# A Rigorous Look at Linear Regression*

Ari Boyarsky†

May 3, 2018

## 1 The Model

Consider the classical OLS model:

$$y_i = x_i\beta + u_i \tag{1}$$

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

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}^1 \tag{2}$$

In this case $X$ is $N \times K$, $\beta$ 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 $\beta$ 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, $u = Y - E[Y|X]$. 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,

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

*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

[1] Forgive us if we do not always note our matricies in boldface.

### 1.1.3 Linear Regression as a Causal Model

Let $Y = g(X, u)$, the effect of of $x$ on $y$ is given by $D_x g(X, u)$, if we add a $\beta_0$ term (sometimes denoted as $\alpha$) 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 $\beta$.

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:

$$\beta = \mathbb{E}[X'X]^{-1}\,\mathbb{E}[XY] \tag{4}$$

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

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

$\square$

**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.** $\beta$ is unbiased.

*Proof.*

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

$\square$

**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 **X'X** is invertible. We will further formalize these assumptions later.

### 1.1.4 Asymptotic Properties of OLS

**Claim.** The OLS estimator is consistent.

*Proof.*

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

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 $u_i$:

$$\operatorname*{plim}_{N\to\infty} \sqrt{N}(\hat{\beta} - \beta) \implies \frac{1}{\sqrt{N}}\sum_{i=0}^{N} x_i' u_i \to_p 0$$

□

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

$$\operatorname*{plim}_{N\to\infty} \sqrt{N}(\hat{\beta} - \beta) \implies \frac{1}{\sqrt{N}}\sum_{i=0}^{N} x_i' u_i \to_p 0$$

Then using the Multivariate CLT, we see:

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

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

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

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

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

In matrix form:

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

Notice, that this variance assumes heteroskedastic errors.