1

Regression Analysis I

Hüseyin Taştan[1]

[1]Department of Economics
Ph.D. Program - Advanced Econometrics
Yildiz Technical University

September 24, 2013

2

Linear Regression Model

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + u_i, \quad i = 1, \ldots, n$$

where y_i is ith observation on the dependent variable and x_{ij}, $j = 1, \ldots, k$, is the ith observation on th jth explanatory variable, and u_i is the ith random error term. There are k unknown parameters and n equations which can also be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & \ldots & x_{1k} \\ 1 & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

3

Linear Regression Model

Using matrix notation:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{12} & \ldots & x_{1k} \\ 1 & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \ldots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

and rewrite

$$\underbrace{\boldsymbol{y}}_{n\times 1} = \underbrace{\boldsymbol{X}}_{n\times k}\underbrace{\boldsymbol{\beta}}_{k\times 1} + \underbrace{\boldsymbol{u}}_{n\times 1}$$

4

Linear Regression Model

Equivalent notation: denote ith observation by

$$\boldsymbol{x}_i = \begin{bmatrix} 1 & x_{i2} & x_{i3} & \ldots & x_{ik} \end{bmatrix}^\top$$

Using this notation CLRM becomes

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + u_i, \quad i = 1, \ldots, n$$

Classical Regression Model

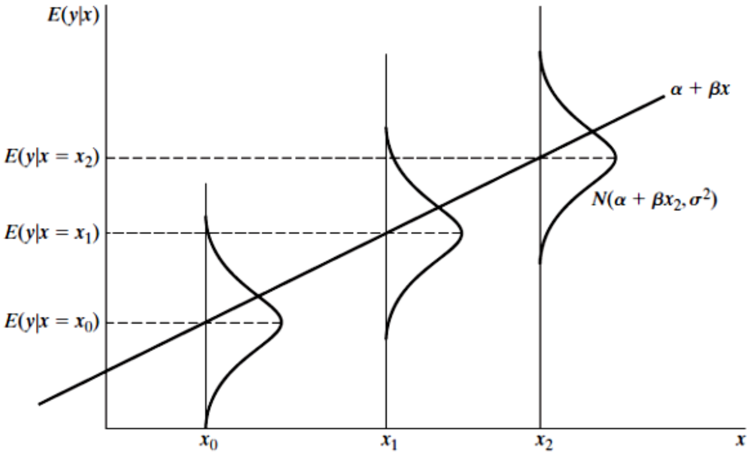

Figure 1 : Population Regression Function

Linear Regression Model

$$\underbrace{\boldsymbol{y}}_{n\times 1} = \underbrace{\boldsymbol{X}}_{n\times k}\underbrace{\boldsymbol{\beta}}_{k\times 1} + \underbrace{\boldsymbol{u}}_{n\times 1}$$

Assumptions of the CLRM (X variables assumed to be stochastic, not fixed in repeated samples)

1. Linearity in parameters: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$
2. rank($\boldsymbol{X}$) = k, (no perfect multicollinearity, columns of X are linearly independent)
3. E $[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{0}_{n\times 1}$, (zero conditional mean for the errors)
4. Var $[\boldsymbol{u}|\boldsymbol{X}] =$ E $\left(\boldsymbol{u}\boldsymbol{u}^\top\right) = \sigma^2\boldsymbol{I}_n$, (no heteroskedasticity, no autocorrelation)
5. $\boldsymbol{u}|\boldsymbol{X} \sim N\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n\right)$, (errors follow multivariate normal distribution)

Adding the last assumption the model is called Classical Normal Linear Regression Model.

Ordinary Least Squares (OLS) Estimator

Sample Regression Function (SRF):

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}$$

$\hat{\boldsymbol{\beta}}$ $k \times 1$: OLS estimator
$\hat{\boldsymbol{u}}$ ise $n \times 1$: residuals
OLS principle: minimize sum of squared residuals:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{b}} SSR(\boldsymbol{b})$$

where

$$SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}}$$

OLS Estimator

$$SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}\right)^2$$

Using the first notation

$$\min_{\hat{\boldsymbol{\beta}}} SSR(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}}$$

or the second notation:

$$\min_{\hat{\boldsymbol{\beta}}} SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}}\right)^2$$

From the first notation expanding SSR we get:

$$\begin{aligned} SSR(\hat{\boldsymbol{\beta}}) &= \hat{\boldsymbol{u}}^\top\hat{\boldsymbol{u}} \\ &= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{y}^\top\boldsymbol{y} - 2\hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{y} + \hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}} \end{aligned}$$

OLS Estimator

We need to evaluate the first derivative of SSR w.r.t. $\hat{\boldsymbol{\beta}}$. Note that the second term is a linear combination and the last is a quadratic form. In general, for a $k \times 1$ vector $\boldsymbol{z}$, and $k \times n$ matrix $\boldsymbol{A}$, and $k \times k$ matrix $\boldsymbol{B}$, we have:

$$\frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial \boldsymbol{z}} = \boldsymbol{A},$$

and

$$\frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial \boldsymbol{z}} = 2\boldsymbol{B}\boldsymbol{z}$$

OLS Estimator

For example, let $\boldsymbol{z} = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$ and

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 2 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\boldsymbol{z}^\top \boldsymbol{A} = \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 0 & 1 & -2 \\ 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} z_2 & z_1 + 2z_2 & -2z_1 \end{bmatrix}$$

First derivatives wrt z_1 and z_2:

$$\frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_1} = \begin{bmatrix} 0 & 1 & -2 \end{bmatrix} \text{ and } \frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_2} = \begin{bmatrix} 1 & 2 & 0 \end{bmatrix}$$

Collecting in a vector

$$\begin{aligned} \frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial \boldsymbol{z}} = &= \begin{bmatrix} \frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_1} \\ \frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_2} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & -2 \\ 1 & 2 & 0 \end{bmatrix} = \boldsymbol{A} \end{aligned}$$

OLS Estimator

Now take the derivative of $\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z}$ wrt $\boldsymbol{z}$ vector.

$$\begin{aligned} \boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z} &= \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\ &= 2z_1^2 + 2z_1 z_2 + 2z_2^2 \end{aligned}$$

1st derivatives:

$$\begin{aligned} \frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial \boldsymbol{z}} &= \begin{bmatrix} \frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial z_1} \\ \frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial z_2} \end{bmatrix} \\ &= \begin{bmatrix} 4z_1 + 2z_2 \\ 2z_1 + 4z_2 \end{bmatrix} \\ &= 2 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 2\boldsymbol{B}\boldsymbol{z} \end{aligned}$$

OLS Estimator

Back to SSR:

$$SSR(\hat{\boldsymbol{\beta}}) = \boldsymbol{y}^\top \boldsymbol{y} - 2\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{y} + \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

Hint: $\hat{\boldsymbol{\beta}} = \boldsymbol{z}$, $\boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{A}$ ve $\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{B}$

OLS FOC:

$$\frac{\partial SSR(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = -2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{0}_k$$

Normal equations:

$$\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{y}$$

Using the second CLRM assumption:

$$rank(\boldsymbol{X}) = rank(\boldsymbol{X}^\top \boldsymbol{X}) = k$$

we can find the inverse of $\boldsymbol{X}^\top \boldsymbol{X}$. Multiplying both sides of normal equations by $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ OLS estimator is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

OLS Estimator

Using the second notation:

$$\min_{\hat{\boldsymbol{\beta}}} SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}\right)^2$$

In this case FOC:

$$\frac{\partial SSR(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = -2 \sum_{i=1}^{n} \boldsymbol{x}_i \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}\right) = \mathbf{0}_k$$

and normal equations

$$\left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right) \hat{\boldsymbol{\beta}} = \sum_{i=1}^{n} \boldsymbol{x}_i y_i$$

OLS estimator can be written as:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right)^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i y_i \\ &\equiv (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \end{aligned}$$

OLS Estimator

$k \times k$ matrix $\boldsymbol{x}_i \boldsymbol{x}_i^\top$ has the following elements:

$$\boldsymbol{x}_i \boldsymbol{x}_i^\top = \begin{bmatrix} 1 \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix} \begin{bmatrix} 1 & x_{i2} & \dots & x_{ik} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_{i2} & x_{i3} & \dots & x_{ik} \\ x_{i2} & x_{i2}^2 & x_{i2}x_{i3} & \dots & x_{i2}x_{ik} \\ x_{i3} & x_{i3}x_{i2} & x_{i3}^2 & \dots & x_{i3}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ik} & x_{ik}x_{i2} & x_{ik}x_{i3} & \dots & x_{ik}^2 \end{bmatrix}$$

We have a sequence of n matrices defined analogously.

OLS Estimator

Summing these matrices

$$\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top = \begin{bmatrix} n & \sum x_{i2} & \sum x_{i3} & \dots & \sum x_{ik} \\ \sum x_{i2} & \sum x_{i2}^2 & \sum x_{i2}x_{i3} & \dots & \sum x_{i2}x_{ik} \\ \sum x_{i3} & \sum x_{i3}x_{i2} & \sum x_{i3}^2 & \dots & \sum n x_{i3}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ik} & \sum x_{ik}x_{i2} & \sum x_{ik}x_{i3} & \dots & \sum x_{ik}^2 \end{bmatrix} = \boldsymbol{X}^\top \boldsymbol{X}$$

This matrix is square, symmetric and positive definite. Similarly,

$$\begin{aligned} \sum_{i=1}^{n} \boldsymbol{x}_i y_i &= \sum_{i=1}^{n} \begin{bmatrix} 1 \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix} y_i = \begin{bmatrix} \sum yi \\ \sum x_{i2} y_i \\ \vdots \\ \sum x_{ik} y_i \end{bmatrix} \\ &= \boldsymbol{X}^\top \boldsymbol{y} \end{aligned}$$

OLS Estimation

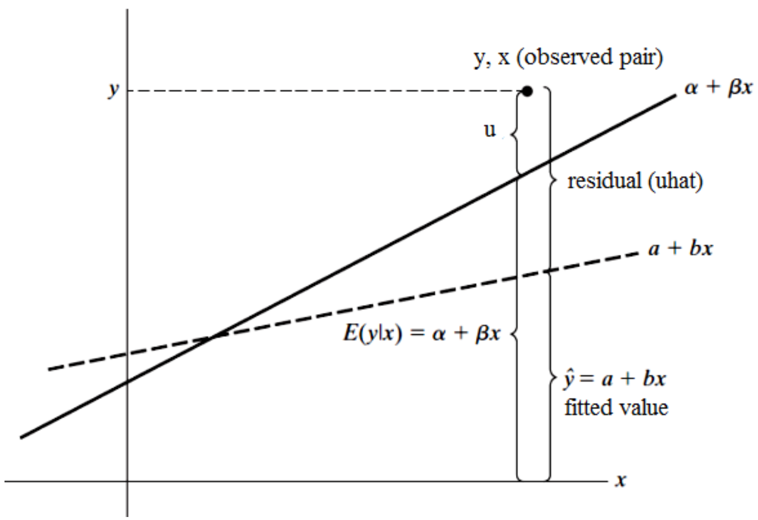

Figure 2 : Sample Regression Function (solid line is population regression function)

OLS Estimator

Example

Consider the following model where there are no explanatory variables:

$$y_i = \beta_1 + u_i, \quad i = 1, \dots, n$$

Now $\boldsymbol{X}$ matrix only has a column of ones $(n \times 1)$ which is denoted by $\boldsymbol{\imath}$:

$$\boldsymbol{\imath} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^\top = \boldsymbol{X}$$

Now the OLS estimator of β_1:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} = (\boldsymbol{\imath}^\top \boldsymbol{\imath})^{-1} \boldsymbol{\imath}^\top \boldsymbol{y}$$

OLS Estimator

Example (cont.d)

Note that

$$\boldsymbol{\imath}^\top \boldsymbol{\imath} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = n$$

and

$$\boldsymbol{\imath}^\top \boldsymbol{y} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} y_i$$

OLS estimator is simply

$$\beta_1 = n^{-1} \sum_{i=1}^{n} y_i \equiv \bar{y}$$

the arithmetic mean.

OLS Estimation

Example

One dummy variable and an intercept model:

$$y_i = \delta_0 + \delta_1 D_i + u_i, \quad i = 1, \dots, n$$

Data consist of 5 observations for simplicity:

$$\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \quad D_i = \begin{cases} 1, & \text{if } y_i \leq 3; \\ 0, & \text{otherwise.} \end{cases} \Leftrightarrow \boldsymbol{D} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Design matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Parameter vector: $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\delta}_0 & \hat{\delta}_1 \end{bmatrix}^\top$ Find the OLS solution.

OLS Estimation

Example (cont.d)

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}$$

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 5/6 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} \sum y_i \\ \sum y_i D_i \end{bmatrix} = \begin{bmatrix} 15 \\ 6 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 5/6 \end{bmatrix} \begin{bmatrix} 15 \\ 6 \end{bmatrix} = \begin{bmatrix} 4.5 \\ -2.5 \end{bmatrix}$$

Sample regression function:

$$\hat{y}_i = 4.5 - 2.5 D_i$$

In this simple example, the intercept is simply the arithmetic mean of the base group (y is greater than 3) ($(4+5)/2 = 4.5$). The arithmetic mean of the other group is 2. The parameter estimate on D is simply the difference between these group means (-2.5).

OLS Estimation

Example (cont.d)

Fitted values and residuals:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4.5 \\ -2.5 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 4.5 \\ 4.5 \end{bmatrix},$$

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 2 \\ 4.5 \\ 4.5 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ -0.5 \\ 0.5 \end{bmatrix}$$

OLS Estimation

Now the **same example** without the intercept:

$$y_i = \gamma_0 D_{i1} + \delta_0 D_{i2} + u_i, \quad i = 1, \ldots, n$$

Now $\boldsymbol{X}$ matrix and inner product matrix:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} \sum y_i D_{i1} \\ \sum y_i D_{i2} \end{bmatrix} = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$$

OLS estimates

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\delta}_0 \end{bmatrix} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ 4.5 \end{bmatrix}$$

Fitted regression equation:

$$\hat{y}_i = 2D_{i1} + 4.5D_{i2}$$

OLS Estimation

Example

Adding an intercept into the previous model we get:

$$y_i = \beta_0 + \gamma_0 D_{i1} + \delta_0 D_{i2} + u_i, \quad i = 1, \ldots, n$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Obviously, the columns of $\boldsymbol{X}$ are not linearly independent: $rank(\boldsymbol{X}) < 3$. Another way of seeing this:

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 5 & 3 & 2 \\ 3 & 3 & 0 \\ 2 & 0 & 2 \end{bmatrix}, \quad |\, \boldsymbol{X}^\top \boldsymbol{X} \,| = 0$$

Again this matrix is **not full rank**.

Exercise

Single explanatory variable and a constant term:

$$y_i = \beta_1 + \beta_2 x_i + u_i, \quad i = 1, \ldots, n$$

using matrix framework show that

$$\hat{\beta}_1 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum y_i x_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\hat{\beta}_2 = \frac{n \sum y_i x_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Geometry of OLS

- Let us focus on numerical properties of OLS estimates.
- These properties are always valid regardless of the data generating process and related assumptions.
- In contrast, statistical properties, such as unbiasedness, efficiency, consistency, asymptotic normality etc, all depend on the validity of certain assumptions.
- To understand the geometry of OLS let us review basic concepts in Euclidean geometry.
- n-vector is defined as a column vector with n elements which can also be represented by a $n \times 1$ matrix.
- Euclidean space in n dimensions is denoted with $\mathbb{E}^n$.
- The set of n-vectors can also be denoted by $\mathbb{R}^n$ (real line, $\mathbb{R}$).
- The difference is that wider set of operations are defined on $\mathbb{E}^n$.

Review of Euclidean Geometry

Figure 3 displays $\boldsymbol{x} = [x_1 \quad x_2]^\top \in \mathbb{E}^2$ which is also known as the Cartesian coordinates.

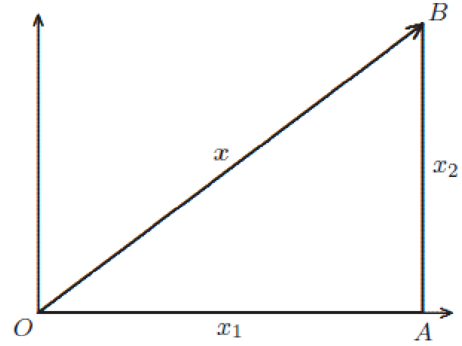

Figure 3 : A vector $\boldsymbol{x}$ in 2-dim space

What is the length of $\boldsymbol{x}$?

Review of Euclidean Geometry

- Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{E}^n$, then the inner product is defined as

$$\boldsymbol{x}^\top \boldsymbol{y} \equiv \boldsymbol{y}^\top \boldsymbol{x}$$

  which is obviously commutative.
- The length of any vector is defined as

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} \equiv \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

- Applying this to the case $n = 2$ one can see the length of the vector is just the hypotenuse formed by the coordinates of vector $\boldsymbol{x}$ (see Figure 3).

Review of Euclidean Geometry

Addition: Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{E}^2$, then

$$\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

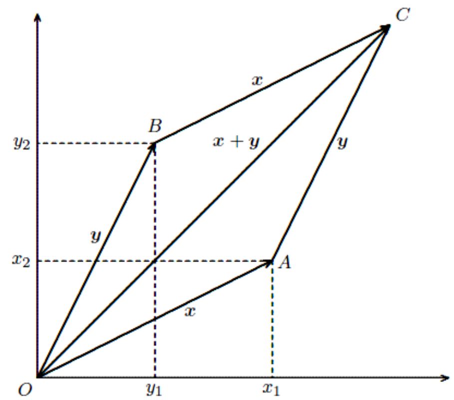

Figure 4 : Addition of two vectors

Review of Euclidean Geometry

Multiplication by a scalar is shown in Figure 5

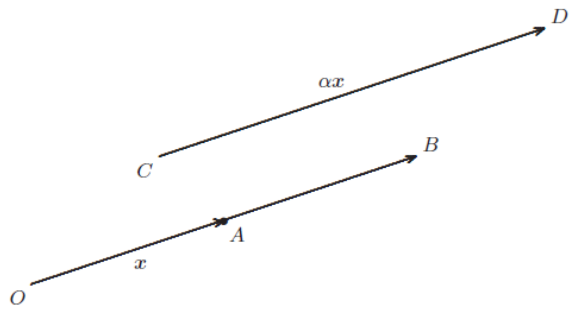

Figure 5 : Multiplication by a scalar

Let α be any scalar, then the length of $\boldsymbol{x}$ is

$$\|\alpha \boldsymbol{x}\| = |\alpha|\sqrt{\boldsymbol{x}^\top \boldsymbol{x}} = |\alpha|\|\boldsymbol{x}\|$$

Review of Euclidean Geometry

- Let us consider two vectors $\boldsymbol{w}$ and $\boldsymbol{z}$ both of length 1:

$$\boldsymbol{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \|\boldsymbol{z}\| = 1$$

  which is represented by a horizontal line in Figure 6.
- We know the length of $\boldsymbol{z}$ is 1 but what are the coordinates?
- Obviously from elementary trigonometry we have

$$z_1 = \cos\theta, \quad z_2 = \sin\theta$$

- From Pythagoras' Theorem

$$z_1^2 + z_2^2 \equiv \|\boldsymbol{z}\|^2 = \cos^2\theta + \sin^2\theta = 1$$

Review of Euclidean Geometry

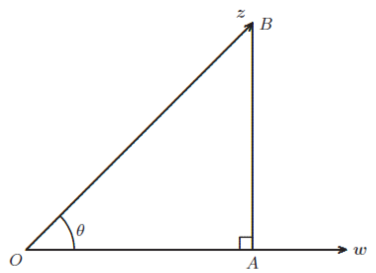

Figure 6 : The Angle between two vectors

Note in Figure 6: line OB represents $\boldsymbol{z}$, the length of AB is $\sin\theta$, the length of OA is $\cos\theta$. What is the scalar product of $\boldsymbol{w}$ and $\boldsymbol{z}$?

$$\boldsymbol{w}^\top \boldsymbol{z} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \cos\theta$$

This result only holds for vectors of length 1.

Review of Euclidean Geometry

More generally, let

$$\boldsymbol{x} = \alpha \boldsymbol{w}, \quad \boldsymbol{y} = \gamma \boldsymbol{z}, \quad \alpha > 0, \ \gamma > 0$$

The lengths are

$$\|\boldsymbol{x}\| = \alpha, \quad \|\boldsymbol{y}\| = \gamma$$

Thus,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \alpha\gamma \boldsymbol{w}^\top \boldsymbol{z}$$

Because $\boldsymbol{x}$ is parallel to $\boldsymbol{w}$, and $\boldsymbol{y}$ is parallel to $\boldsymbol{z}$, the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ is the same as that between $\boldsymbol{w}$ and $\boldsymbol{z}$, ie, θ. Therefore,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos\theta$$

Cosine of the angle between two vectors measures how close two vectors are in terms of their directions.

Review of Euclidean Geometry

- Recall
$$-1 \le \cos\theta \le 1$$
and if θ is measured in radians we have
$$\cos 0 = 1, \quad \cos(\pi/2) = 0, \quad \cos\pi = -1$$
- Thus, $\cos\theta = 1$ for vectors that are parallel, 0 for vectors that are right angles (90°) to each other, and -1 for vectors that point in directly opposite directions.
- If the angle is $\pi/2$ (in radians) or 90° (in degrees) then the inner product is zero:
$$\boldsymbol{x}^\top \boldsymbol{y} = 0.$$
- This kind of vectors are said to be **orthogonal**.

Review of Euclidean Geometry

From
$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos\theta$$
and
$$-1 \le \cos\theta \le 1$$
we have

Definition (Cauchy-Schwartz inequality)

$$|\boldsymbol{x}^\top \boldsymbol{y}| \le \|\boldsymbol{x}\| \|\boldsymbol{y}\|$$

This says that the inner product can never be greater than the product of the lengths of vectors. If $\boldsymbol{x}$ and $\boldsymbol{y}$ parallel to each other the inequality becomes the equality.

Geometry of OLS

- In the regression model
$$\underbrace{\boldsymbol{y}}_{n\times 1} = \underbrace{\boldsymbol{X}}_{n\times k} \underbrace{\boldsymbol{\beta}}_{k\times 1} + \underbrace{\boldsymbol{u}}_{n\times 1}$$
the dependent variable and each column of $\boldsymbol{X}$ can be thought of as vectors in $\mathbb{E}^n$.
- Further, the elements of n-vector can be thought of as the coordinates of a point in $\mathbb{E}^n$.
- Note that there are three vectors in the regression model: $\boldsymbol{y}$, $\boldsymbol{X\beta}$ and $\boldsymbol{u}$.
- A **subspace** of $\mathbb{E}^n$ can be defined in terms of a set of **basis** vectors.
- We are particularly interested in subspaces defined by the columns of $\boldsymbol{X}$ as the basis.
- There are k basis vectors in $\boldsymbol{X}$: $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$. The subspace associated with these basis vectors is denoted $\mathcal{S}(\boldsymbol{X})$ or $\mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$.

Geometry of OLS

- The subspace $\mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$ consists of every vector that can be formed as a linear combination of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$. This can be formally defined as follows
$$\mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) \equiv \left\{ \boldsymbol{z} \in \mathbb{E}^n : \sum_{i=1}^{k} b_i \boldsymbol{x}_1, \quad b_i \in \mathbb{R} \right\}.$$
This is called **subspace spanned by** $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ or the **column space** of $\boldsymbol{X}$.
- The **orthogonal complement** of $\mathcal{S}(\boldsymbol{X})$, denoted $\mathcal{S}^\perp(\boldsymbol{X})$, is the set of all vectors $\boldsymbol{w}$ in $\mathbb{E}^n$ that are orthogonal to everything in $\mathcal{S}(\boldsymbol{X})$. This implies that for every $\boldsymbol{z}$ in $\mathcal{S}(\boldsymbol{X})$, $\boldsymbol{w}^\top \boldsymbol{z} = 0$:
$$\mathcal{S}^\perp(\boldsymbol{X}) \equiv \left\{ \boldsymbol{w} \in \mathbb{E}^n : \ \boldsymbol{w}^\top \boldsymbol{z} = 0 \right\}.$$
- If the dimension of $\mathcal{S}(\boldsymbol{X})$ is k then the dimension of $\mathcal{S}^\perp(\boldsymbol{X})$ is $n - k$. Figure 7 shows the case where $n = 2, k = 1$.

Geometry of OLS

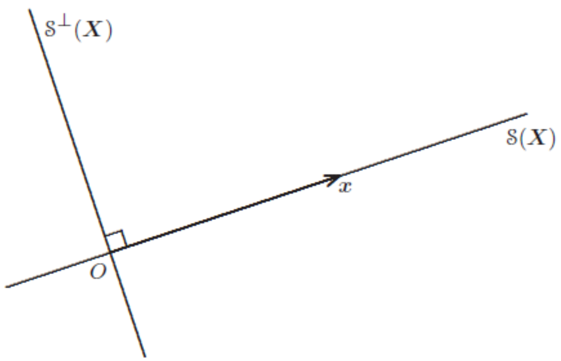

Figure 7 : Span of $\boldsymbol{X}$ and its orthogonal complement $(n = 2, k = 1)$

Geometry of OLS

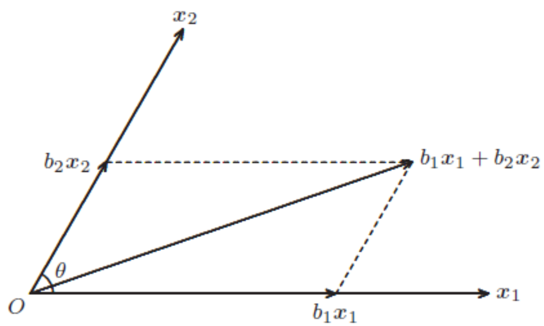

Figure 8 : 2-dimensional subspace

Figure 8 displays a 2-dimensional subspace.

Geometry of OLS

- Figure 9 represents linear regression model geometrically.
- The horizontal direction is chosen for the vector $\boldsymbol{X\beta}$, and $\boldsymbol{y}$ and $\boldsymbol{u}$ are shown in the plane. Notice that $\boldsymbol{u}$ the error vector, is not orthogonal to $\boldsymbol{X}\beta$.
- In order for the OLS estimator to be defined the square matrix $\boldsymbol{X}^\top\boldsymbol{X}$ must be invertible, ie nonsingular (full rank). This is another way of saying that the columns of $\boldsymbol{X}$ must be linearly independent. $\boldsymbol{x}_j$ is said to be linearly dependent if we can write it as a linear combination of other vectors in $\boldsymbol{X}$:

$$x_j = \sum_{i \neq j} c_i \boldsymbol{x}_i$$

Or,

$$\boldsymbol{Xb} = \boldsymbol{0}_n$$

- If this is the case then premultiplying by $\boldsymbol{X}^\top$

$$\boldsymbol{X}^\top\boldsymbol{Xb} = \boldsymbol{0}.$$

implying that $\boldsymbol{X}^\top\boldsymbol{X}$ cannot be inverted.
- If the k columns of $\boldsymbol{X}$ are not linearly independent then they will span a subspace of dimension less than k. This is called the rank of $\boldsymbol{X}$.

Geometry of OLS

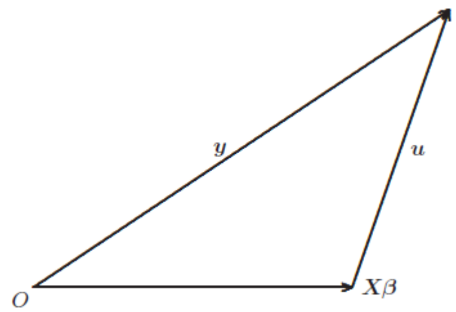

Figure 9 : Geometry of linear regression model

Geometry of OLS

- Note that every $\boldsymbol{X\beta}$ belongs to $\mathcal{S}(\boldsymbol{X})$, span of $\boldsymbol{X}$.
- The vector $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ constructed using the OLS estimator $\hat{\boldsymbol{\beta}}$ belongs to this subspace. $\hat{\boldsymbol{\beta}}$ was obtained by solving the system of equations represented by

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$$

The ith element is

$$\boldsymbol{x}_i^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \langle \boldsymbol{x}_i^\top, \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \rangle$$

- The vector $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is orthogonal to all of the regressors.
- In other words the residual vector, $\hat{\boldsymbol{u}}$, is orthogonal to all the regressors. This is depicted in Figure 10.

Geometry of OLS

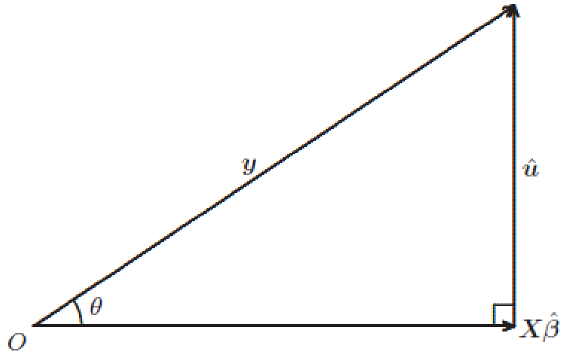

Figure 10 : Residuals and fitted values

Residual vector is orthogonal to all of the regressors:

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$$

Geometry of OLS

- Adding a third dimension we obtain Figure 11.
- There are two regressors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ which together span the horizontal plane labeled $\mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2)$. The shortest distance from $\boldsymbol{y}$ to the horizontal plane is obtained by dropping a perpendicular. Minimizing SSR accomplishes this.
- Using Pythagoras' theorem we can write

$$\|\boldsymbol{y}\|^2 = \|\boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 + \|\hat{\boldsymbol{u}}\|^2$$

or

$$\boldsymbol{y}^\top\boldsymbol{y} = \hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

i.e.

$$TSS = ESS + SSR$$

TSS = Total Sum of Squares; ESS = Explained Sum of Square; SSR = Residual Sum of Squares

Geometry of OLS

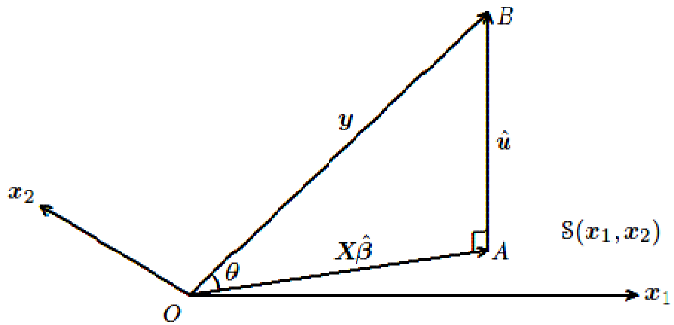

Figure 11 : $\boldsymbol{y}$ projected on two regressors

45

Geometry of OLS

- When we estimate a linear regression model we implicitly map the regressand $\boldsymbol{y}$ into a vector of fitted values $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ and a vector of residuals $\hat{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$.
- These mappings are examples of **orthogonal projection**s.
- A **projection** is a mapping that takes each point $\mathbb{E}^n$ into a point in a subspace of $\mathbb{E}^n$ while leaving all points in that subspace unchanged.
- An **orthogonal projection** maps any point into the point of the subspace that is closest to it.

46

Geometry of OLS

- An orthogonal projection onto a given subspace can be performed by premultiplying the vector to be projected by a suitable **projection matrix**.
- In OLS

$$\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$$
$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

- Obviously,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$
$$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y}$$

  $\boldsymbol{P}$ projects onto $\mathcal{S}(\boldsymbol{X})$ and when applied to $\boldsymbol{y}$, yields the fitted values.
- Similarly, when $\boldsymbol{y}$ is premultiplied by $\boldsymbol{M}$ it yields residuals:

$$\boldsymbol{M}\boldsymbol{y} = \left(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{P}\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{u}}$$

47

Geometry of OLS

- Note that

$$\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}, \quad \boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}, \quad \boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}, \quad \boldsymbol{M}\boldsymbol{M} = \boldsymbol{M},$$
$$\boldsymbol{P} + \boldsymbol{M} = \boldsymbol{I}, \quad \boldsymbol{P}\boldsymbol{M} = \boldsymbol{0}$$

- Both $\boldsymbol{P}$ and $\boldsymbol{M}$ are symmetric and idempotent matrices. Orthogonal decomposition can be written as

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y} + \boldsymbol{M}\boldsymbol{y}$$

  which can be represented by a right-angled triangle where $\boldsymbol{y}$ is the hypotenuse and $\boldsymbol{P}\boldsymbol{y}$ and $\boldsymbol{M}\boldsymbol{y}$ are the other two sides.
- Using Pythagoras' theorem

$$\boldsymbol{y}^\top\boldsymbol{y} = \hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

  i.e.

$$TSS = ESS + SSR$$

48

Coefficient of Determination, R^2

- The decomposition $TSS = ESS + SSR$ can be written as

$$\|\boldsymbol{y}\|^2 = \|\boldsymbol{P}\boldsymbol{y}\|^2 + \|\boldsymbol{M}\boldsymbol{y}\|^2.$$

- Using this we can define uncentered R^2

$$R_u^2 = \frac{ESS}{TSS} = \frac{\|\boldsymbol{P}\boldsymbol{y}\|^2}{\|\boldsymbol{y}\|^2} = 1 - \frac{\|\boldsymbol{M}\boldsymbol{y}\|^2}{\|\boldsymbol{y}\|^2} = 1 - \frac{SSR}{TSS} = \cos^2\theta$$

  where θ is the angle between $\boldsymbol{y}$ and $\boldsymbol{P}\boldsymbol{y}$.
- However, R_u^2 is not invariant to the changes in units of measurement. Instead centered R^2 is more widely used:

$$R^2 = \frac{\|\boldsymbol{P}\boldsymbol{M}_\imath\boldsymbol{y}\|^2}{\|\boldsymbol{M}_\imath\boldsymbol{y}\|^2}$$

  where

$$\boldsymbol{M}_\imath = \boldsymbol{I} - \boldsymbol{P}_\imath = \boldsymbol{I} - \boldsymbol{\imath}(\boldsymbol{\imath}^\top\boldsymbol{\imath})^{-1}\boldsymbol{\imath}^\top$$

  where $\boldsymbol{\imath}$ is n-vector of ones. Premultiplying $\boldsymbol{y}$ by $\boldsymbol{M}_\imath$ we obtain deviations from arithmetic mean. Since $-1 \leq \cos\theta \leq 1$ we have $0 \leq R^2 \leq 1$. Regression equation must contain a constant term.

Decomposition of y

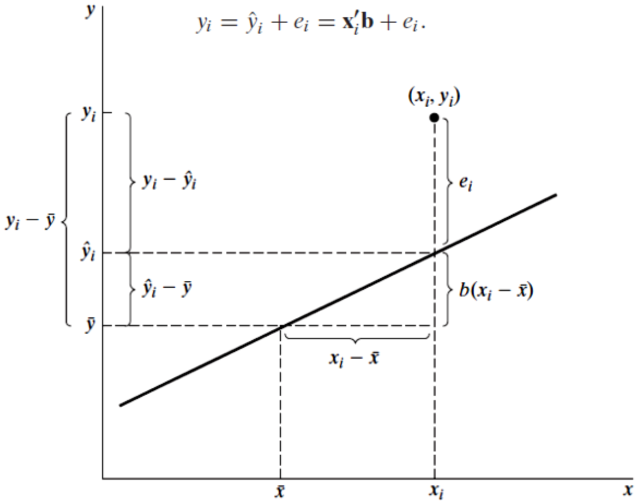

Figure 12 : Decomposition of y_i. Note that e_i is the ith residual (see Greene, p.40)

Goodness-of-Fit: R-Squared

- The coefficient of determination, R^2, is simply an estimate of "how much variation in y is explained by $x_1, x_2, \ldots, x_k$ in the population".
- A low R^2 value does not automatically imply that the classical assumptions fail.
- As the number of explanatory variables (k) increases, R^2 always increases (it never decreases). Thus, R^2 has a limited role in choosing between alternative models.
- The relative change in the R-squared when variables added to an equation may be very helpful (e.g. F-statistic for exclusion restrictions depends on the difference in R^2s).

Adjusted R-Squared: $\bar{R}^2$

- Recall the definition of R^2:

$$R^2 = 1 - \frac{SSR}{TSS}$$

- Dividing the numerator and denominator by n:

$$R^2 = 1 - \frac{SSR/n}{TSS/n} = 1 - \frac{\sigma_u^2}{\sigma_y^2}$$

- Since TSS/n and SSR/n are biased estimators of respective population variances we will instead use:

$$\frac{TSS}{n-1}, \quad \frac{SSR}{n-k}$$

Adjusted R-Squared: $\bar{R}^2$

- Adjusted R-squared is defined as

$$\bar{R}^2 = 1 - \frac{SSR/(n-k)}{TSS/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k} = 1 - \frac{(n-1)\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{(n-k)\boldsymbol{y}^\top \boldsymbol{M}_\imath \boldsymbol{y}}$$

- Adjusted R^2, or R-bar squared may increase or decrease when a new variable is added to the regression. Recall that, in contrast, R^2 never decreases.
- The reason is that when a new variable is added, while SSR decreases, the degrees of freedom $(n-k)$ also decreases.
- Basically, it imposes a penalty for adding additional variables to a model. $SSR/(n-k)$ can go up or down.
- When a new x variable is added, R-bar square increases if, and only if, the t statistic on the new variable is greater than one in absolute value.
- Extension: when a group of x variables is added, $\bar{R}^2$ increases if, and only if, the F statistic for joint significance of the new variables is greater than 1.

Frisch-Waugh-Lovell (FWL) Theorem

- Rewrite the model by grouping regressors:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \quad \boldsymbol{X} = [\underbrace{\boldsymbol{X}_1}_{n\times k_1} \vdots \underbrace{\boldsymbol{X}_2}_{n\times k_2}]$$

- Define

$$\boldsymbol{M}_1 = \boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top$$

- $\boldsymbol{M}_1\boldsymbol{y}$ gives the residuals from the regression of $\boldsymbol{y}$ on $\boldsymbol{X}_1$
- $\boldsymbol{M}_1\boldsymbol{X}_2$ gives the residuals from the regression of $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$
- FWL theorem states that $\hat{\boldsymbol{\beta}}_2$ can be obtained by regressing $\boldsymbol{M}_1\boldsymbol{y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}_2$$

- The effect of $\boldsymbol{X}_1$ is partialed out

Linear Transformations of Regressors

- What happens to OLS estimates, fitted values and residuals when take a linear transformation of explanatory variables?
- Let $\boldsymbol{A}$ be any nonsingular $k \times k$ matrix of constants and

$$\tilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{A}$$

$$\boldsymbol{y} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{u}$$

- It can be shown that

$$\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^\top = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{P}$$

- Thus the fitted values and residuals will remain the same, whereas OLS estimates will change.
- Residuals and fitted values are invariant under nonsingular transformations of the explanatory variables (eg consider changing units of measurement).

Influential Observations

- OLS parameter estimators are just a weighted average of the elements of the vector $\boldsymbol{y}$.
- Defining the ith row $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ as $\boldsymbol{c}_i$ we see that

$$\hat{\beta}_i = \boldsymbol{c}_i\boldsymbol{y}$$

- As may be obvious from this relationship some of the observations may have much more influence than the other observations. This may be seen in Figure 11.

Influential Observations

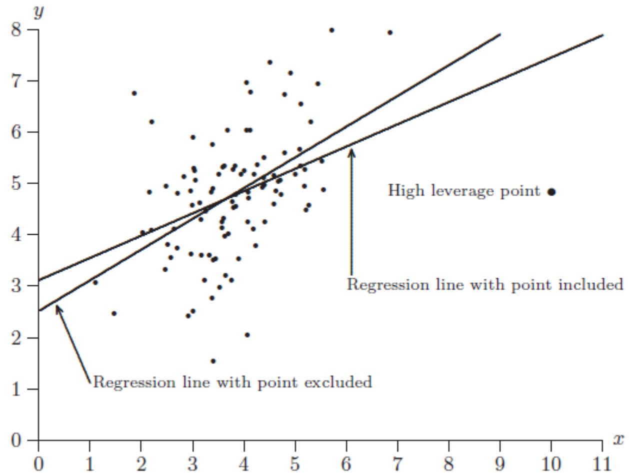

Figure 13 : An influential observation

Influential Observations

- Influential observations can be diagnosed by inspection the diagonal elements of the projection matrix $\boldsymbol{P}$, which is sometimes called the **hat matrix**. These observations are called high leverage or leverage points.
- The ith diagonal element is usually denoted h_i. Sum of these elements is just the trace of $\boldsymbol{P}$:

$$\sum_{i=1}^{n} h_i = tr(\boldsymbol{P}) = tr(\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top) = tr((\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X})$$

$$= tr(\boldsymbol{I}_k) = k$$

  where we used the property of trace:

$$tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB}) = tr(\boldsymbol{BCA}).$$

- Average of h_i is just k/n. If the elements of h_i are close to their average value then no observations has very much leverage (balanced design). If some of h_i are much larger than k/n then they may be influential observations (unbalanced design).