Estimating the Regression Model by Least Squares

Henrique Veras

PIMES/UFPE

Today

We now examine the Least Squares as an **estimator** of the parameters of the linear regression model.

We start by analysing the question "*why should we use least squares*"?

We will compare the LS estimator to other candidates based on their **statistical properties**:

1. Unbiasedness
2. Efficiency
3. Consistency

Population orthogonality

Recall assumption A3: $E[\varepsilon_i|\mathbf{X}] = 0$

By iterated expectations, $E[\varepsilon] = E_x E[\varepsilon_i|\mathbf{X}] = E_x[0] = 0$.

Also, $cov(\mathbf{x}, \varepsilon) = cov[\mathbf{x}, E[\varepsilon_i|\mathbf{X}]] = cov(\mathbf{x}, 0) = 0$, so $\mathbf{x}$ and ε are uncorrelated.

From these results we can find that

$$E[\mathbf{X}\mathbf{y}] = E[\mathbf{X}'\mathbf{X}]\beta$$

Population orthogonality

Now recall the FOC of the LS problem: $\mathbf{X'y} = \mathbf{X'Xb}$. Dividing both sides by n and writing it as a summation:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i y_i} = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i' x_i}\right)\mathbf{b}$$

Notice that this is the sample counterpart of the population condition $E[\mathbf{Xy}] = E[\mathbf{X'X}]\beta$.

Statistical Properties of the LS Estimator

An **estimator** is a strategy for using the sample data that are drawn from a population.

The **properties** of that estimator are descriptions of how it can be expected to behave when it is appied to a sample of data.

Unbiasedness

The least squares estimator is **unbiased** in every sample:

$$E[\mathbf{b}|\mathbf{X}] = \beta$$

Moreover,

$$E[\mathbf{b}] = E_x[E[\mathbf{b}|\mathbf{X}]] = E_x[\beta] = \beta$$

This is to say that the Least Squares estimator has expectation β.

Moreover, when we average this over the possible values of $\mathbf{X}$, the unconditional mean is also β.

Omitted Variable Bias (OVB)

Suppose the true population model is given by

$$\mathbf{y} = \mathbf{X}\beta + \gamma z + \varepsilon$$

If we estimate $\mathbf{y}$ on $\mathbf{X}$ only, without the *relevant* variable z, the estimator is

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \gamma z + \varepsilon) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\gamma z + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\end{aligned}$$

Omitted Variable Bias (OVB)

The expected value is given by

$$\begin{aligned} E[\mathbf{b}|\mathbf{X}, z] &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\gamma z \\ &= \beta + \mathbf{p}_{X.z}\gamma, \end{aligned}$$

where $\mathbf{p}_{X.z} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'z$. What does it represent? What happens if $\mathbf{X}$ and z are orthogonal?

Based on the FWL theorem and corollary 3.2.1, we can write

$$E[b_k|\mathbf{X}, z] = \beta_k + \gamma \left(\frac{cov(z, x_k|\text{all other x's})}{var(x_k|\text{all other x's})} \right)$$

An Example

Suppose we are interested in estimating the returns to education regression model below:

$$Income = \beta_0 + \beta_1 Educ + \beta_2 age + \beta_3 age^2 + \beta_4 Abil + \varepsilon$$

What is the sign of the bias if we estimate the model above without the $(unobserved)$ $Abil$?

An Example

The sign of the bias will depend on the signs of γ and $cov(z, x_k|\text{all other x's})$:

$$E[b_1|\mathbf{X}, z] = \beta_1 + \gamma \left(\frac{cov(Abil, Educ|age, age^2)}{var(Educ|age, age^2)} \right)$$

Thus, if $\gamma > 0$ and $cov(Abil, Educ|age, age^2) > 0$, b_1 will be biased upward:

$$E[b_1|\mathbf{X}, z] > \beta_1$$

Notice, however, that in some circumstances, the sign of the conditional covariance might not be obvious!

What happens if we include irrelevant variables instead?

Variance of the Least Squares Estimator

If Assumption A4 holds, the variance of the Least Squares estimator is given by

$$Var(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

If we wish to find a sample estimate of $Var(\mathbf{b}|\mathbf{X})$, we need to estimate the (unknown) population parameter σ^2.

Recall:

1. σ^2 is the variance of the error term: $\sigma^2 = E[\varepsilon_i^2|\mathbf{X}]$
2. e_i is the estimate of ε_i

Variance of the Least Squares Estimator

A natural estimator for σ^2 would then be $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$.

However, we would also need to estimate K parameters β, which would distort σ^2.

An **unbiased** estimator for σ^2 is

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-K}$$

Like $\mathbf{b}$, s^2 is unbiased unconditionally because

$$E[s^2] = E_X[E[s^2|\mathbf{X}]] = E_X[\sigma^2] = \sigma^2$$

Variance of the Least Squares Estimator

The **standard error of the regression** is $s = \sqrt{s^2}$.

The variance of the Least Squares Estimator can thus be estimated by

$$\hat{Var}(\mathbf{b}|X) = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

$\hat{Var}(\mathbf{b}|X)$ is the sample estimate of the *sampling variance* of the LS estimator.

Notice that the k-th diagonal element of this matrix is $[s^2(\mathbf{X}'\mathbf{X}_{kk})^{-1}]^{1/2}$, the standard error of the estimator b_k.

The Gauss-Markov Theorem

THEOREM 4.2 Gauss–Markov Theorem
In the linear regression model with given regressor matrix $\mathbf{X}$, (1) *the least squares estimator,* $\mathbf{b}$, *is the minimum variance linear unbiased estimator of* $\boldsymbol{\beta}$ and (2) *for any vector of constants* $\mathbf{w}$, *the minimum variance linear unbiased estimator of* $\mathbf{w}'\boldsymbol{\beta}$ *is* $\mathbf{w}'\mathbf{b}$.

Table of Contents