Two-Variable Regression Model: The Problem of Estimation

$\hat{Y}_i = E(Y_i | X_i)$

The Two-Variable PRF:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

The Two-Variable SRF:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

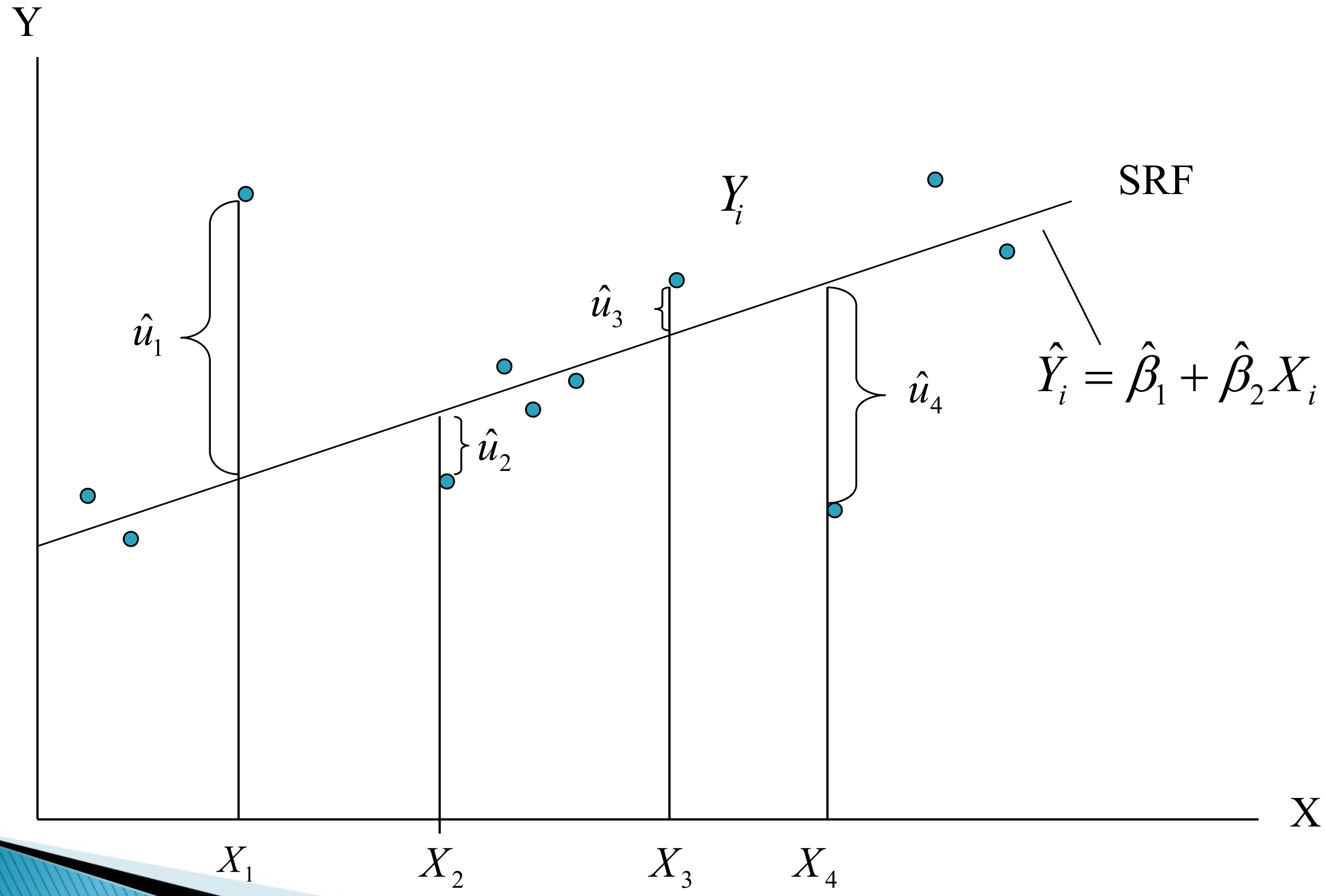
$$= \hat{Y}_i + \hat{u}_i$$

$\hat{Y}_i$ is the estimated (conditional mean) value of Y_i

$$\hat{u}_i = Y_i - \hat{Y}_i$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

Ordinary Least Squares (OLS)



$\sum \hat{u}_i = \sum (Y_i - \hat{Y}_i)$ ← This is not a very good criterion

If we adopt the criterion of minimizing $\sum \hat{u}_i$

Figure shows that the residuals $\hat{u}_2$ and $\hat{u}_3$ as well as the residuals $\hat{u}_1$ and $\hat{u}_4$ receive the same weight in the sum $(\hat{u}_1 + \hat{u}_2 + \hat{u}_3 + \hat{u}_4)$, although the first two residuals are much closer to the SRF than the latter two.

All the residuals receive equal importance no matter how close or how widely scattered the individual observations are from the SRF.

A consequence of this is that it is quite possible that the algebraic sum of the $\hat{u}_i$ is small (even zero) although the $\hat{u}_i$ are widely scattered about the SRF.

To see this, let $\hat{u}_1, \hat{u}_2, \hat{u}_3,$ and $\hat{u}_4$ in figure assume that values of 10, 2, −2, and −10 respectively.

The algebraic sum of these residuals is zero although $\hat{u}_1$ and $\hat{u}_4$ are scatter more widely around SRF than $\hat{u}_2$ and $\hat{u}_3$

we can avoid this problem if we adopt the least-squares criterion, which states that the SRF can be fixed in such a way that

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$$
$$= \sum (Y_i - \underline{\hat{\beta}_1 - \hat{\beta}_2 X_i})^2$$

is as small as possible, where $\hat{u}_i^2$ are the squared residuals.

By squaring $\hat{u}_i$, this method gives more weight to residuals such as $\hat{u}_1$ and $\hat{u}_4$ in figure than the residuals $\hat{u}_2$ and $\hat{u}_3$.

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$E(\hat{\beta}_2) = \beta_2$

$E(\hat{\beta}_1) = \beta_1$

Ordinary Least Squares (OLS)

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

The principle or the method of least squares chooses $\hat{\beta}_1 and \hat{\beta}_2$ in such a manner that, for a given sample or set of data, $\sum \hat{u}_i^2$ is as small as possible

$$\sum \hat{u}_i{}^2 = \mathrm{f}(\hat{\beta}_1, \hat{\beta}_2)$$

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial\hat{\beta}_1} = -2\sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) = -2\sum \hat{u}_i$$

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial\hat{\beta}_2} = -2\sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right) X_i = -2\sum \hat{u}_i X_i$$

Setting these equation to zero

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \text{PRF}$$

parameters

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad \text{SRF}$$

Estimators

$$= \hat{Y}_i + \hat{u}_i$$

$$\hat{u}_i = Y_i - \hat{Y}_i$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

$$= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

$$\frac{\partial(\sum \hat{u}_i)}{\partial \hat{\beta}_1} = -2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0 \quad (1)$$

$$\sum Y_i - \sum \hat{\beta}_1 - \sum \hat{\beta}_2 X_i = 0 \quad (2)$$

$$\sum Y_i - n\hat{\beta}_1 - \hat{\beta}_2 \sum X_i = 0 \quad (3)$$

$$\sum Y_i - \hat{\beta}_2 \sum X_i = n\hat{\beta}_1 \quad (4)$$

$$\frac{\sum Y_i}{n} - \hat{\beta}_2 \frac{\sum X_i}{n} = \hat{\beta}_1 \quad (5)$$

$$\bar{Y} - \hat{\beta}_2 \bar{X} = \hat{\beta}_1 \quad (6)$$

$$\bar{X} = \frac{1}{n} \sum X_i \Rightarrow \sum X_i = n\bar{X}$$

$$\bar{Y} = \frac{1}{n} \sum Y \Rightarrow \sum Y = n\bar{Y}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\frac{\partial\left(\sum \hat{u}_i\right)^2}{\partial \hat{\beta}_2} = -2\sum\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)(X_i) = 0$$ (1)

(2)

$$= \sum Y_i X_i - \hat{\beta}_1 \sum X_i - \hat{\beta}_2 \sum X_i^2 \quad (*)$$

(3)

$$= \sum Y_i X_i - \left(\bar{Y} - \hat{\beta}_2 \bar{X}\right)\sum X_i - \hat{\beta}_2 \sum X_i^2$$

(4)

$$= \sum Y_i X_i - \bar{Y}\sum X_i + \hat{\beta}_2 \bar{X} \sum X_i - \hat{\beta}_2 \sum X_i^2$$

(5)

$$= \sum Y_i X_i - \frac{\sum Y_i}{n}\sum X_i + \hat{\beta}_2 \frac{\sum X_i}{n}\sum X_i - \hat{\beta}_2 \sum X_i^2$$

(6)

$$= n\sum Y_i X_i - \sum Y_i \sum X_i + \hat{\beta}_2 \sum X_i^2 - \hat{\beta}_2 n \sum X_i^2$$

$$\hat{\beta}_2 n \sum X_i^2 - \hat{\beta}_2 \sum X_i^2 = n\sum Y_i X_i - \sum Y_i \sum X_i$$ (7)

$$\hat{\beta}_2 = \frac{n\sum Y_i X_i - \sum Y_i \sum X_i}{n\sum X_i^2 - \sum X_i^2}$$ ✓

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

Where n is the sample size. These simultaneous equations are known as the **Normal Equation**

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$ ✱

$$= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$ ✱

$$= \frac{\sum x_i y_i}{\sum x_i^2}$$ ✱

$x_i = X_i - \bar{X}$

$y_i = Y_i - \bar{Y}$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$= \frac{\sum x_i Y_i}{\sum X_i^2 - n\bar{X}^2}$$

$$= \frac{\sum X_i y_i}{\sum X_i^2 - n\bar{X}^2}$$

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

$$= \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\frac{\sum Y_i}{n} \qquad \frac{\sum X_i}{n}$$