OLS Regression Models

Daniel Cullen

October 4, 2017

1 OLS and the "Classic" Assumptions

The "classic" OLS assumptions are necessary to prove unbiasedness and calculate the variance of the OLS estimator. Note that we did not have to assume any of these prior to our derivation of the OLS estimator.

$$
\begin{array}{rll}
\underline{A1}: & y_i = \beta x_i + \varepsilon_i \text{ is the true DGP} & \\
\underline{A2}: & x_i \text{ is nonrandom} & \\
\underline{A3}: & \mathbb{E}[\varepsilon_i] = 0 \ \forall i & \\
\underline{A4}: & Var(\varepsilon_i) = \sigma^2 \ \ i = 1, \ldots, n & (\text{"homoskedastic"}) \\
 & Cov(\varepsilon_i, \varepsilon_j) = 0 \ \forall i \neq j & (\text{"no correlation"}) \\
\text{(sometimes)} \ \underline{A5}: & \varepsilon_i \sim Normally &
\end{array}
$$

You might wonder what "sometimes" means. If we make assumption $A5$, then we'll have

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right).$$

If we don't then we'll have (using the Central Limit Theorem)

$$\widehat{\beta} \sim^A N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right).$$

The difference is whether or not we think our sample size is large enough for the distribution to converge. If we don't think so, assuming that the errors are normal will do the trick. If we do think so, then we might avoid making the assumption altogether. Now lets calculate the mean and variance of the OLS estimator. First, the mean.

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\beta}^{OLS}\right] &= \mathbb{E}\left[\frac{\sum\limits_i y_i x_i}{\sum\limits_i x_i^2}\right] \\
&= \mathbb{E}\left[\frac{\sum\limits_i (\beta x_i + \varepsilon_i) x_i}{\sum\limits_i x_i^2}\right] && \text{(A1)} \\
&= \mathbb{E}\left[\beta + \frac{\sum\limits_i x_i \varepsilon_i}{\sum\limits_i x_i^2}\right] \\
&= \beta + \frac{\sum\limits_i x_i \mathbb{E}[\varepsilon_i]}{\sum\limits_i x_i^2} && \text{(A2)} \\
&= \beta && \text{(A3)}
\end{aligned}
$$

That is, the OLS estimator is unbiased. Now we find the variance.

$$
\begin{aligned}
Var\left(\widehat{\beta}^{OLS}\right) &= Var\left(\frac{\sum\limits_i x_i y_i}{\sum\limits_i x_i^2}\right) \\
&= Var\left(\frac{\sum\limits_i x_i(\beta x_i + \varepsilon_i)}{\sum\limits_i x_i^2}\right) && \text{(A1)} \\
&= Var\left(\beta + \frac{\sum\limits_i x_i \varepsilon_i}{\sum\limits_i x_i^2}\right) \\
&= Var\left(\frac{\sum\limits_i x_i \varepsilon_i}{\sum\limits_i x_i^2}\right) && \text{(the variance of a constant is zero)} \\
&= \frac{1}{\left(\sum\limits_i x_i^2\right)^2} Var\left(\sum_i x_i \varepsilon_i\right) && \text{(pull constant out and square)} \\
&= \frac{1}{\left(\sum\limits_i x_i^2\right)^2} \left[\sum_i x_i^2 Var(\varepsilon_i) + 2\sum_{i \neq j} x_i x_j Cov(\varepsilon_i, \varepsilon_j)\right] && \text{(apply general formula; A2)} \\
&= \frac{1}{\left(\sum\limits_i x_i^2\right)^2} \left[\sigma^2 \sum_i x_i^2\right] && \text{(A4)} \\
&= \frac{\sigma^2}{\sum\limits_i x_i^2}
\end{aligned}
$$

2 Population Value of the OLS parameter

Given the model,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

we can derive the population value of β_1.

$$\begin{aligned}
Cov(x_i, y_i) &= Cov(x_i, \beta_0 + \beta_1 x_i + \varepsilon_i) && \text{(plugging in for } y_i\text{)}\\
Cov(x_i, y_i) &= Cov(x_i, \beta_0) + Cov(x_i, \beta_1 x_i) + Cov(x_i, \varepsilon_i) && \text{(property of } Cov\text{)}\\
Cov(x_i, y_i) &= 0 + \beta_1 Cov(x_i, x_i) + Cov(x_i, \varepsilon_i) && (\beta_0 \text{ \& } \beta_1 \text{ are constants})\\
Cov(x_i, y_i) &= \beta_1 Var(x_i) + Cov(x_i, \varepsilon_i) && \text{(property of } Cov\text{)}\\
Cov(x_i, y_i) &= \beta_1 Var(x_i) + 0 && \text{(assumption)}\\
\implies \beta_1 &= \frac{Cov(x_i, y_i)}{Var(x_i)}
\end{aligned}$$

3 Why do we use OLS?

We might ask ourselves, why are we using OLS instead of some other estimator. Restricting ourselves might seem severe. After all, we often have reason to believe that variables are related in *nonlinear* ways. However, we should not fret. That OLS is a *linear* model simply means that it is linear in the coefficients (not the variables). Consider the following:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \varepsilon_i.$$

Many people would say that this is not a linear regression. However, what if we simply called $x_{2i} = x_{1i}^2$ (the computer doesn't really care what we call the variables; all it does it take the numbers and doesn't interpret whether or not they are nonlinear). Thus we can write our problem as

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i.$$

In fact we can do this for any transformation of our variables (*logs*, for instance).

In addition to the above, there is a very important theorem that explains why OLS might be the preferred. Colloquially, if the classic OLS assumptions introduced earlier hold, OLS is the best in the class of linear estimators.

Gauss-Markov Theorem: If $\mathbb{E}[\varepsilon_i] = 0$, $Var(\varepsilon_i) = \sigma^2 < \infty \; \forall \, i$, and $Cov(\varepsilon_i, \varepsilon_j) = 0 \; forall \; i \neq j$, then the Ordinary Least Squares estimator is the Best Linear Unbiased Estimator (BLUE), where best means "lowest variance."