Econometricks: Short guides to econometrics

Trick 04: The Least Squares Estimator

D. Rostam-Afschar

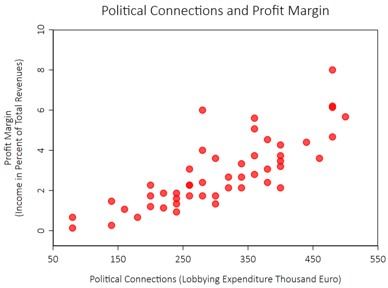
Content

Content

1. What is the Relationship between Two Variables?
2. The Econometric Model
3. Estimation with OLS
4. Properties of the OLS Estimator in the Small and in the Large
5. Politically Connected Firms: Causality or Correlation?

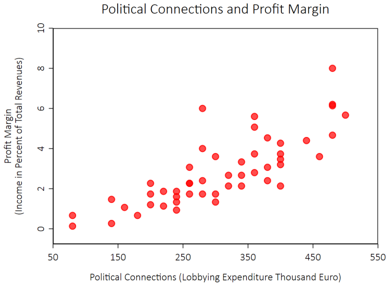
Political Connections and Firms

Firm profits increase with the degree of political connections



Political Connections and Firms

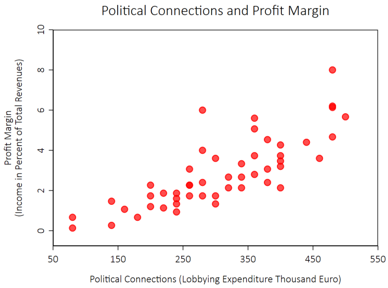
Firm profits increase with the degree of political connections



- Learn how to represent relationships between two or more variables

Political Connections and Firms

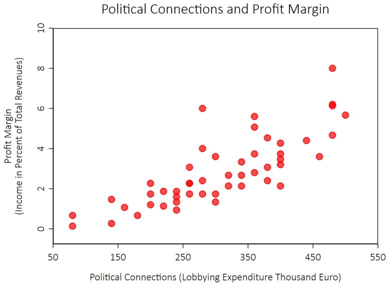
Firm profits increase with the degree of political connections



- Learn how to represent relationships between two or more variables
- How to quantify and predict effects of shocks and policy changes

Political Connections and Firms

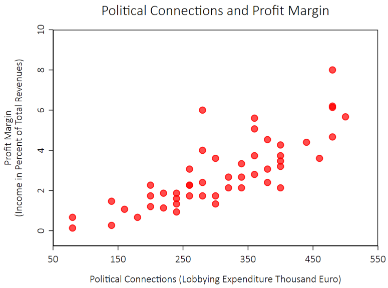
Firm profits increase with the degree of political connections



- Learn how to represent relationships between two or more variables
- How to quantify and predict effects of shocks and policy changes
- Show properties of the OLS estimator in small & large samples

Political Connections and Firms

Firm profits increase with the degree of political connections

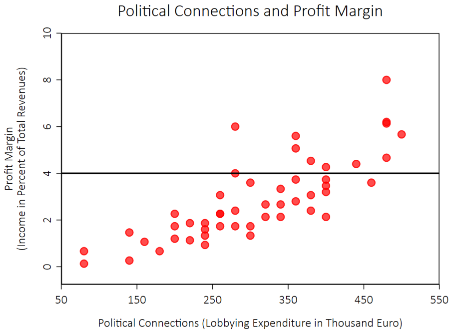


- Learn how to represent relationships between two or more variables
- How to quantify and predict effects of shocks and policy changes
- Show properties of the OLS estimator in small & large samples
- Apply Monte Carlo Simulations to assess properties of OLS

Content

Specification of a Linear Regression

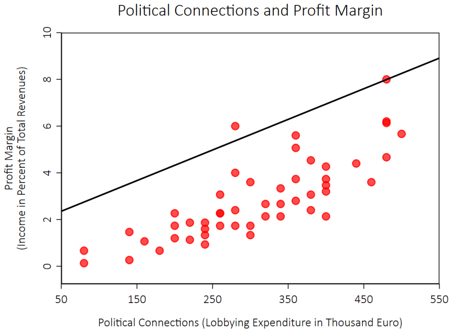
- dependent variable
  y_i = profits of firm i
- explanatory variables
  $x_{i1}, \dots, x_{iK}$ $k = 1, \dots K$
  political connections, other firm characteristics
- $x_{i0} = 1$ is a constant
- parameters to be estimated
  $\beta_0, \beta_1, \dots, \beta_K$ are $K + 1$
- u_i is called the error term



$$y_i = (\beta_0 = 4) + (\beta_1 = 0)x_{i1} + u_i.$$

Specification of a Linear Regression

- dependent variable
  y_i = profits of firm i
- explanatory variables
  $x_{i1}, \ldots, x_{iK}$ $k = 1, \ldots K$
  political connections, other firm characteristics
- $x_{i0} = 1$ is a constant
- parameters to be estimated
  $\beta_0, \beta_1, \ldots, \beta_K$ are $K + 1$
- u_i is called the error term



$$y_i = (\beta_0 = 2.36) + (\beta_1 = 0.01)x_{i1} + u_i.$$

How Were the Data Generated?

The *data generating process* is fully described by a set of assumptions.

The Five Assumptions of the Econometric Model

How Were the Data Generated?

The *data generating process* is fully described by a set of assumptions.

The Five Assumptions of the Econometric Model

- LRM1: Linearity

How Were the Data Generated?

The *data generating process* is fully described by a set of assumptions.

The Five Assumptions of the Econometric Model

- LRM1: Linearity
- LRM2: Simple random sampling

How Were the Data Generated?

The *data generating process* is fully described by a set of assumptions.

The Five Assumptions of the Econometric Model

- LRM1: Linearity
- LRM2: Simple random sampling
- LRM3: Exogeneity

How Were the Data Generated?

The *data generating process* is fully described by a set of assumptions.

The Five Assumptions of the Econometric Model

- LRM1: Linearity
- LRM2: Simple random sampling
- LRM3: Exogeneity
- LRM4: Error variance

The *data generating process* is fully described by a set of assumptions.

The Five Assumptions of the Econometric Model

- LRM1: Linearity
- LRM2: Simple random sampling
- LRM3: Exogeneity
- LRM4: Error variance
- LRM5: Identifiability

Data Generating Process: Linearity

Definition

LRM1: Linearity.

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + u_i \textit{ and } E(u_i) = 0.$$

Definition

LRM1: Linearity.

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + u_i \text{ and } E(u_i) = 0.$$

LRM1 assumes that the

- functional relationship is linear in parameters β_k

Data Generating Process: Linearity

Definition

LRM1: Linearity.

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + u_i \text{ and } E(u_i) = 0.$$

LRM1 assumes that the

- functional relationship is linear in parameters β_k
- error term u_i enters additively

Definition

LRM1: Linearity.

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + u_i \text{ and } E(u_i) = 0.$$

LRM1 assumes that the

- functional relationship is linear in parameters β_k
- error term u_i enters additively
- parameters β_k are constant across individual firms i and $j \neq i$.

Anscombe's Quartet

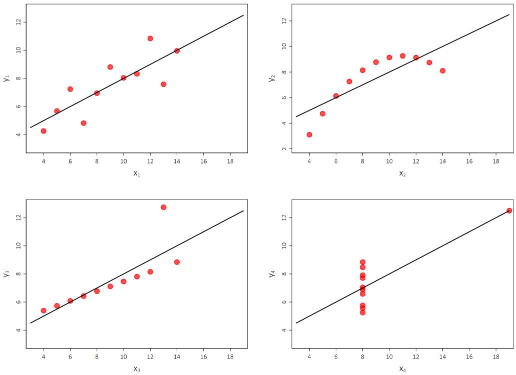


Figure 1: All four sets are identical when examined using linear statistics, but very different when graphed. Correlation between x and y is 0.816. Linear Regression y = 3.00 + 0.50x.

Data Generating Process: Random Sampling

Definition

LRM2: Simple Random Sampling.

$$\{x_{i1}, \ldots, x_{iK}, y_i\}_{i=1}^{N} \quad \textit{i.i.d. (independent and identically distributed)}$$

Data Generating Process: Random Sampling

Definition

LRM2: Simple Random Sampling.

$$\{x_{i1}, \ldots, x_{iK}, y_i\}_{i=1}^{N} \quad \textit{i.i.d. (independent and identically distributed)}$$

LRM2 means that

- observation i has no information content for observation $j \neq i$

Data Generating Process: Random Sampling

Definition

LRM2: Simple Random Sampling.

$$\{x_{i1}, \ldots, x_{iK}, y_i\}_{i=1}^{N} \quad \textit{i.i.d. (independent and identically distributed)}$$

LRM2 means that

- observation i has no information content for observation $j \neq i$
- all observations i come from the same distribution

Data Generating Process: Random Sampling

Definition

LRM2: Simple Random Sampling.

$$\{x_{i1}, \ldots, x_{iK}, y_i\}_{i=1}^{N} \quad \textit{i.i.d. (independent and identically distributed)}$$

LRM2 means that

- observation i has no information content for observation $j \neq i$
- all observations i come from the same distribution

This assumption is guaranteed by simple random sampling provided there is no systematic non-response or truncation.

Density of Population and Truncated Sample

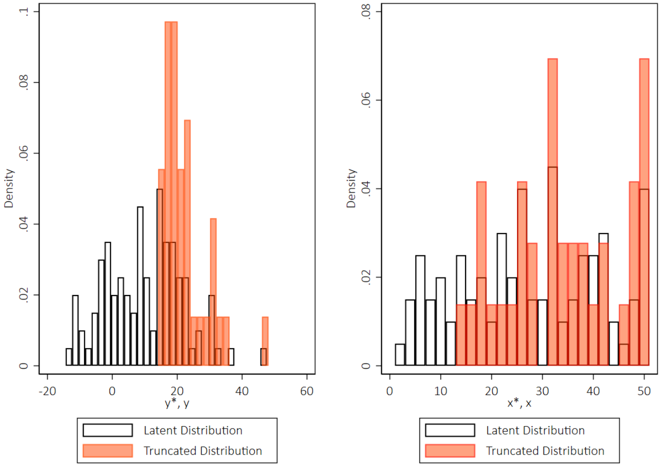


Figure 2: Distribution of a dependent variable and an independent variable truncated at $y^* = 15$.

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i \big| x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i \middle| x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)

$$u_i \perp x_{ik} \quad \forall k \quad (independent), pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)

$$u_i \perp x_{ik} \quad \forall k \quad (independent), pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

LRM3b means that the error term is independent of the explanatory variables.

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)

$$u_i \perp x_{ik} \quad \forall k \quad \textit{(independent)}, pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

LRM3b means that the error term is independent of the explanatory variables.

c)

$$E(u_i | x_{i1}, \ldots, x_{iK}) = E(u_i) = 0 \quad \textit{(mean independent)}$$

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)

$$u_i \perp x_{ik} \quad \forall k \quad (independent), pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

LRM3b means that the error term is independent of the explanatory variables.

c)

$$E(u_i | x_{i1}, \ldots, x_{iK}) = E(u_i) = 0 \quad (mean\ independent)$$

LRM3c states that the mean of the error term is independent of explanatory variables.

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)
$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)
$$u_i \perp x_{ik} \quad \forall k \quad (independent), pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

LRM3b means that the error term is independent of the explanatory variables.

c)
$$E(u_i | x_{i1}, \ldots, x_{iK}) = E(u_i) = 0 \quad (mean\ independent)$$

LRM3c states that the mean of the error term is independent of explanatory variables.

d)
$$cov(x_{ik}, u_i) = 0 \quad \forall k \quad (uncorrelated)$$

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)

$$u_i \perp x_{ik} \quad \forall k \quad (independent), pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

LRM3b means that the error term is independent of the explanatory variables.

c)

$$E(u_i | x_{i1}, \ldots, x_{iK}) = E(u_i) = 0 \quad (mean\ independent)$$

LRM3c states that the mean of the error term is independent of explanatory variables.

d)

$$cov(x_{ik}, u_i) = 0 \quad \forall k \quad (uncorrelated)$$

LRM3d means that the error term and the explanatory variables are uncorrelated.

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)

$$u_i \perp x_{ik} \quad \forall k \quad (independent), pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

LRM3b means that the error term is independent of the explanatory variables.

c)

$$E(u_i | x_{i1}, \ldots, x_{iK}) = E(u_i) = 0 \quad (mean\ independent)$$

LRM3c states that the mean of the error term is independent of explanatory variables.

d)

$$cov(x_{ik}, u_i) = 0 \quad \forall k \quad (uncorrelated)$$

LRM3d means that the error term and the explanatory variables are uncorrelated.

Data Generating Process: Exogeneity

Definition

LRM3: Exogeneity.

a)

$$u_i | x_{i1}, \ldots, x_{iK} \sim N(0, \sigma_i^2)$$

LRM3a assumes that the error term is normally distributed conditional on the explanatory variables.

b)

$$u_i \perp x_{ik} \quad \forall k \quad (independent), pdf_{u,x}(u_i x_{ik}) = pdf_u(u_i) pdf_x(x_{ik})$$

LRM3b means that the error term is independent of the explanatory variables.

c)

$$E(u_i | x_{i1}, \ldots, x_{iK}) = E(u_i) = 0 \quad (mean\ independent)$$

LRM3c states that the mean of the error term is independent of explanatory variables.

d)

$$cov(x_{ik}, u_i) = 0 \quad \forall k \quad (uncorrelated)$$

LRM3d means that the error term and the explanatory variables are uncorrelated.

LRM3a or LRM3b imply LRM3c and LRM3d. LRM3c implies LRM3d.

(Conditional) Mean Independence

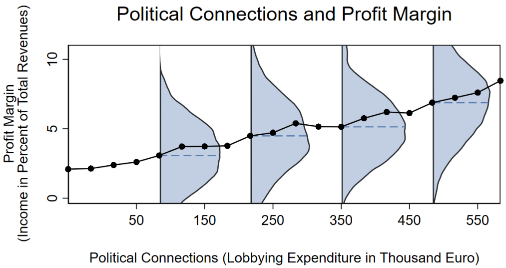


Figure 3: Distributions of the dependent variable conditional on values of an independent variable.

Weaker exogeneity assumption if interest only in, say, x_{i1}:

Conditional Mean Independence

$E(u_i | x_{i1}, x_{i2}, \ldots, x_{iK}) = E(u_i | x_{i2}, \ldots, x_{iK})$

Given the control variables $x_{i2}, \ldots, x_{iK}$, the mean of u_i does not depend on the variable of interest x_{i1}.

Data Generating Process: Error Variance

Definition

LRM4: Error Variance.

Data Generating Process: Error Variance

Definition

LRM4: Error Variance.

a)
$$V(u_i|x_{i1},\ldots,x_{iK}) = \sigma^2 < \infty \quad \textit{(homoskedasticity)}$$

Data Generating Process: Error Variance

Definition

LRM4: Error Variance.

a)
$$V(u_i|x_{i1},\dots,x_{iK}) = \sigma^2 < \infty \quad \textit{(homoskedasticity)}$$

LRM4a means that the variance of the error term is a constant.

Data Generating Process: Error Variance

Definition

LRM4: Error Variance.

a)

$$V(u_i | x_{i1}, \ldots, x_{iK}) = \sigma^2 < \infty \quad \textit{(homoskedasticity)}$$

LRM4a means that the variance of the error term is a constant.

b)

$$V(u_i | x_{i1}, \ldots, x_{iK}) = \sigma_i^2 = g(x_{i1}, \ldots, x_{iK}) < \infty \quad \textit{(cond. heteroskedasticity)}$$

Data Generating Process: Error Variance

Definition

LRM4: Error Variance.

a)

$$V(u_i|x_{i1},\ldots,x_{iK}) = \sigma^2 < \infty \quad \textit{(homoskedasticity)}$$

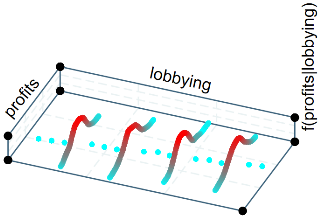
LRM4a means that the variance of the error term is a constant.

b)

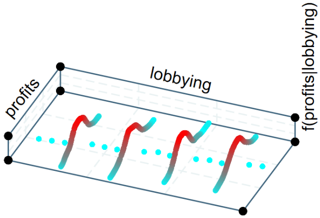
$$V(u_i|x_{i1},\ldots,x_{iK}) = \sigma_i^2 = g(x_{i1},\ldots,x_{iK}) < \infty \quad \textit{(cond. heteroskedasticity)}$$

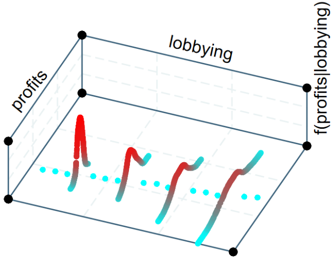
LRM4b allows the variance of the error term to depend on a function g of the explanatory variables.

Heteroskedasticity



Heteroskedasticity





Heteroskedasticity

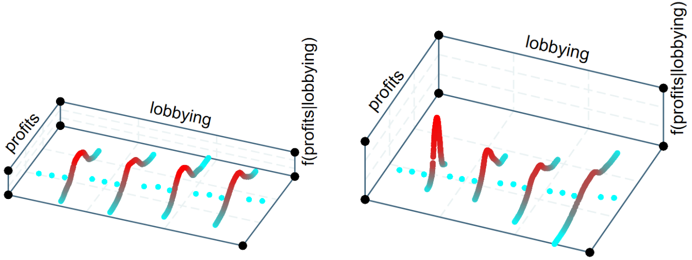


Figure 4: The simple regression model under homo- and heteroskedasticity. $Var(profits|lobbying, employees)$ increasing with *lobbying*.

Data Generating Process: Identifiability

Definition

LRM5: Identifiability.

Data Generating Process: Identifiability

Definition

LRM5: Identifiability.

$$(x_{i0}, x_{i1}, \ldots, x_{iK}) \textit{ are not linearly dependent}$$

Definition

LRM5: Identifiability.

$$(x_{i0}, x_{i1}, \ldots, x_{iK}) \text{ are not linearly dependent}$$

$$0 < V(x_{ik}) < \infty \quad \forall k > 0$$

Data Generating Process: Identifiability

Definition

LRM5: Identifiability.

$$(x_{i0}, x_{i1}, \ldots, x_{iK}) \textit{ are not linearly dependent}$$

$$0 < V(x_{ik}) < \infty \quad \forall k > 0$$

LRM5 assumes that

Definition

LRM5: Identifiability.

$$(x_{i0}, x_{i1}, \ldots, x_{iK}) \textit{ are not linearly dependent}$$

$$0 < V(x_{ik}) < \infty \quad \forall k > 0$$

LRM5 assumes that

- the regressors are not *perfectly collinear*, i.e. no variable is a linear combination of the others

Data Generating Process: Identifiability

Definition

LRM5: Identifiability.

$$(x_{i0}, x_{i1}, \ldots, x_{iK}) \text{ are not linearly dependent}$$

$$0 < V(x_{ik}) < \infty \quad \forall k > 0$$

LRM5 assumes that

- the regressors are not *perfectly collinear*, i.e. no variable is a linear combination of the others
- all regressors (but the constant) have strictly positive variance both in expectations and in the sample and not too many extreme values.

Data Generating Process: Identifiability

Definition

LRM5: Identifiability.

$$(x_{i0}, x_{i1}, \ldots, x_{iK}) \textit{ are not linearly dependent}$$

$$0 < V(x_{ik}) < \infty \quad \forall k > 0$$

LRM5 assumes that

- the regressors are not *perfectly collinear*, i.e. no variable is a linear combination of the others
- all regressors (but the constant) have strictly positive variance both in expectations and in the sample and not too many extreme values.

LRM5 means that every explanatory variable adds additional information.

The Identifying Variation from x_{ik}

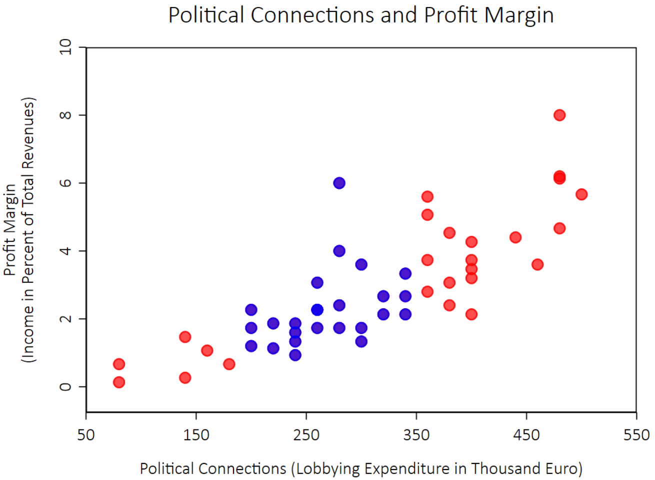


Figure 5: The number of red and blue dots is the same. Using which would you get a more accurate regression line?

Content

Estimation with OLS

Ordinary least squares (OLS) minimizes the squared distances (SD) between the observed and the predicted dependent variable y:

Ordinary least squares (OLS) minimizes the squared distances (SD) between the observed and the predicted dependent variable y:

$$\min_{\beta_0,\ldots,\beta_K} SD(\beta_0, \ldots, \beta_K),$$

Ordinary least squares (OLS) minimizes the squared distances (SD) between the observed and the predicted dependent variable y:

$$\min_{\beta_0,\ldots,\beta_K} SD(\beta_0, \ldots, \beta_K),$$

$$\text{where } SD = \sum_{i=1}^{N} [y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK})]^2.$$

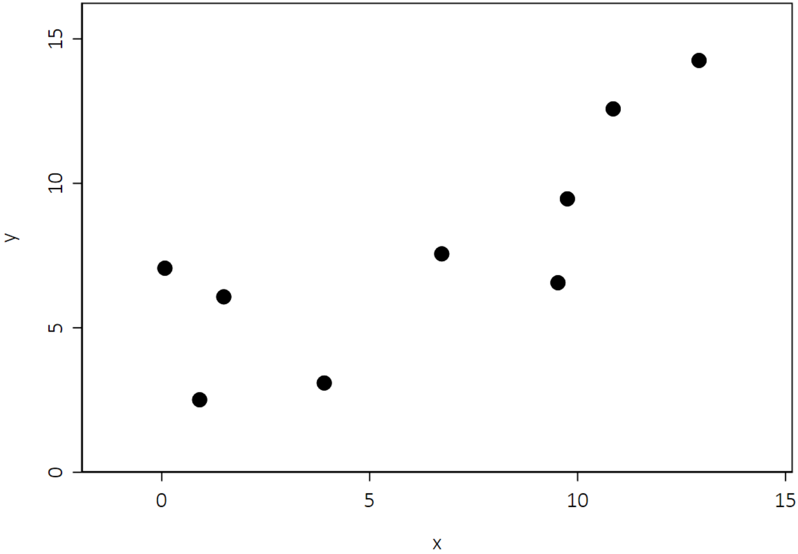
Estimation with OLS

Ordinary least squares (OLS) minimizes the squared distances (SD) between the observed and the predicted dependent variable y:

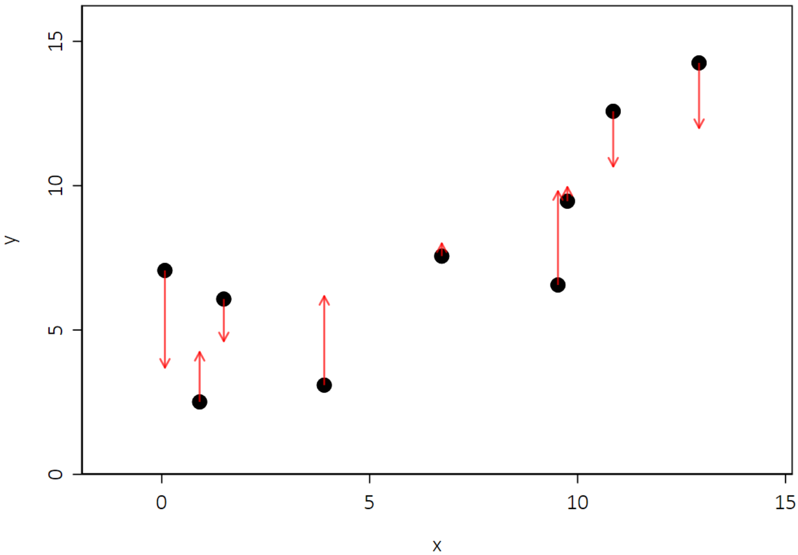
$$\min_{\beta_0,\ldots,\beta_K} SD(\beta_0, \ldots, \beta_K),$$

$$\text{where } SD = \sum_{i=1}^{N} [y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK})]^2.$$

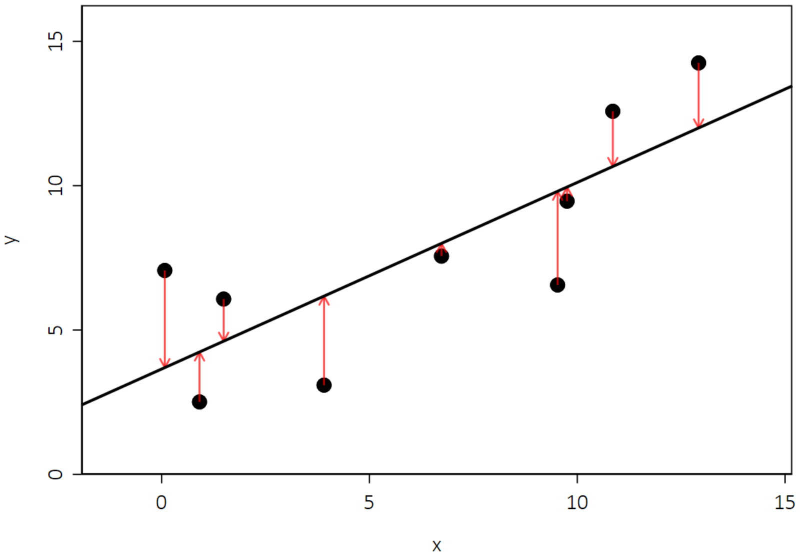
How to Describe the Relationship Best?



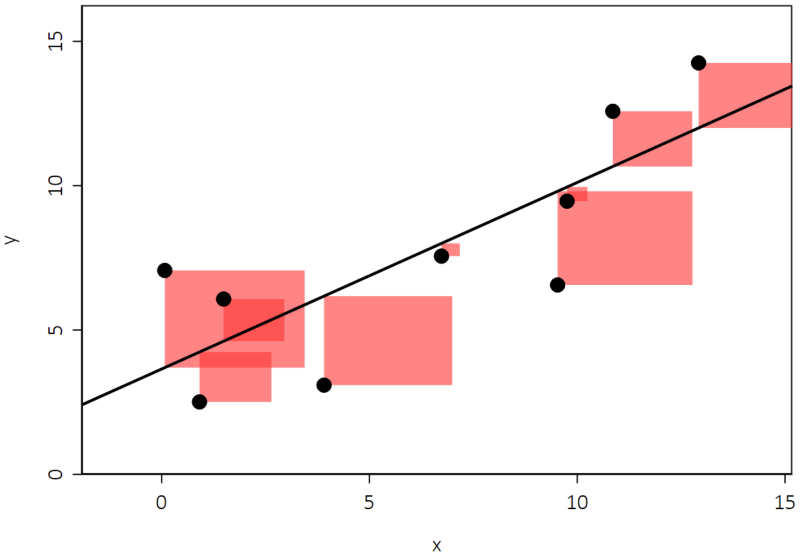
How to Describe the Relationship Best?



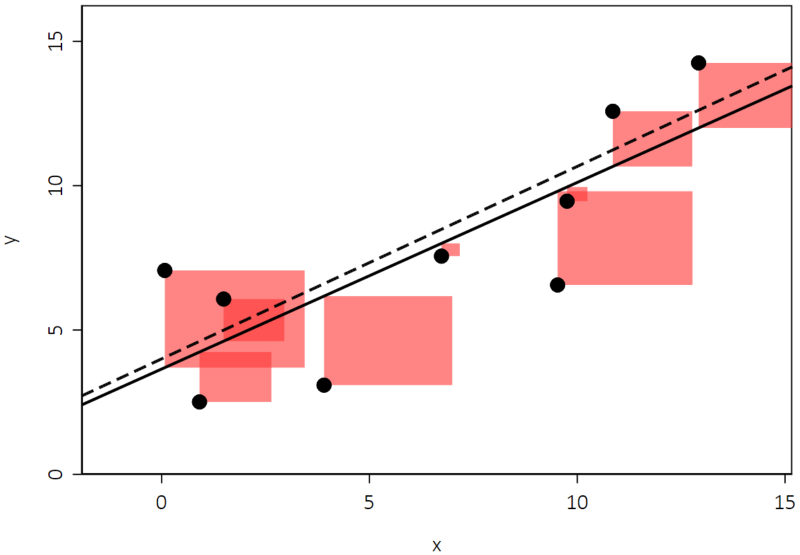
How to Describe the Relationship Best?



How to Describe the Relationship Best?



How to Describe the Relationship Best?



Invention of OLS

Legendre to Jacobi (Paris, 30 November 1827, Plackett, 1972): *"...How can Mr. Gauss have dared to tell you that the greater part of your theorems were known to him...?*

Invention of OLS

Legendre to Jacobi (Paris, 30 November 1827, Plackett, 1972): *"...How can Mr. Gauss have dared to tell you that the greater part of your theorems were known to him...?*

... this is the same man ... who wanted to appropriate in 1809 the method of least squares published in 1805.

Invention of OLS

Legendre to Jacobi (Paris, 30 November 1827, Plackett, 1972): "*...How can Mr. Gauss have dared to tell you that the greater part of your theorems were known to him...?*

... this is the same man ... who wanted to appropriate in 1809 the method of least squares published in 1805.

— Other examples will be found in other places, but a man of honour should refrain from imitating them."

Legendre to Jacobi (Paris, 30 November 1827, Plackett, 1972): *"...How can Mr. Gauss have dared to tell you that the greater part of your theorems were known to him...?*

... this is the same man ... who wanted to appropriate in 1809 the method of least squares published in 1805.

— Other examples will be found in other places, but a man of honour should refrain from imitating them."

Figure 6: Watercolor caricature of Legendre by Boilly (1820), the only existing portrait known.

Invention of OLS

Legendre to Jacobi (Paris, 30 November 1827, Plackett, 1972): *"...How can Mr. Gauss have dared to tell you that the greater part of your theorems were known to him...?*

... this is the same man ... who wanted to appropriate in 1809 the method of least squares published in 1805.

— Other examples will be found in other places, but a man of honour should refrain from imitating them."

Figure 7: Portrait of Gauss by Jensen (1840).

Estimation with OLS

For the bivariate regression model, the OLS estimators of β_0 and β_1 are

Estimation with OLS

For the bivariate regression model, the OLS estimators of β_0 and β_1 are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimation with OLS

For the bivariate regression model, the OLS estimators of β_0 and β_1 are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (x_{i1} - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_{i1} - \bar{x})^2} = \frac{cov(x, y)}{var(x)}$$

Estimation with OLS

For the bivariate regression model, the OLS estimators of β_0 and β_1 are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (x_{i1} - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_{i1} - \bar{x})^2} = \frac{cov(x, y)}{var(x)}$$

$$\hat{\beta}_1 = cov(x, y)/(s_x s_x) = R s_y / s_x,$$

where $R \equiv cov(x, y)/(s_x s_y)$ is **Pearson's correlation coefficient** with s_z denoting the standard deviation of z.

OLS estimator Measures Linear Correlation

Equivalently,

$$R = s_x/s_y\hat{\beta}_1 = \frac{\hat{\beta}_1\sum_{i=1}^{N}(x_{i1} - \bar{x})}{\sum_{i=1}^{N}(y_i - \bar{y})} = \frac{\sum_{i=1}^{N}(\hat{\beta}_1 x_{i1} - \hat{\beta}_1\bar{x})}{\sum_{i=1}^{N}(y_i - \bar{y})}.$$

Squaring gives

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{N}\hat{u}_i^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}.$$

R^2 as measure of the **goodness of fit**:
The fit improves with the fraction of the sample variation in y that is explained by the x.

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

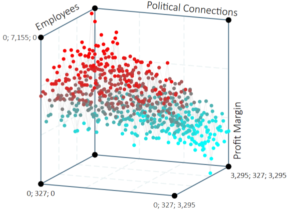


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

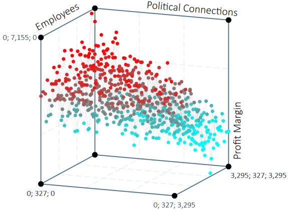


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

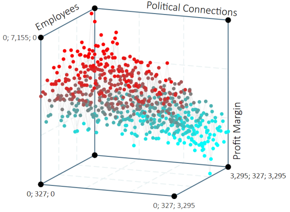


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

Given the OLS estimator, we can predict the

- dependent variable by $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_K x_{iK}$

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

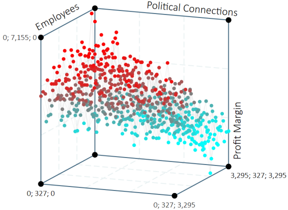


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

Given the OLS estimator, we can predict the

- dependent variable by $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_K x_{iK}$

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

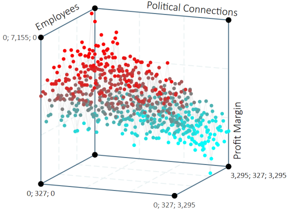


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

Given the OLS estimator, we can predict the

- dependent variable by $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_K x_{iK}$
- the error term by $\hat{u}_i = y_i - \hat{y}_i$.

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

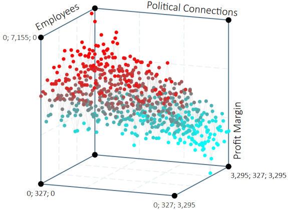


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

Given the OLS estimator, we can predict the

- dependent variable by $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_K x_{iK}$
- the error term by $\hat{u}_i = y_i - \hat{y}_i$.

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

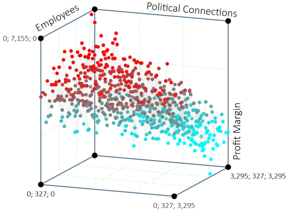


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

Given the OLS estimator, we can predict the

- dependent variable by $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_K x_{iK}$
- the error term by $\hat{u}_i = y_i - \hat{y}_i$.

$\hat{u}_i$ is called the *residual*.

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

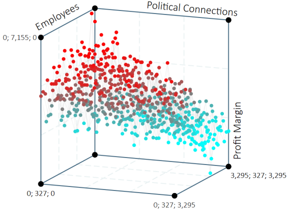


Figure 8: Scatter cloud visualized with GRAPH3D for Stata.

Given the OLS estimator, we can predict the

- dependent variable by $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_K x_{iK}$
- the error term by $\hat{u}_i = y_i - \hat{y}_i$.

$\hat{u}_i$ is called the *residual*.

Adjusted $R^2 = 1 - \frac{N-1}{N-K-1} \frac{\sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$.

The Case with K Explanatory Variables

The more general case with K explanatory variables is

$$\underset{(K+1)\times 1}{\hat{\beta}} = \underset{(K+1)\times(K+1)}{(X'X)^{-1}} \underset{(K+1)\times N}{X'} \underset{N\times 1}{y}$$

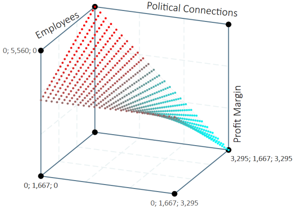


Figure 9: OLS surface visualized with GRAPH3D for Stata.

Given the OLS estimator, we can predict the

- dependent variable by $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_K x_{iK}$
- the error term by $\hat{u}_i = y_i - \hat{y}_i$.

$\hat{u}_i$ is called the *residual*.

Adjusted $R^2 = 1 - \frac{N-1}{N-K-1} \frac{\sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$.

Content

Properties of the OLS Estimator

- *Small sample properties of* $\hat{\beta}$

- *Small sample properties of* $\hat{\beta}$
    - unbiased

Properties of the OLS Estimator

- *Small sample properties of* $\hat{\beta}$
  - unbiased
  - normally distributed

Properties of the OLS Estimator

- *Small sample properties of* $\hat{\beta}$
  - unbiased
  - normally distributed
  - efficient

Properties of the OLS Estimator

- *Small sample properties of* $\hat{\beta}$
  - unbiased
  - normally distributed
  - efficient
- *Large sample properties of* $\hat{\beta}$

Properties of the OLS Estimator

- *Small sample properties of* $\hat{\beta}$
  - unbiased
  - normally distributed
  - efficient
- *Large sample properties of* $\hat{\beta}$
  - consistent

Properties of the OLS Estimator

- *Small sample properties of* $\hat{\beta}$
  - unbiased
  - normally distributed
  - efficient

- *Large sample properties of* $\hat{\beta}$
  - consistent
  - approx. normal

- *Small sample properties of* $\hat{\beta}$
  - unbiased
  - normally distributed
  - efficient
- *Large sample properties of* $\hat{\beta}$
  - consistent
  - approx. normal
  - asymptotically efficient

Small Sample Properties

Figure 10: What is a small sample?
Source: Familien-Duell
Grundy Light Entertainment.

Small Sample Properties

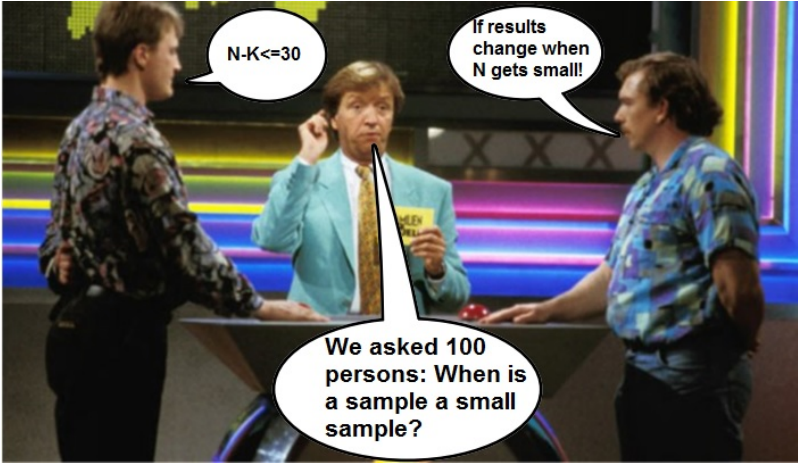


Figure 11: What is a small sample? (Wooldridge, 2009, p. 755): "But large sample approximations have been known to work well for sample sizes as small as $N = 20$." *Source:* Familien-Duell Grundy Light Entertainment.

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1,

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2,

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a,

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

- The OLS estimator of β is **unbiased**.

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

- The OLS estimator of β is **unbiased**.

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

- The OLS estimator of β is **unbiased**.

$$E(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \beta_k.$$

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

- The OLS estimator of β is **unbiased**.

$$E(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \beta_k.$$

- The OLS estimator is (multivariate) **normally distributed**.

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

- The OLS estimator of β is **unbiased**.

$$E(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \beta_k.$$

- The OLS estimator is (multivariate) **normally distributed**.

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

- The OLS estimator of β is **unbiased**.

$$E(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \beta_k.$$

- The OLS estimator is (multivariate) **normally distributed**.

$$\hat{\beta}_k | x_{11}, \ldots, x_{NK} \sim N(\beta_k, V(\hat{\beta}_k)).$$

Unbiasedness and Normality of $\hat{\beta}_k$

Assuming LRM1, LRM2, LRM3a, LRM4, and LRM5,
the following properties can be established even for small samples.

- The OLS estimator of β is **unbiased**.

$$E(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \beta_k.$$

- The OLS estimator is (multivariate) **normally distributed**.

$$\hat{\beta}_k | x_{11}, \ldots, x_{NK} \sim N(\beta_k, V(\hat{\beta}_k)).$$

- Under homoskedasticity (LRM4a)
  the variance $\widehat{V}(\hat{\beta}_k | x_{11}, \ldots, x_{NK})$ can be **unbiasedly** estimated.

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

$$\widehat{V} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \text{ with}$$

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

$$\widehat{V} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \text{ with}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N - K - 1}.$$

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

$$\widehat{V} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \text{ with}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N - K - 1}.$$

- Gauß-Markov-Theorem: under homoskedasticity (LRM4a) $\hat{\beta}_k$ is the **BLUE** (best linear unbiased estimator, e.g., non-linear least squares biased).

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

$$\widehat{V} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \text{ with}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N - K - 1}.$$

- Gauß-Markov-Theorem: under homoskedasticity (LRM4a) $\hat{\beta}_k$ is the **BLUE** (best linear unbiased estimator, e.g., non-linear least squares biased).
- $\widehat{V}(\hat{\beta}_k)$ inflates with

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

$$\widehat{V} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \text{ with}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}\hat{u}_i^2}{N - K - 1}.$$

- Gauß-Markov-Theorem: under homoskedasticity (LRM4a) $\hat{\beta}_k$ is the **BLUE** (best linear unbiased estimator, e.g., non-linear least squares biased).
- $\widehat{V}(\hat{\beta}_k)$ inflates with
  - **micronumerosity** (small sample size)

Variance of $\hat{\beta}_k$ and Efficiency

- For the bivariate regression model, it is estimated as

$$\widehat{V} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \text{ with}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}\hat{u}_i^2}{N - K - 1}.$$

- Gauß-Markov-Theorem: under homoskedasticity (LRM4a) $\hat{\beta}_k$ is the **BLUE** (best linear unbiased estimator, e.g., non-linear least squares biased).
- $\widehat{V}(\hat{\beta}_k)$ inflates with
  - **micronumerosity** (small sample size)
  - **multicollinearity** (high (but not perfect) correlation between two or more of the independent variables).

Unbiasedness

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

Unbiasedness

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

$$\mathrm{E}[\hat{\beta}] = E_{X,u}\Big[(X'X)^{-1}X'(X\beta + u)\Big]$$

Unbiasedness

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

$$\mathrm{E}[\hat{\beta}] = E_{X,u}\Big[(X'X)^{-1}X'(X\beta + u)\Big]$$

$$= \beta + E_{X,u}\Big[(X'X)^{-1}X'u\Big]$$

Unbiasedness

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

$$\mathrm{E}[\hat{\beta}] = E_{X,u}\Big[(X'X)^{-1}X'(X\beta + u)\Big]$$

$$= \beta + E_{X,u}\Big[(X'X)^{-1}X'u\Big]$$

$$= \beta + E_{X}\Big[E_{u|X}\Big[(X'X)^{-1}X'u|X\Big]\Big]$$

Unbiasedness

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

$$\mathrm{E}[\hat{\beta}] = E_{X,u}\Big[(X'X)^{-1}X'(X\beta + u)\Big]$$

$$= \beta + E_{X,u}\Big[(X'X)^{-1}X'u\Big]$$

$$= \beta + E_X\Big[E_{u|X}\Big[(X'X)^{-1}X'u|X\Big]\Big]$$

$$= \beta + E_X\Big[(X'X)^{-1}X'E_{u|X}[u|X]\Big]$$

Unbiasedness

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

$$
\begin{aligned}
\mathrm{E}[\hat{\beta}] &= E_{X,u}\Big[(X'X)^{-1}X'(X\beta + u)\Big] \\
&= \beta + E_{X,u}\Big[(X'X)^{-1}X'u\Big] \\
&= \beta + E_{X}\Big[E_{u|X}\Big[(X'X)^{-1}X'u|X\Big]\Big] \\
&= \beta + E_{X}\Big[(X'X)^{-1}X'E_{u|X}[u|X]\Big] \\
&= \beta,
\end{aligned}
$$

Unbiasedness

- The OLS estimator of β is *unbiased*.
  Plug $y = X\beta + u$ into the formula for $\hat{\beta}$ and then use the law of iterated expectation to first take expectation with respect to u conditional on X and then take the unconditional expectation:

$$
\begin{aligned}
\mathrm{E}[\hat{\beta}] &= E_{X,u}\left[(X'X)^{-1}X'(X\beta + u)\right] \\
&= \beta + E_{X,u}\left[(X'X)^{-1}X'u\right] \\
&= \beta + E_X\left[E_{u|X}\left[(X'X)^{-1}X'u|X\right]\right] \\
&= \beta + E_X\left[(X'X)^{-1}X'E_{u|X}[u|X]\right] \\
&= \beta,
\end{aligned}
$$

  where $E[u|X] = 0$ by assumptions of the model. $\square$

Variance

- The OLS estimator β has variance $\widehat{V}(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \sigma^2 (X'X)^{-1}$

Variance

- The OLS estimator β has variance $\widehat{V}(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \sigma^2 (X'X)^{-1}$

Variance

- The OLS estimator β has variance
  $\widehat{V}(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \sigma^2 (X'X)^{-1}$
  Let $\sigma^2 I$ denote the covariance matrix of u. Then,

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E\left[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\right]$$

- The OLS estimator β has variance $\widehat{V}(\hat{\beta}_k|x_{11},\ldots,x_{NK}) = \sigma^2(X'X)^{-1}$
  Let $\sigma^2 I$ denote the covariance matrix of u. Then,

$$E[\,(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = E\Big[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\Big]$$

$$= E\Big[(X'X)^{-1}X'uu'X(X'X)^{-1}\Big]$$

Variance

- The OLS estimator β has variance
  $\widehat{V}(\hat{\beta}_k|x_{11},\ldots,x_{NK}) = \sigma^2(X'X)^{-1}$
  Let $\sigma^2 I$ denote the covariance matrix of u. Then,

$$E[\,(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = E\Big[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\Big]$$

$$= E\Big[(X'X)^{-1}X'uu'X(X'X)^{-1}\Big]$$

$$= E\Big[(X'X)^{-1}X'\sigma^2X(X'X)^{-1}\Big]$$

Variance

- The OLS estimator β has variance $\widehat{V}(\hat{\beta}_k|x_{11},\ldots,x_{NK}) = \sigma^2(X'X)^{-1}$
  Let $\sigma^2 I$ denote the covariance matrix of u. Then,

$$
\begin{aligned}
E[\,(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] &= E\Big[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\Big] \\
&= E\Big[(X'X)^{-1}X'uu'X(X'X)^{-1}\Big] \\
&= E\Big[(X'X)^{-1}X'\sigma^2 X(X'X)^{-1}\Big] \\
&= E\Big[\sigma^2(X'X)^{-1}X'X(X'X)^{-1}\Big]
\end{aligned}
$$

- The OLS estimator β has variance $\widehat{V}(\hat{\beta}_k | x_{11}, \ldots, x_{NK}) = \sigma^2 (X'X)^{-1}$
  Let $\sigma^2 I$ denote the covariance matrix of u. Then,

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E\left[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\right]$$

$$= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right]$$

$$= E\left[(X'X)^{-1}X'\sigma^2 X(X'X)^{-1}\right]$$

$$= E\left[\sigma^2 (X'X)^{-1}X'X(X'X)^{-1}\right]$$

$$= \sigma^2 (X'X)^{-1},$$

where we used the fact that $\hat{\beta} - \beta$ is just an affine transformation of u by the matrix $(X'X)^{-1}X'$. $\square$

Estimator for Variance

For a simple linear regression model, where $\beta = [\beta_0, \beta_1]'$ (β_0 is the y-intercept and β_1 is the slope), one obtains

$$\sigma^2(X'X)^{-1} = \sigma^2 \left(\sum x_i x_i' \right)^{-1}$$

Estimator for Variance

For a simple linear regression model, where $\beta = [\beta_0, \beta_1]'$ (β_0 is the y-intercept and β_1 is the slope), one obtains

$$\sigma^2(X'X)^{-1} = \sigma^2 \left(\sum x_i x_i' \right)^{-1}$$

$$= \sigma^2 \left(\sum (1, x_i)'(1, x_i) \right)^{-1}$$

Estimator for Variance

For a simple linear regression model, where $\beta = [\beta_0, \beta_1]'$ (β_0 is the y-intercept and β_1 is the slope), one obtains

$$\sigma^2(X'X)^{-1} = \sigma^2 \left(\sum x_i x_i' \right)^{-1}$$

$$= \sigma^2 \left(\sum (1, x_i)'(1, x_i) \right)^{-1}$$

$$= \sigma^2 \left(\sum \begin{pmatrix} 1 x_i \\ x_i x_i^2 \end{pmatrix} \right)^{-1}$$

Estimator for Variance

For a simple linear regression model, where $\beta = [\beta_0, \beta_1]'$ (β_0 is the y-intercept and β_1 is the slope), one obtains

$$\sigma^2(X'X)^{-1} = \sigma^2 \left(\sum x_i x_i' \right)^{-1}$$

$$= \sigma^2 \left(\sum (1, x_i)'(1, x_i) \right)^{-1}$$

$$= \sigma^2 \left(\sum \begin{pmatrix} 1 x_i \\ x_i x_i^2 \end{pmatrix} \right)^{-1}$$

$$= \sigma^2 \begin{pmatrix} N \sum x_i \\ \sum x_i \sum x_i^2 \end{pmatrix}^{-1}$$

Estimator for Variance

For a simple linear regression model, where $\beta = [\beta_0, \beta_1]'$ (β_0 is the y-intercept and β_1 is the slope), one obtains

$$\sigma^2(X'X)^{-1} = \sigma^2 \left(\sum x_i x_i'\right)^{-1}$$

$$= \sigma^2 \left(\sum (1, x_i)'(1, x_i)\right)^{-1}$$

$$= \sigma^2 \left(\sum \begin{pmatrix} 1 x_i \\ x_i x_i^2 \end{pmatrix}\right)^{-1}$$

$$= \sigma^2 \begin{pmatrix} N \sum x_i \\ \sum x_i \sum x_i^2 \end{pmatrix}^{-1}$$

$$= \sigma^2 \cdot \frac{1}{N \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 - \sum x_i \\ -\sum x_i N \end{pmatrix}$$

Estimator for Variance

For a simple linear regression model, where $\beta = [\beta_0, \beta_1]'$ (β_0 is the y-intercept and β_1 is the slope), one obtains

$$
\begin{aligned}
\sigma^2 (X'X)^{-1} &= \sigma^2 \left(\sum x_i x_i' \right)^{-1} \\
&= \sigma^2 \left(\sum (1, x_i)'(1, x_i) \right)^{-1} \\
&= \sigma^2 \left(\sum \begin{pmatrix} 1 x_i \\ x_i x_i^2 \end{pmatrix} \right)^{-1} \\
&= \sigma^2 \begin{pmatrix} N \sum x_i \\ \sum x_i \sum x_i^2 \end{pmatrix}^{-1} \\
&= \sigma^2 \cdot \frac{1}{N \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 - \sum x_i \\ - \sum x_i N \end{pmatrix} \\
&= \sigma^2 \cdot \frac{1}{N \sum_{i=1}^{N} (x_i - \bar{x})^2} \begin{pmatrix} \sum x_i^2 - \sum x_i \\ - \sum x_i N \end{pmatrix}
\end{aligned}
$$

Estimator for Variance

For a simple linear regression model, where $\beta = [\beta_0, \beta_1]'$ (β_0 is the y-intercept and β_1 is the slope), one obtains

$$
\begin{aligned}
\sigma^2 (X'X)^{-1} &= \sigma^2 \left(\sum x_i x_i' \right)^{-1} \\
&= \sigma^2 \left(\sum (1, x_i)'(1, x_i) \right)^{-1} \\
&= \sigma^2 \left(\sum \begin{pmatrix} 1 x_i \\ x_i x_i^2 \end{pmatrix} \right)^{-1} \\
&= \sigma^2 \begin{pmatrix} N \sum x_i \\ \sum x_i \sum x_i^2 \end{pmatrix}^{-1} \\
&= \sigma^2 \cdot \frac{1}{N \sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \sum x_i^2 - \sum x_i \\ - \sum x_i N \end{pmatrix} \\
&= \sigma^2 \cdot \frac{1}{N \sum_{i=1}^{N} (x_i - \bar{x})^2} \begin{pmatrix} \sum x_i^2 - \sum x_i \\ - \sum x_i N \end{pmatrix} \\
Var(\beta_1) &= \frac{\sigma^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}.
\end{aligned}
$$

Monte Carlo Simulations show the distribution of the estimate.
Suppose the data generating process is

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

Monte Carlo Simulations show the distribution of the estimate.
Suppose the data generating process is

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

Parameter Values for Simulations

Monte Carlo Simulations show the distribution of the estimate. Suppose the data generating process is

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

- $\beta_0 = 2.00$

Parameter Values for Simulations

Monte Carlo Simulations show the distribution of the estimate. Suppose the data generating process is

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

- $\beta_0 = 2.00$
- $\beta_1 = 0.5$

Parameter Values for Simulations

Monte Carlo Simulations show the distribution of the estimate. Suppose the data generating process is

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

- $\beta_0 = 2.00$
- $\beta_1 = 0.5$
- $u_i \sim N(0.00, 1.00)$

Monte Carlo Simulations show the distribution of the estimate.
Suppose the data generating process is

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

- $\beta_0 = 2.00$
- $\beta_1 = 0.5$
- $u_i \sim N(0.00, 1.00)$
- $N = 3$, $N = 5$, $N = 10$,
  $N = 25$, $N = 100$, $N = 1000$

Parameter Values for Simulations

Monte Carlo Simulations show the distribution of the estimate. Suppose the data generating process is

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i.$$

- $\beta_0 = 2.00$
- $\beta_1 = 0.5$
- $u_i \sim N(0.00, 1.00)$
- $N = 3, N = 5, N = 10,$
  $N = 25, N = 100, N = 1000$

Try it yourself...

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

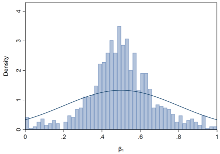
How to Establish Asymptotic Properties of $\hat{\beta}_k$?

Law of Large Numbers

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

Law of Large Numbers
As N increases, the distribution of $\hat{\beta}_k$ becomes more tightly centered around β_k.

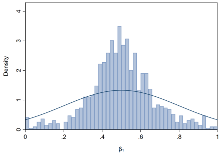
How to Establish Asymptotic Properties of $\hat{\beta}_k$?



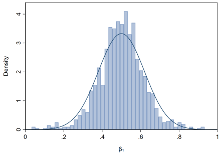
(a) N=3

Law of Large Numbers
As N increases, the distribution of $\hat{\beta}_k$ becomes more tightly centered around β_k.

Law of Large Numbers
As N increases, the distribution of $\hat{\beta}_k$ becomes more tightly centered around β_k.



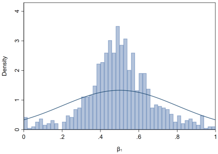
(a) N=3



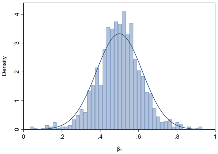
(b) N=5

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

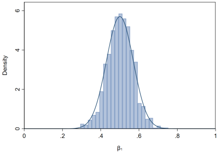
Law of Large Numbers
As N increases, the distribution of $\hat{\beta}_k$ becomes more tightly centered around β_k.



(a) N=3



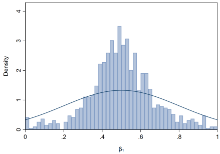
(b) N=5



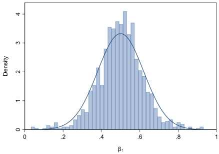
(c) N=10

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

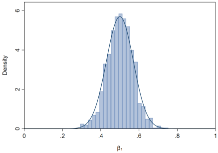
Law of Large Numbers
As N increases, the distribution of $\hat{\beta}_k$ becomes more tightly centered around β_k.



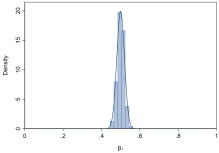
(a) N=3



(b) N=5



(c) N=10



(d) N=100

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

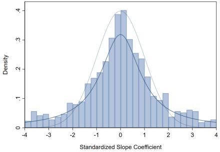
How to Establish Asymptotic Properties of $\hat{\beta}_k$?

Central Limit Theorem

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

Central Limit Theorem
As N increases, the distribution of $\hat{\beta}_k$ becomes normal (starting from a t-distribution).

How to Establish Asymptotic Properties of $\hat{\beta}_k$?



(a) N=3

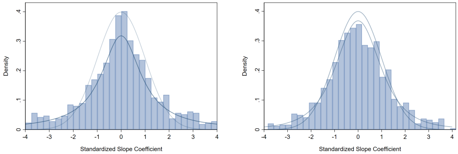
Central Limit Theorem

As N increases, the distribution of $\hat{\beta}_k$ becomes normal (starting from a t-distribution).

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

Central Limit Theorem

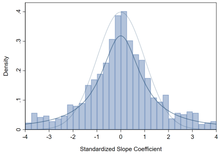
As N increases, the distribution of $\hat{\beta}_k$ becomes normal (starting from a t-distribution).



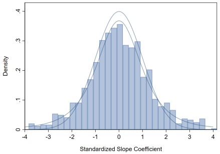
(a) N=3 (b) N=5

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

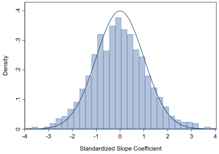
Central Limit Theorem
As N increases, the distribution of $\hat{\beta}_k$ becomes normal (starting from a t-distribution).



(a) N=3



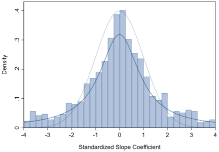
(b) N=5



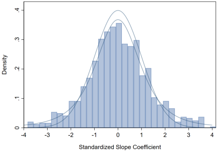
(c) N=10

How to Establish Asymptotic Properties of $\hat{\beta}_k$?

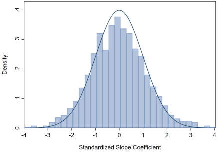
Central Limit Theorem
As N increases, the distribution of $\hat{\beta}_k$ becomes normal (starting from a t-distribution).



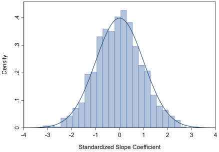
(a) N=3



(b) N=5



(c) N=10



(d) N=100

Consistency, Asymptotically Normality

Assuming LRM1,

Consistency, Asymptotically Normality

Assuming LRM1, LRM2,

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d,

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b,

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

$$plim\hat{\beta}_k = \beta_k.$$

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

$$plim\hat{\beta}_k = \beta_k.$$

That is, for all $\varepsilon > 0$

$$\lim_{N\to\infty} \Pr\left(|\hat{\beta}_k - \beta_k| > \varepsilon\right) = 0.$$

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

$$plim\hat{\beta}_k = \beta_k.$$

  That is, for all $\varepsilon > 0$

$$\lim_{N\to\infty} \Pr\left(|\hat{\beta}_k - \beta_k| > \varepsilon\right) = 0.$$

- The OLS estimator is **asymptotically normally distributed**

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

$$plim\hat{\beta}_k = \beta_k.$$

  That is, for all $\varepsilon > 0$

$$\lim_{N \to \infty} \Pr\left(|\hat{\beta}_k - \beta_k| > \varepsilon\right) = 0.$$

- The OLS estimator is **asymptotically normally distributed**

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

  $$plim\hat{\beta}_k = \beta_k.$$

  That is, for all $\varepsilon > 0$

  $$\lim_{N\to\infty} \Pr\left(|\hat{\beta}_k - \beta_k| > \varepsilon\right) = 0.$$

- The OLS estimator is **asymptotically normally distributed**

  $$\sqrt{N}(\hat{\beta}_k - \beta_k) \xrightarrow{d} N(0, Avar(\hat{\beta}_k) \times N)$$

  (Avar means asymptotic variance)

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

$$plim \hat{\beta}_k = \beta_k.$$

  That is, for all $\varepsilon > 0$

$$\lim_{N \to \infty} \Pr\left(|\hat{\beta}_k - \beta_k| > \varepsilon\right) = 0.$$

- The OLS estimator is **asymptotically normally distributed**

$$\sqrt{N}(\hat{\beta}_k - \beta_k) \xrightarrow{d} N(0, Avar(\hat{\beta}_k) \times N)$$

  (Avar means asymptotic variance)
- The OLS estimator is **approximately normally distributed**

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

  $$plim \hat{\beta}_k = \beta_k.$$

  That is, for all $\varepsilon > 0$

  $$\lim_{N \to \infty} \Pr\left(|\hat{\beta}_k - \beta_k| > \varepsilon\right) = 0.$$

- The OLS estimator is **asymptotically normally distributed**

  $$\sqrt{N}(\hat{\beta}_k - \beta_k) \xrightarrow{d} N(0, Avar(\hat{\beta}_k) \times N)$$

  (Avar means asymptotic variance)
- The OLS estimator is **approximately normally distributed**

Consistency, Asymptotically Normality

Assuming LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5 the following properties can be established using law of large numbers and central limit theorem for large samples.

- The OLS estimator is **consistent**:

$$plim\hat{\beta}_k = \beta_k.$$

  That is, for all $\varepsilon > 0$

$$\lim_{N\to\infty} \Pr\left(|\hat{\beta}_k - \beta_k| > \varepsilon\right) = 0.$$

- The OLS estimator is **asymptotically normally distributed**

$$\sqrt{N}(\hat{\beta}_k - \beta_k) \xrightarrow{d} N(0, Avar(\hat{\beta}_k) \times N)$$

  (Avar means asymptotic variance)
- The OLS estimator is **approximately normally distributed**

$$\hat{\beta}_k \overset{A}{\sim} N\left(\beta_k, Avar(\hat{\beta}_k)\right)$$

Efficiency and Asymptotic Variance

For the bivariate regression under LRM4a (homoskedasticity) it can be **consistently** estimated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_{i1} - \bar{x})^2},$$

Efficiency and Asymptotic Variance

For the bivariate regression under LRM4a (homoskedasticity) it can be **consistently** estimated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_{i1} - \bar{x})^2},$$

with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2}.$$

For the bivariate regression under LRM4a (homoskedasticity) it can be **consistently** estimated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(x_{i1} - \bar{x})^2},$$

with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}\hat{u}_i^2}{N-2}.$$

Under LRMb (heteroskedasticity), $Avar(\hat{\beta})$ can be **consistently** estimated as the *robust* or *Eicker-Huber-White* estimator.

For the bivariate regression under LRM4a (homoskedasticity) it can be **consistently** estimated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} (x_{i1} - \bar{x})^2},$$

with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2}.$$

Under LRMb (heteroskedasticity), $Avar(\hat{\beta})$ can be **consistently** estimated as the *robust* or *Eicker-Huber-White* estimator.

The robust variance estimator is calculated as

Efficiency and Asymptotic Variance

For the bivariate regression under LRM4a (homoskedasticity) it can be **consistently** estimated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} (x_{i1} - \bar{x})^2},$$

with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2}.$$

Under LRMb (heteroskedasticity), $Avar(\hat{\beta})$ can be **consistently** estimated as the *robust* or *Eicker-Huber-White* estimator.

The robust variance estimator is calculated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\sum_{i=1}^{N} \hat{u}_i^2 (x_{i1} - \bar{x})^2}{\left[\sum_{i=1}^{N} (x_{i1} - \bar{x})^2\right]}.$$

Efficiency and Asymptotic Variance

For the bivariate regression under LRM4a (homoskedasticity) it can be **consistently** estimated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} (x_{i1} - \bar{x})^2},$$

with

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2}.$$

Under LRMb (heteroskedasticity), $Avar(\hat{\beta})$ can be **consistently** estimated as the *robust* or *Eicker-Huber-White* estimator.

The robust variance estimator is calculated as

$$\widehat{Avar}(\hat{\beta}_1) = \frac{\sum_{i=1}^{N} \hat{u}_i^2 (x_{i1} - \bar{x})^2}{\left[\sum_{i=1}^{N} (x_{i1} - \bar{x})^2\right]}.$$

Note: In practice we can almost never be sure that the errors are homoskedastic and should therefore always use robust standard errors.

Sketch of Proof for Asymptotic Properties

- The OLS estimator of $\hat{\beta}$ is consistent and asymptotic normal
  Estimator $\hat{\beta}$ can be written as: $\hat{\beta} = \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'y =$
  $\beta + \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'u = \beta \ + \ \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$

Sketch of Proof for Asymptotic Properties

- The OLS estimator of $\hat{\beta}$ is consistent and asymptotic normal Estimator $\hat{\beta}$ can be written as: $\hat{\beta} = \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'y =$
  $\beta + \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'u = \beta \ + \ \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$

Sketch of Proof for Asymptotic Properties

- The OLS estimator of $\hat{\beta}$ is consistent and asymptotic normal
  Estimator $\hat{\beta}$ can be written as: $\hat{\beta} = \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'y =$
  $\beta + \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'u = \beta \ + \ \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$

  We can use the law of large numbers to establish that :
  $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \ \xrightarrow{p}\ \mathrm{E}[x_i x_i'] = \frac{Q_{xx}}{N}, \qquad \frac{1}{N}\sum_{i=1}^{N} x_i u_i \ \xrightarrow{p}\ \mathrm{E}[x_i u_i] = 0$

Sketch of Proof for Asymptotic Properties

- The OLS estimator of $\hat{\beta}$ is consistent and asymptotic normal
  Estimator $\hat{\beta}$ can be written as: $\hat{\beta} = \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'y =$
  $\beta + \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'u = \beta \; + \; \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$
  We can use the law of large numbers to establish that :
  $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \;\xrightarrow{p}\; \mathrm{E}[x_i x_i'] = \frac{Q_{xx}}{N}, \qquad \frac{1}{N}\sum_{i=1}^{N} x_i u_i \;\xrightarrow{p}\; \mathrm{E}[x_i u_i] = 0$
  By Slutsky's theorem and continuous mapping theorem these results can be combined to establish consistency of estimator $\hat{\beta}$:
  $\hat{\beta} \;\xrightarrow{p}\; \beta + Q_{xx}^{-1} \cdot 0 = \beta$

Sketch of Proof for Asymptotic Properties

- The OLS estimator of $\hat{\beta}$ is consistent and asymptotic normal
  Estimator $\hat{\beta}$ can be written as: $\hat{\beta} = \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'y =$
  $\beta + \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'u = \beta \ + \ \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$

  We can use the law of large numbers to establish that :
  $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \ \xrightarrow{p} \ \mathrm{E}[x_i x_i'] = \frac{Q_{xx}}{N}, \qquad \frac{1}{N}\sum_{i=1}^{N} x_i u_i \ \xrightarrow{p} \ \mathrm{E}[x_i u_i] = 0$

  By Slutsky's theorem and continuous mapping theorem these results can be combined to establish consistency of estimator $\hat{\beta}$:
  $\hat{\beta} \ \xrightarrow{p} \ \beta + Q_{xx}^{-1} \cdot 0 = \beta$

  The central limit theorem tells us that: $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i u_i \ \xrightarrow{d} \ \mathcal{N}(0,\ V)$,
  where $V = \mathrm{Var}[x_i u_i] = \mathrm{E}[\,u_i^2 x_i x_i'\,] = \mathrm{E}\left[\ \mathrm{E}[u_i^2|x_i]\ x_i x_i'\,\right] = \sigma^2 \frac{Q_{xx}}{N}$

- The OLS estimator of $\hat{\beta}$ is consistent and asymptotic normal
  Estimator $\hat{\beta}$ can be written as: $\hat{\beta} = \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'y =$
  $\beta + \left(\frac{1}{N}X'X\right)^{-1}\frac{1}{N}X'u = \beta \ + \ \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$

  We can use the law of large numbers to establish that :
  $\frac{1}{N}\sum_{i=1}^{N} x_i x_i' \ \xrightarrow{p} \ \mathrm{E}[x_i x_i'] = \frac{Q_{xx}}{N}, \qquad \frac{1}{N}\sum_{i=1}^{N} x_i u_i \ \xrightarrow{p} \ \mathrm{E}[x_i u_i] = 0$
  By Slutsky's theorem and continuous mapping theorem these results can be combined to establish consistency of estimator $\hat{\beta}$:
  $\hat{\beta} \ \xrightarrow{p} \ \beta + Q_{xx}^{-1} \cdot 0 = \beta$

  The central limit theorem tells us that: $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i u_i \ \xrightarrow{d} \ \mathcal{N}(0, \ V)$,
  where $V = \mathrm{Var}[x_i u_i] = \mathrm{E}[\, u_i^2 x_i x_i' \,] = \mathrm{E}\left[\ \mathrm{E}[u_i^2|x_i]\ x_i x_i' \,\right] = \sigma^2 \frac{Q_{xx}}{N}$
  Applying Slutsky's theorem again we'll have:
  $\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i u_i\right) \ \xrightarrow{d}$
  $Q_{xx}^{-1} N \cdot \mathcal{N}(0, \sigma^2 \frac{Q_{xx}}{N}) = \mathcal{N}(0, \sigma^2 Q_{xx}^{-1} N)$ $\square$

OLS Properties in the Small and in the Large

Set of assumptions	(1)	(2)	(3)	(4)	(5)	(6)
LRM1: linearity		f u	l f	i l	l e	d
LRM2: simple random sampling		f u	l f	i l	l e	d
LRM5: identifiability		f u	l f	i l	l e	d
LRM4: error variance						
- LRM4a: homoskedastic	✓	✓	✓	×	×	×
- LRM4b: heteroskedastic	×	×	×	✓	✓	✓
LRM3: exogeneity						
- LRM3a: normality	✓	×	×	✓	×	×
- LRM3b: independent	✓	✓	×	×	×	×
- LRM3c: mean indep.	✓	✓	✓	✓	✓	×
- LRM3d: uncorrelated	✓	✓	✓	✓	✓	✓
Small sample properties of $\hat{\beta}$						
- unbiased	✓	✓	✓	✓	✓	×
- normally distributed	✓	×	×	✓	×	×
- efficient	✓	✓	✓	×	×	×
Large sample properties of $\hat{\beta}$						
- consistent	✓	✓	✓	✓	✓	✓
- approx. normal	✓	✓	✓	✓	✓	✓
- asymptotically efficient	✓	✓	✓	×	×	×

- *Notes:* ✓ = fulfilled, × = violated

Tests in Small Samples I

Assume LRM1,

Tests in Small Samples I

Assume LRM1, LRM2,

Assume LRM1, LRM2, LRM3a,

Tests in Small Samples I

Assume LRM1, LRM2, LRM3a, LRM4a,

Tests in Small Samples I

Assume LRM1, LRM2, LRM3a, LRM4a, and LRM5.

Tests in Small Samples I

Assume LRM1, LRM2, LRM3a, LRM4a, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the t-**test**.
If the null hypotheses is true, the t-statistic

Assume LRM1, LRM2, LRM3a, LRM4a, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the t-**test**.
If the null hypotheses is true, the t-statistic

$$t = \frac{\hat{\beta}_k - q}{\widehat{se}(\hat{\beta}_k)} \sim t_{N-K-1}$$

Tests in Small Samples I

Assume LRM1, LRM2, LRM3a, LRM4a, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the t-**test**.
If the null hypotheses is true, the t-statistic

$$t = \frac{\hat{\beta}_k - q}{\widehat{se}(\hat{\beta}_k)} \sim t_{N-K-1}$$

follows a t-distribution with $N - K - 1$ degrees of freedom. The standard error is $\widehat{se}(\hat{\beta}_k) = \sqrt{\hat{V}(\hat{\beta}_k)}$.

Assume LRM1, LRM2, LRM3a, LRM4a, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the t-**test**.
If the null hypotheses is true, the t-statistic

$$t = \frac{\hat{\beta}_k - q}{\widehat{se}(\hat{\beta}_k)} \sim t_{N-K-1}$$

follows a t-distribution with $N - K - 1$ degrees of freedom. The standard error is $\widehat{se}(\hat{\beta}_k) = \sqrt{\hat{V}(\hat{\beta}_k)}$.

For example, to perform a two-sided test of H_0 against the alternative hypotheses $H_A : \beta_k \neq q$ on the 5% significance level, we calculate the t-statistic and compare its absolute value to the 0.975-quantile of the t-distribution. With $N = 30$ and $K = 2$, H_0 is rejected if $|t| > 2.052$.

Tests in Small Samples II

A null hypotheses of the form $H_0 : r_{j1}\beta_1 + \ldots + r_{jK}\beta_K = q_j$, in matrix notation $H_0 : R\beta = q$, with J linear restrictions $j = 1 \ldots J$ is jointly tested with the F-**test**.

Tests in Small Samples II

A null hypotheses of the form $H_0 : r_{j1}\beta_1 + \ldots + r_{jK}\beta_K = q_j$, in matrix notation $H_0 : R\beta = q$, with J linear restrictions $j = 1 \ldots J$ is jointly tested with the F-**test**.

If the null hypotheses is true, the F-statistic follows an F distribution with J numerator degrees of freedom and $N - K - 1$ denominator degrees of freedom:

Tests in Small Samples II

A null hypotheses of the form $H_0 : r_{j1}\beta_1 + \ldots + r_{jK}\beta_K = q_j$, in matrix notation $H_0 : R\beta = q$, with J linear restrictions $j = 1 \ldots J$ is jointly tested with the F-**test**.

If the null hypotheses is true, the F-statistic follows an F distribution with J numerator degrees of freedom and $N - K - 1$ denominator degrees of freedom:

$$F = \frac{\left(R\hat{\beta} - q\right)' \left[R\hat{V}(\hat{\beta}|X)R'\right]^{-1} \left(R\hat{\beta} - q\right)}{J} \sim F_{J,N-K-1}.$$

Tests in Small Samples II

A null hypotheses of the form $H_0: r_{j1}\beta_1 + \ldots + r_{jK}\beta_K = q_j$, in matrix notation $H_0: R\beta = q$, with J linear restrictions $j = 1\ldots J$ is jointly tested with the F-**test**.

If the null hypotheses is true, the F-statistic follows an F distribution with J numerator degrees of freedom and $N-K-1$ denominator degrees of freedom:

$$F = \frac{\left(R\hat{\beta}-q\right)'\left[R\hat{V}(\hat{\beta}|X)R'\right]^{-1}\left(R\hat{\beta}-q\right)}{J} \sim F_{J,N-K-1}.$$

For example, to perform a two-sided test of H_0 against the alternative hypotheses $H_A: r_{j1}\beta_1 + \ldots + r_{jK}\beta_K \neq q_j$ for all j at the 5% significance level, we calculate the F-statistic and compare it to the 0.95-quantile of the F-distribution.

With $N = 30, K = 2$ and $J = 2$, H_0 is rejected if $F > 3.35$. We cannot perform two-sided F-tests because the F distribution has one tail.

Tests in Small Samples III

Only under homoskedasticity (LRM4a), the F-statistic can also be computed as

Only under homoskedasticity (LRM4a), the F-statistic can also be computed as

$$F = \frac{(R^2 - R^2_{\text{restricted}})/J}{(1 - R^2)/(N - K - 1)} \sim F_{J,N-K-1},$$

Tests in Small Samples III

Only under homoskedasticity (LRM4a), the F-statistic can also be computed as

$$F = \frac{(R^2 - R^2_{\text{restricted}})/J}{(1 - R^2)/(N - K - 1)} \sim F_{J,N-K-1},$$

where $R^2_{\text{restricted}}$ is estimated by restricted least squares which minimizes $SD(\beta)$ s.t. $r_{j1}\beta_1 + \ldots + r_{jK}\beta_K \neq q_j$ for all j.

Tests in Small Samples III

Only under homoskedasticity (LRM4a), the F-statistic can also be computed as

$$F = \frac{(R^2 - R^2_{\text{restricted}})/J}{(1 - R^2)/(N - K - 1)} \sim F_{J,N-K-1},$$

where $R^2_{\text{restricted}}$ is estimated by restricted least squares which minimizes $SD(\beta)$ s.t. $r_{j1}\beta_1 + \ldots + r_{jK}\beta_K \neq q_j$ for all j.

Exclusionary restrictions of the form $H_0 : \beta_k = 0, \beta_m = 0, \ldots$ are a special case of $H_0 : r_{j1}\beta_1 + \ldots + r_{jK}\beta_K = q_j$ for all j. In this case, restricted least squares is simply estimated as a regression were the explanatory variables $k, m, \ldots$ are excluded, e.g. a regression with a constant only.

If the F distribution has degrees of freedom (df) 1 as the numerator df, and $N - K - 1$ as the denominator df, then it can be shown that $t^2 = F(1, N - K - 1)$.

Assuming LRM1,

Confidence Intervals in Small Samples

Assuming LRM1, LRM2,

Assuming LRM1, LRM2, LRM3a,

Assuming LRM1, LRM2, LRM3a, LRM4a, and

Assuming LRM1, LRM2, LRM3a, LRM4a, and LRM5, Assuming LRM1, LRM2, LRM3a, LRM4a, and LRM5,we can construct confidence intervals for a particular coefficient β_k.

Assuming LRM1, LRM2, LRM3a, LRM4a, and LRM5,we can construct confidence intervals for a particular coefficient β_k.The $(1-\alpha)$ confidence interval is given by

$$\left(\hat{\beta}_k - t_{(1-\alpha/2),(N-K-1)}\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + t_{(1-\alpha/2),(N-K-1)}\widehat{se}(\hat{\beta}_k)\right),$$

Assuming LRM1, LRM2, LRM3a, LRM4a, and LRM5,we can construct confidence intervals for a particular coefficient β_k.The $(1-\alpha)$ confidence interval is given by

$$\left(\hat{\beta}_k - t_{(1-\alpha/2),(N-K-1)}\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + t_{(1-\alpha/2),(N-K-1)}\widehat{se}(\hat{\beta}_k)\right),$$

where $t_{(1-\alpha/2),(N-K-1)}$ is the $(1-\alpha/2)$ quantile of the t-distribution with $(N-K-1)$ degrees of freedom.

Assuming LRM1, LRM2, LRM3a, LRM4a, and LRM5,we can construct confidence intervals for a particular coefficient β_k.The $(1-\alpha)$ confidence interval is given by

$$\left(\hat{\beta}_k - t_{(1-\alpha/2),(N-K-1)}\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + t_{(1-\alpha/2),(N-K-1)}\widehat{se}(\hat{\beta}_k)\right),$$

where $t_{(1-\alpha/2),(N-K-1)}$ is the $(1-\alpha/2)$ quantile of the t-distribution with $(N-K-1)$ degrees of freedom. For example, the 95% confidence interval with $N=30$ and $K=2$ is $\left(\hat{\beta}_k - 2.052\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + 2.052\widehat{se}(\hat{\beta}_k)\right)$.

Recall: α is the maximum acceptable probability of a Type I error.

Null hypothesis (H_0)	is valid (Innocent)	is invalid (Guilty)
Reject H_0	**Type I ($\alpha = 0.05$) error**	Correct outcome
I think he is guilty!	False positive Convicted!	True positive Convicted!
Don't reject H_0	Correct outcome	**Type II (β) error**
I think he is innocent!	True negative Freed!	False negative Freed!

Assume LRM1,

Asymptotic Tests

Assume LRM1, LRM2,

Assume LRM1, LRM2, LRM3d,

Asymptotic Tests

Assume LRM1, LRM2, LRM3d, LRM4a or LRM4b, Assume LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5.

Assume LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the z-**test**. If the null hypotheses is true, the z-statistic

Assume LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the z-**test**. If the null hypotheses is true, the z-statistic

$$z = \frac{\hat{\beta}_k - q}{\widehat{se}(\hat{\beta}_k)} \overset{A}{\sim} N(0, 1)$$

Assume LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the z-**test**. If the null hypotheses is true, the z-statistic

$$z = \frac{\hat{\beta}_k - q}{\widehat{se}(\hat{\beta}_k)} \overset{A}{\sim} N(0,1)$$

follows approximately the standard normal distribution. The standard error is $\widehat{se}(\hat{\beta}_k) = \sqrt{\widehat{Avar}(\hat{\beta}_k)}$.

Asymptotic Tests

Assume LRM1, LRM2, LRM3d, LRM4a or LRM4b, and LRM5. A simple null hypotheses of the form $H_0 : \beta_k = q$ is tested with the z-**test**. If the null hypotheses is true, the z-statistic

$$z = \frac{\hat{\beta}_k - q}{\widehat{se}(\hat{\beta}_k)} \overset{A}{\sim} N(0,1)$$

follows approximately the standard normal distribution. The standard error is $\widehat{se}(\hat{\beta}_k) = \sqrt{\widehat{Avar}(\hat{\beta}_k)}$.

For example, to perform a two sided test of H_0 against the alternative hypotheses $H_A : \beta_k \neq q$ on the 5% significance level, we calculate the z-statistic and compare its absolute value to the 0.975-quantile of the standard normal distribution. H_0 is rejected if $|z| > 1.96$.

We talk about the Wald test later...

Confidence Intervals in Large Samples

Assuming LRM1,

Confidence Intervals in Large Samples

Assuming LRM1, LRM2,

Assuming LRM1, LRM2, LRM3d,

Confidence Intervals in Large Samples

Assuming LRM1, LRM2, LRM3d, LRM5,

Assuming LRM1, LRM2, LRM3d, LRM5, and LRM4a or LRM4b, Assuming LRM1, LRM2, LRM3d, LRM5, and LRM4a or LRM4b, we can construct confidence intervals for a particular coefficient β_k.

Assuming LRM1, LRM2, LRM3d, LRM5, and LRM4a or LRM4b, we can construct confidence intervals for a particular coefficient β_k. The $(1-\alpha)$ confidence interval is given by

$$\left(\hat{\beta}_k - z_{(1-\alpha/2)}\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + z_{(1-\alpha/2)}\widehat{se}(\hat{\beta}_k)\right)$$

Assuming LRM1, LRM2, LRM3d, LRM5, and LRM4a or LRM4b, we can construct confidence intervals for a particular coefficient β_k. The $(1-\alpha)$ confidence interval is given by

$$\left(\hat{\beta}_k - z_{(1-\alpha/2)}\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + z_{(1-\alpha/2)}\widehat{se}(\hat{\beta}_k)\right)$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

Assuming LRM1, LRM2, LRM3d, LRM5, and LRM4a or LRM4b, we can construct confidence intervals for a particular coefficient β_k. The $(1-\alpha)$ confidence interval is given by

$$\left(\hat{\beta}_k - z_{(1-\alpha/2)}\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + z_{(1-\alpha/2)}\widehat{se}(\hat{\beta}_k)\right)$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

For example, the 95% confidence interval is $\left(\hat{\beta}_k - 1.96\widehat{se}(\hat{\beta}_k), \hat{\beta}_k + 1.96\widehat{se}(\hat{\beta}_k)\right)$.

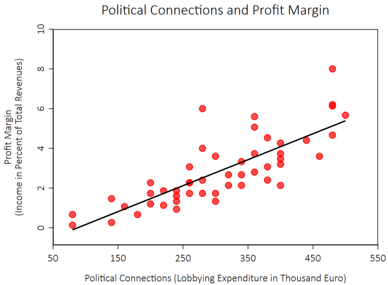
OLS Properties in the Small and in the Large

Set of assumptions	(1)	(2)	(3)	(4)	(5)	(6)
LRM1: linearity	fulfilled					
LRM2: simple random sampling	fulfilled					
LRM5: identifiability	fulfilled					
LRM4: error variance						
- LRM4a: homoskedastic	✓	✓	✓	×	×	×
- LRM4b: heteroskedastic	×	×	×	✓	✓	✓
LRM3: exogeneity						
- LRM3a: normality	✓	×	×	✓	×	×
- LRM3b: independent	✓	✓	×	×	×	×
- LRM3c: mean indep.	✓	✓	✓	✓	✓	×
- LRM3d: uncorrelated	✓	✓	✓	✓	✓	✓
Small sample properties of $\hat{\beta}$						
- unbiased	✓	✓	✓	✓	✓	×
- normally distributed	✓	×	×	✓	×	×
- efficient	✓	✓	✓	×	×	×
t-test, F-test	✓	×	×	×	×	×
Large sample properties of $\hat{\beta}$						
- consistent	✓	✓	✓	✓	✓	✓
- approx. normal	✓	✓	✓	✓	✓	✓
- asymptotically efficient	✓	✓	✓	×	×	×
z-test, Wald test	✓	✓	✓	✓*	✓*	✓*

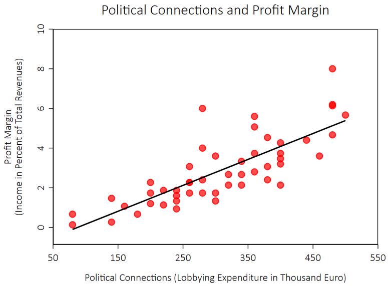
- *Notes:* ✓ = fulfilled, × = violated, * = corrected standard errors.

Content

Arguments **For** Causality of Effect



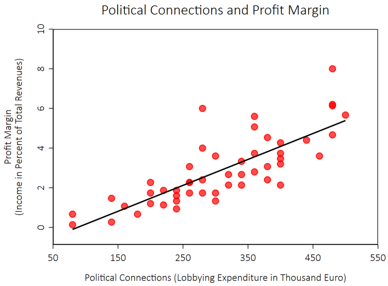
Arguments **For** Causality of Effect



Econometric methods need to address concerns, including:

- **Misspecification:** Results robust to different functional forms

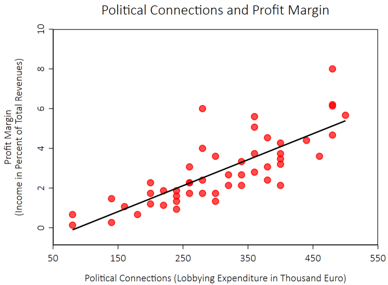
Arguments **For** Causality of Effect



Econometric methods need to address concerns, including:

- **Misspecification:** Results robust to different functional forms
- **Errors-in-variables:** little concern with administrative data

Arguments **For** Causality of Effect



Econometric methods need to address concerns, including:

- **Misspecification:** Results robust to different functional forms
- **Errors-in-variables:** little concern with administrative data
- **External validity:** Similar effect found in independent studies.

- **Omitted variable bias:**
  e.g., business acumen
  $\rightarrow$ Panel data models

Arguments **Against** Causality of Effect

- **Omitted variable bias:**
  e.g., business acumen
  $\rightarrow$ Panel data models
- **Sample selection bias:**
  lobbying expenditures only observed if in transparency register.
  $\rightarrow$ Selection correction models

Arguments **Against** Causality of Effect

- **Omitted variable bias:**
  e.g., business acumen
  $\rightarrow$ Panel data models
- **Sample selection bias:**
  lobbying expenditures only observed if in transparency register.
  $\rightarrow$ Selection correction models
- **Simultaneous causality:**

Arguments **Against** Causality of Effect

- **Omitted variable bias:**
  e.g., business acumen
  $\rightarrow$ Panel data models
- **Sample selection bias:**
  lobbying expenditures only observed if in transparency register.
  $\rightarrow$ Selection correction models
- **Simultaneous causality:**
  - profits may be higher because of political connections

Arguments **Against** Causality of Effect

- **Omitted variable bias:**
  e.g., business acumen
  $\rightarrow$ Panel data models
- **Sample selection bias:**
  lobbying expenditures only observed if in transparency register.
  $\rightarrow$ Selection correction models
- **Simultaneous causality:**
  - profits may be higher because of political connections
  - firms may become connected because of their high profits

Arguments **Against** Causality of Effect

- **Omitted variable bias:**
  e.g., business acumen
  $\rightarrow$ Panel data models
- **Sample selection bias:**
  lobbying expenditures only observed if in transparency register.
  $\rightarrow$ Selection correction models
- **Simultaneous causality:**
  - profits may be higher because of political connections
  - firms may become connected because of their high profits

Arguments **Against** Causality of Effect

- **Omitted variable bias:**
  e.g., business acumen
  $\rightarrow$ Panel data models
- **Sample selection bias:**
  lobbying expenditures only observed if in transparency register.
  $\rightarrow$ Selection correction models
- **Simultaneous causality:**
  - profits may be higher because of political connections
  - firms may become connected because of their high profits

All of those concerns may be addressed with $\rightarrow$instrumental variable models. What would be a good instrument/experiment?

ANGRIST, J. D., AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Chapter 3.

CAMERON, A. C., AND P. K. TRIVEDI (2005): *Microeconometrics: Methods and Applications*. Cambridge University Press, 3 edn., Section 4.1–4.5, 4.8–4.9.

HILL, R. C., W. E. GRIFFITHS, AND G. C. LIM (2010): *Principles of Econometrics*. John Wiley & Sons, 4 edn., Chapters 1–6, 8, 10.3.

KENNEDY, P. (2008): *A Guide to Econometrics*. Blackwell Publishing, 6 edn., Chapter 3–11, 14, in particular 7, 8.1–8.3.

PLACKETT, R. L. (1972): "Studies in the History of Probability and Statistics. XXIX: The discovery of the method of least squares," *Biometrika*, 59(2), 239–251.

ROSTAM-AFSCHAR, D., AND R. JESSEN (2014): "GRAPH3D: Stata module to draw colored, scalable, rotatable 3D plots," Statistical Software Components, Boston College Department of Economics.

References II

Stock, J. H., and M. W. Watson (2012): *Introduction to Econometrics*. Pearson Addison-Wesley, 3 edn., Chapters 4–9, 12.

Verbeek, M. (2012): *A Guide to Modern Econometrics*. John Wiley & Sons, 3 edn., Chapter 2–5.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Sections 4.1–4.23 and 6.24, Chapter 5.

——— (2009): *Introductory Econometrics: A Modern Approach*. Cengage Learning, 4 edn., Chapters 1–8, 15.