

A Rigorous Look at Linear Regression*

Ari Boyarsky[†]

October 31, 2018

1 The Model

Consider the classical OLS model:

$$y_i = x_i\beta + u_i \quad (1)$$

Where $i \in \{1, \dots, N\}$.

In this model we call y_i our outcome/response/dependent variable, $x_i = (x_{i,1}, \dots, x_{i,K})$ is our predictor/explanatory/independent/regressor variable. u_i is our residual value. We attempt to estimate our coefficient, β . In matrix form,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}^1 \quad (2)$$

In this case X is $N \times K$, β is $K \times 1$, u is $N \times 1$, and y is $N \times 1$.

1.1 Ordinary Least Squares

One process by which we estimate the β terms is called ordinary least squares. To provide some motivation for this we will consider some interpretations of linear regression and return to the OLS estimator.

1.1.1 Linear Regression as a Conditional Expectation

Let $\mathbb{E}[Y|X] = X'\beta$ then, $u = Y - \mathbb{E}[Y|X]$. Notice, that this implies a linear relationship. If this the relationship is indeed linear and we satisfy the $\mathbb{E}[u|X] = 0$ assumption then our "approximation" would be exact that is $\mathbb{E}[Xu] = 0$

1.1.2 Linear Regression as a Linear Approximation

Suppose that we have moment existence. Then,

$$\hat{\beta} = \min_{\hat{\beta} \in \mathbb{R}^{k+1}} \mathbb{E}[(Y|X] - X'\hat{\beta})^2] \iff \min_{\hat{\beta} \in \mathbb{R}^{k+1}} \mathbb{E}[(Y - X'\hat{\beta})^2] \quad (3)$$

*These notes are based on classes taught by Azeem Shaikh, and Stephane Bonhome at the University of Chicago. All mistakes are my own.

[†]aboyarsky@uchicago.edu

¹Forgive us if we do not always note our matrices in boldface.

1.1.3 Linear Regression as a Causal Model

Let $Y = g(X, u)$, the effect of x on y is given by $D_x g(X, u)$, if we add a β_0 term (sometimes denoted as α) we can have that $\mathbb{E}[Xu] = 0$. Then, assuming the same context that we had before and assuming moment existence we can solve for β .

To do this we need to make one more assumption that is $\mathbb{E}[X'X]$ is full rank, that is it is invertible.

Claim. (The OLS estimator) Then, we claim that:

$$\beta = \mathbb{E}[X'X]^{-1} \mathbb{E}[XY] \quad (4)$$

Proof. Recall, by assumption $\mathbb{E}[X|u] = 0$, $u = Y - X'\beta$ such that:

$$\mathbb{E}[X'Y] = \mathbb{E}[X'X]\beta \implies \beta = \mathbb{E}[X'X]^{-1} \mathbb{E}[X'Y]$$

□

Remark. Thus under the above assumptions, we are presented with the familiar OLS estimator.

Next, we present the proof of unbiasedness of the OLS estimator.

Claim. β is unbiased.

Proof.

$$\begin{aligned} \mathbb{E}[\beta|X] &= \mathbb{E}[\mathbb{E}[X'X]^{-1} \mathbb{E}[X'Y]|X] \\ &= \mathbb{E}\left[\left(\sum_{i=0}^N x'_i x_i\right)^{-1} \sum_{i=0}^N x'_i y_i | X\right] \\ &= \left(\sum_{i=0}^N x'_i x_i\right)^{-1} \sum_{i=0}^N x'_i \mathbb{E}[y_i | X] \\ &= \left(\sum_{i=0}^N x'_i x_i\right)^{-1} \sum_{i=0}^N x'_i x_i \beta = \beta \end{aligned}$$

□

Remark. For now, take notice of the assumptions that we made to arrive at the above result. First, we assume $\mathbb{E}[y_i|X] = X'\beta$ which implies $\mathbb{E}[u_i|x_i] = 0$. We also assume $\mathbf{X}'\mathbf{X}$ is invertible. We will further formalize these assumptions later.

1.1.4 Asymptotic Properties of OLS

Claim. The OLS estimator is consistent.

Proof.

$$\begin{aligned}
\hat{\beta} &= \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-1} \frac{1}{N} \sum_{i=0}^N x'_i y_i \\
&= \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-1} \frac{1}{N} \sum_{i=0}^N x'_i (x_i \beta + u_i) \\
&= \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-1} \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \beta + \frac{1}{N} \sum_{i=0}^N x'_i u_i \right) \\
&= \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-1} \left(\frac{1}{N} \sum_{i=0}^N x'_i u_i \right) + \beta \\
\implies \sqrt{N}(\hat{\beta} - \beta) &= \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=0}^N x'_i u_i \right) \tag{5}
\end{aligned}$$

Equation (5) allows us to consider the asymptotic consistency of our estimator, we expect of course that it goes to 0 as our sample grows. This falls from our assumption on the orthogonality of x_i and u_i :

$$\text{plim}_{N \rightarrow \infty} \sqrt{N}(\hat{\beta} - \beta) \implies \frac{1}{\sqrt{N}} \sum_{i=0}^N x'_i u_i \rightarrow_p 0$$

□

We can also consider the limiting distribution of the OLS estimator, since

$$\text{plim}_{N \rightarrow \infty} \sqrt{N}(\hat{\beta} - \beta) \implies \frac{1}{\sqrt{N}} \sum_{i=0}^N x'_i u_i \rightarrow_p 0$$

Then using the Multivariate CLT, we see:

$$\sqrt{N}(\hat{\beta} - \beta) \rightarrow_d \mathcal{N}(0, V)$$

Now we can solve for V , notice, $V = \text{Var}(\sqrt{N}(\hat{\beta} - \beta)) = \mathbb{E}[\sqrt{N}(\hat{\beta} - \beta)^2] - \mathbb{E}[\sqrt{N}(\hat{\beta} - \beta)]^2 = \mathbb{E}[\sqrt{N}(\hat{\beta} - \beta)^2]$. Then,

$$\mathbb{E}[\sqrt{N}(\hat{\beta} - \beta)^2] = \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-2} \left(\frac{1}{\sqrt{N}} \sum_{i=0}^N x'_i u_i \right)^2$$

Which we commonly right as (This is the famous White Formula):

$$\left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=0}^N u_i^2 x'_i x_i \right) \left(\frac{1}{N} \sum_{i=0}^N x'_i x_i \right)^{-1}$$

In matrix form:

$$(X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

Notice, that this variance implies heteroskedastic errors under the homoskedasticity assumption this simplifies to $(X'X)^{-1}\sigma^2$.