest β_0 and β_1. How do we get $\hat{\beta}_0$ and $\hat{\beta}_1$?

1 Derivation of OLS Estimators

There are two equivalent ways to obtain $\hat{\beta}_0$ and $\hat{\beta}_1$ given the 3 assumptions above. The first way is the *Method of Moments*; the second is the *Sum of Squared Residuals*.

1.1 Method of Moments

$\mathbb{E}[u|x] = 0$ implies that $\mathbb{E}[u] = 0$ and $\mathbb{E}[xu] = 0$[1]. From (1) we can write:

$$\mathbb{E}[u] = \mathbb{E}[y - \beta_0 - \beta_1 x] = 0$$
$$\mathbb{E}[xu] = \mathbb{E}[x(y - \beta_0 - \beta_1 x)] = 0$$

These equations hold in the population. Their sample analogues are:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$
$$\frac{1}{n}\sum_{i=1}^{n}\left(x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)\right) = 0$$

*Alexandros Theloudis: Department of Economics, University College London, Gower Street, WC1E 6BT London (email: alexandros.theloudis.10@ucl.ac.uk).

[1] From the Law of Iterated Expectations: $\mathbb{E}(u) = \mathbb{E}[\mathbb{E}(u|x)] = \mathbb{E}[0] = 0$. Also: $\mathbb{E}(xu) = \mathbb{E}[\mathbb{E}(xu|x)] = \mathbb{E}[x\mathbb{E}(u|x)] = \mathbb{E}[x0] = 0$

where the summation over i implies summation over all sample observations.

Let's take the first equation:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)&=0\\
\frac{1}{n}\sum_{i=1}^{n}y_i-\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_1 x_i&=n\frac{1}{n}\hat{\beta}_0\\
\frac{1}{n}\sum_{i=1}^{n}y_i-\hat{\beta}_1\frac{1}{n}\sum_{i=1}^{n}x_i&=\hat{\beta}_0
\end{aligned}$$

(as $\hat{\beta}_1$ is just a constant number it can exit the summation)

$$\bar{y}-\hat{\beta}_1\bar{x}=\hat{\beta}_0 \tag{2}$$

To reach the last line, we make use of the formula for calculating a sample average: for any generic variable z: $\bar{z}=\frac{1}{n}\sum_{i=1}^{n}z_i$.

Working on the second equation, we can write:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(x_i(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)\right)&=0\\
\overset{(2)}{\Leftrightarrow}\ \frac{1}{n}\sum_{i=1}^{n}\left(x_i(y_i-(\bar{y}-\hat{\beta}_1\bar{x})-\hat{\beta}_1 x_i)\right)&=0\\
\sum_{i=1}^{n}\left(x_i y_i-x_i\bar{y}+x_i\hat{\beta}_1\bar{x}-\hat{\beta}_1 x_i^2\right)&=0\\
\sum_{i=1}^{n}\left(x_i(y_i-\bar{y})-\hat{\beta}_1 x_i(x_i-\bar{x})\right)&=0\\
\sum_{i=1}^{n}\hat{\beta}_1 x_i(x_i-\bar{x})&=\sum_{i=1}^{n}x_i(y_i-\bar{y})\\
\hat{\beta}_1\sum_{i=1}^{n}x_i(x_i-\bar{x})&=\sum_{i=1}^{n}x_i(y_i-\bar{y})\\
\hat{\beta}_1&=\frac{\sum_{i=1}^{n}x_i(y_i-\bar{y})}{\sum_{i=1}^{n}x_i(x_i-\bar{x})}
\end{aligned} \tag{3}$$

Expression (3) is the OLS estimator $\hat{\beta}_1$ if $\sum_{i=1}^{n}x_i(x_i-\bar{x})\neq 0$. This is guaranteed by Assumption *SLR.3*. Substituting (3) into (2) we can get the OLS estimator $\hat{\beta}_0$. Having obtained $\hat{\beta}_0$ and $\hat{\beta}_1$ we can then get:

- Predicted/fitted values: $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1 x_i$
- OLS regression line: $\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x$
- Residuals: $\hat{u}_i=y_i-\hat{y}_i$

1.2 Sum of Squared Residuals

An alternative (but equivalent) way to obtain the OLS estimators is by minimizing the sum of squared residuals (SSR hereafter). What really is a residual $\hat{u}_i$? Think about observation i with (x_i, y_i); for x_i the aforementioned OLS regression line predicts a y-value of $\hat{y}_i$. How far is $\hat{y}_i$ from the actual y_i? This information is given by $\hat{u}_i$! In other words, $\hat{u}_i$ is the vertical distance between

y_i and $\hat{y}_i$; it can be both positive or negative depending on whether the actual point lies above or below the OLS regression line.

If we want our OLS regression line to fit the data well, then we must minimize the distances $\hat{u}_i$, $\forall i$. How do we do that? One way would be to minimize the sum of residuals $\sum_{i=1}^{n} \hat{u}_i$. But that sum is by definition equal to 0. Instead we can minimize the SSR: the smaller this sum is, the closer the OLS regression line is to our sample observations.

Analytically:

$$\min \sum_{i=1}^{n} \hat{u}_i^2 \equiv \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \equiv \min \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

We need to find $\hat{\beta}_0$ and $\hat{\beta}_1$ so that the above SSR is minimized. Assuming that the above function is well behaved, we will derive the first order conditions with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and set them equal to 0. For convenience we should open up the above expression:

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 = \sum_{i=1}^{n} \left(y_i^2 + (\hat{\beta}_0 + \hat{\beta}_1 x_i)^2 - 2y_i(\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)$$

The first order conditions are:

- {with respect to $\hat{\beta}_0$}:

$$\begin{aligned} \sum_{i=1}^{n} 2\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) - \sum_{i=1}^{n} 2y_i &= 0 \\ n\hat{\beta}_0 + \hat{\beta}_1 n\bar{x} &= n\bar{y} \\ \bar{y} - \hat{\beta}_1 \bar{x} &= \hat{\beta}_0 \end{aligned}$$

  Notice that we now actually reached equation (2) above.

- {with respect to $\hat{\beta}_1$}:

$$\begin{aligned} \sum_{i=1}^{n} 2\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) x_i - \sum_{i=1}^{n} 2y_i x_i &= 0 \\ \overset{(2)}{\Leftrightarrow} \sum_{i=1}^{n} \left(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i\right) x_i &= \sum_{i=1}^{n} y_i x_i \\ \sum_{i=1}^{n} \bar{y} x_i + \sum_{i=1}^{n} \hat{\beta}_1 x_i (x_i - \bar{x}) &= \sum_{i=1}^{n} y_i x_i \\ \sum_{i=1}^{n} \hat{\beta}_1 x_i (x_i - \bar{x}) &= \sum_{i=1}^{n} x_i (y_i - \bar{y}) \\ \hat{\beta}_1 \sum_{i=1}^{n} x_i (x_i - \bar{x}) &= \sum_{i=1}^{n} x_i (y_i - \bar{y}) \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})} \end{aligned}$$

  Now notice that this is actually the equation we found for $\hat{\beta}_1$ in (3). It should now be obvious that the two ways of obtaining the OLS estimators are equivalent. As before, to obtain an equation for $\hat{\beta}_0$, one only needs to replace $\hat{\beta}_1$ in (2) with the expression in (3).

2 Unbiasedness

Every time we draw a new random sample from the population, the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ will be different. The question we are asking now is: does the expected value of these estimates equal the unknown true value for β_0 and β_1 in the population or not? The answer is yes. It will turn out that $\mathbb{E}[\hat{\beta}_0] = \beta_0$ and $\mathbb{E}[\hat{\beta}_1] = \beta_1$ and thus we will say that the OLS estimates are *unbiased*.

2.1 Proof for $\hat{\beta}_1$

Following the lecture notes, we will set $s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$. From (3) we have:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i(y_i - \bar{y})}{\sum_{i=1}^n x_i(x_i - \bar{x})}$$

Notice that:

$$\sum_{i=1}^n \bar{x}(x_i - \bar{x}) = \sum_{i=1}^n \bar{x}x_i - \sum_{i=1}^n \bar{x}\bar{x} = n\bar{x}\bar{x} - n\bar{x}\bar{x} = 0$$

Similarly we can show that $\sum_{i=1}^n \bar{y}(x_i - \bar{x}) = 0$ and $\sum_{i=1}^n \bar{x}(y_i - \bar{y}) = 0$. Working on (3) we get:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i(y_i - \bar{y})}{\sum_{i=1}^n x_i(x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i(y_i - \bar{y}) - \sum_{i=1}^n \bar{x}(y_i - \bar{y})}{\sum_{i=1}^n x_i(x_i - \bar{x}) - \sum_{i=1}^n \bar{x}(x_i - \bar{x})} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{s_x^2} \\
\overset{(1)}{\Leftrightarrow} \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{s_x^2} \\
&= \beta_0 \underbrace{\frac{\sum_{i=1}^n (x_i - \bar{x})}{s_x^2}}_{=0} + \beta_1 \underbrace{\frac{\sum_{i=1}^n (x_i - \bar{x})x_i}{s_x^2}}_{=1} + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{s_x^2} \\
&= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{s_x^2}
\end{aligned}$$

Now let's take expectations on both side conditional on the data $x_1, x_2, \ldots, x_n$ we have available:

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_1 | x_1, \ldots, x_n] &= \mathbb{E}\left[\beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{s_x^2} \ | x_1, \ldots, x_n\right] \\
&= \mathbb{E}\left[\beta_1 | x_1, \ldots, x_n\right] + \mathbb{E}\left[\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{s_x^2} \ | x_1, \ldots, x_n\right]
\end{aligned}$$

(because the expectation operator is linear)

$$= \beta_1 + \mathbb{E}\left[\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{s_x^2} \ | x_1, \ldots, x_n\right]$$

(because β_1 is just a number)

$$= \beta_1 + s_x^{-2} \sum_{i=1}^n (x_i - \bar{x}) \mathbb{E}\left[u_i \ | x_1, \ldots, x_n\right]$$

(because conditional on $x_1, \ldots, x_n$ the expected value of any function of x_i is the function itself)

$$= \beta_1 + s_x^{-2} \sum_{i=1}^n (x_i - \bar{x}) \mathbb{E}\left[u_i \ | x_i\right]$$

$$\Leftrightarrow \quad \mathbb{E}[\hat{\beta}_1 | x_1, \ldots, x_n] = \beta_1 \tag{4}$$

(because of *SLR.4*). We are not done though. We have to prove that the un-conditional expectation of $\hat{\beta}_1$ is equal to β_1. By the Law of Iterated Expectations:

$$\mathbb{E}\left[\hat{\beta}_1\right] = \mathbb{E}\left[\mathbb{E}[\hat{\beta}_1|x_1,\ldots,x_n]\right] \overset{(4)}{=} \mathbb{E}\left[\beta_1\right] = \beta_1$$

2.2 Proof for $\hat{\beta}_0$

From (2) we have that

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Notice that from (1) $\bar{y} = \beta_0 + \beta_1\bar{x} + \bar{u}$, so that (2) now becomes:

$$\begin{aligned}\hat{\beta}_0 &= \beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x} \\ &= \beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}\end{aligned}$$

Now let's take expectations on both side conditional on the data $x_1, x_2, \ldots, x_n$ we have available:

$$\begin{aligned}\mathbb{E}[\hat{\beta}_0|x_1,\ldots,x_n] &= \mathbb{E}\left[\beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}|x_1,\ldots,x_n\right] \\ &= \mathbb{E}\left[\beta_0|x_1,\ldots,x_n\right] + \mathbb{E}\left[(\beta_1 - \hat{\beta}_1)\bar{x}|x_1,\ldots,x_n\right] + \mathbb{E}\left[\bar{u}|x_1,\ldots,x_n\right]\end{aligned}$$

(because the expectation operator is linear)

$$= \beta_0 + \bar{x}\mathbb{E}\left[(\beta_1 - \hat{\beta}_1)|x_1,\ldots,x_n\right] + \mathbb{E}\left[\bar{u}|x_1,\ldots,x_n\right]$$

(because β_0 is just a number; also conditional on $x_1, \ldots, x_n$ $\bar{x}$ is non-random)

$$= \beta_0 + \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} u_i|x_i\right]$$

(because it has already been proved that $\mathbb{E}\left[\hat{\beta}_1\right] = \beta_1$)

$$\begin{aligned}&= \beta_0 + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[u_i|x_i\right] \\ \Leftrightarrow \quad \mathbb{E}[\hat{\beta}_0|x_1,\ldots,x_n] &= \beta_0 \qquad (5)\end{aligned}$$

(from *SLR.4*). As before, by the Law of Iterated Expectations one can show that $\mathbb{E}[\hat{\beta}_0] = \beta_0$.

3 Variance of OLS estimator

Here we impose an additional assumption:

$$Var[u_i|x_i] = \sigma^2, \qquad \forall\ i$$

This is Assumption *SLR.5* according to the lecture notes. Recall from before that:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{s_x^2}$$

We are now interested in the conditional variance $Var\left[\hat{\beta}_1|x_1,\ldots,x_n\right]$:

$$Var\left[\hat{\beta}_1|x_1,\ldots,x_n\right] = Var\left[\beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{s_x^2}|x_1,\ldots,x_n\right]$$

$$= Var\left[\beta_1|x_1,\ldots,x_n\right] + Var\left[\frac{\sum_{i=1}^n (x_i-\bar{x})u_i}{s_x^2}|x_1,\ldots,x_n\right]$$
$$+ 2Cov\left[\beta_1, \frac{\sum_{i=1}^n (x_i-\bar{x})u_i}{s_x^2}|x_1,\ldots,x_n\right]$$

(because by the properties of the variance, $Var(a+b) = Var(a) + Var(b) + 2Cov(a,b)$)

$$= Var\left[\frac{\sum_{i=1}^n (x_i-\bar{x})u_i}{s_x^2}|x_1,\ldots,x_n\right]$$

(because β_1 is just a number so it doesn't vary or covary with any random variable)

$$= (\frac{1}{s_x^2})^2 \sum_{i=1}^n (x_i-\bar{x})^2 Var\left[u_i|x_1,\ldots,x_n\right]$$

(because conditional on $x_1,\ldots,x_n$, any function of x is treated as a constant number; by the properties of the variance: $Var(c\cdot z) = c^2 Var(z)$ where c is a constant and z is a random variable)

$$= (\frac{1}{s_x^2})^2 \sum_{i=1}^n (x_i-\bar{x})^2 Var\left[u_i|x_i\right]$$
$$= (\frac{1}{s_x^2})^2 \sum_{i=1}^n (x_i-\bar{x})^2 \sigma^2$$

(because of Assumption *SLR.5*)

$$\Leftrightarrow \quad Var\left[\hat{\beta}_1|x_1,\ldots,x_n\right] = \frac{1}{s_x^2}\sigma^2 \tag{6}$$

as $\sum_{i=1}^n (x_i-\bar{x})^2 = s_x^2$. The derivation of the conditional variance of $\hat{\beta}_0$ follows a same logic. *Can you derive it?*

4 Goodness-of-Fit (R^2)

How much of the variation in y is explained by variation in x? If we are interested in that question, we are after the coefficient of determination R^2. R^2 gives us a sense of the *goodness-of-fit* of our regression; i.e. it informs us about what fraction of the variation in y is due to variation in x.

What do we mean by saying variation in y? The squared distance of y_i from the sample mean $\bar{y}$ informs us about the spread of y_i and is denoted by $\sum_{i=1}^n (y_i-\bar{y})^2$. Notice that if we divide this expression by $n-1$ we get the sample variance for y_i. For what follows we will work on $\sum_{i=1}^n (y_i-\bar{y})^2$. As $y_i = \hat{u}_i + \hat{y}_i$ we can write:

$$\sum_{i=1}^n (y_i-\bar{y})^2 = \sum_{i=1}^n (\hat{u}_i + \hat{y}_i - \bar{y})^2$$
$$= \sum_{i=1}^n \left(\hat{u}_i^2 + (\hat{y}_i-\bar{y})^2 + 2\hat{u}_i(\hat{y}_i-\bar{y})\right)$$
$$= \sum_{i=1}^n \hat{u}_i^2 + \sum_{i=1}^n (\hat{y}_i-\bar{y})^2 + 2\sum_{i=1}^n \hat{u}_i(\hat{y}_i-\bar{y}) \tag{7}$$

The last expression in (7) is 0. To see why, we can use $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ and write:

$$\sum_{i=1}^n \hat{u}_i(\hat{y}_i-\bar{y}) = \sum_{i=1}^n \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \sum_{i=1}^{n} \hat{u}_i \bar{y} \\
&= \hat{\beta}_0 \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{n} \hat{u}_i x_i - \bar{y} \sum_{i=1}^{n} \hat{u}_i \\
&= 0
\end{aligned}
$$

where the last line comes from our sample moment conditions $\sum_{i=1}^{n} \hat{u}_i = 0$ and $\sum_{i=1}^{n} \hat{u}_i x_i = 0$. Hence: $\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y}) = 0$. Going back to (7) we now have:

$$
\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} \hat{u}_i^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2
$$
$$
SST = SSR + SSE
$$

where:

- SST: Total sum of squares (variation in y)
- SSR: Sum of squared residuals (unexplained variation in y)
- SSE: Explained sum of squares (explained variation in y by x)

Dividing across by SST we get:

$$
\begin{aligned}
1 &= \frac{SSR}{SST} + \frac{SSE}{SST} \\
\frac{SSE}{SST} &= 1 - \frac{SSR}{SST} \\
\Leftrightarrow \quad R^2 &\equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}
\end{aligned}
$$

R^2 is the fraction of sample variation in y that is explained by x.